

Forward-Backward EM-TV methods for inverse problems with Poisson noise

Christoph Brune¹, Martin Burger², Alex Sawatzky²,
Thomas Kösters³ and Frank Wübbeling²

¹ Department of Mathematics, University of California Los Angeles, CA
90095-1555, USA

² Institute for Computational and Applied Mathematics, University of Münster,
Orleans-Ring 10, 48149 Münster, Germany

³ European Institute for Molecular Imaging, Mendelstr. 11, 48149 Münster,
Germany

E-mail:

¹brune@math.ucla.edu, ²martin.burger@wwu.de, ²alex.sawatzky@wwu.de,

³tkoesters@uni-muenster.de, ²frank.wuebbeling@math.uni-muenster.de

Abstract. We address the task of reconstructing images corrupted by Poisson noise, which is important in various applications, such as fluorescence microscopy, positron emission tomography (PET), or astronomical imaging. In this work, we focus on reconstruction strategies, combining the expectation-maximization (EM) algorithm and total variation (TV) based regularization, and present a detailed analysis as well as numerical results.

Recently extensions of the well known EM/Richardson-Lucy algorithm received increasing attention for inverse problems with Poisson data. However, most algorithms for regularizations like TV lead to convergence problems for large regularization parameters, cannot guarantee positivity, and rely on additional approximations (like smoothed TV).

The goal of this work is to provide an accurate, robust and fast FB-EM-TV method for computing cartoon reconstructions facilitating post-segmentation and further image quantifications. Motivated by several applications we provide a statistical modeling of inverse problems with Poisson noise in terms of Bayesian MAP estimation and relate it to the continuous variational setting. We focus on minimizing the energy functional with the Kullback-Leibler divergence as the data fidelity and TV as regularization, subject to non-negativity constraints. Our proposed FB-EM-TV minimization algorithm is a semi-implicit, alternating two step method consisting of an EM step and the solution of a weighted ROF problem. The method can be reinterpreted as a modified forward-backward (FB) splitting strategy known from convex optimization. First of all, we establish the well-posedness of the variational problem under general conditions, in particular we give a proof of existence, uniqueness and stability. Under certain assumptions on the given data, we can prove positivity preservation of our iteration method. A damped variant of the FB-EM-TV algorithm, interpreted as a splitting strategy with modified time steps, is the key step towards global convergence. In addition, we present a Bregman-FB-EM-TV strategy, extending the FB-EM-TV framework, which corrects the natural loss of contrast using TV via iterative Bregman distance regularization.

Finally, we illustrate the performance of the proposed algorithms and confirm the analytical concepts by 2D and 3D synthetic and real-world results in optical nanoscopy and positron emission tomography.

1. Introduction

Image reconstruction is a fundamental problem in several areas of applied sciences, such as medical imaging, optical microscopy, or astronomy. A prominent example is positron emission tomography (PET), a biomedical imaging technique in nuclear medicine that generates images of living organisms by visualization of weak radioactively marked pharmaceuticals, so-called tracers. Due to the possibility of measuring temporal tracer uptake (from list-mode data), this modality is particularly suitable for investigating physiological and biochemical processes. Another application of image reconstruction is fluorescence microscopy. It represents an important technique for investigating biological living cells at nanoscales. In this type of applications image reconstruction arises in terms of deconvolution problems, where undesired blurring effects are caused by diffraction of light.

Mathematically, image reconstruction in those applications can be formulated as a linear inverse and ill-posed problem. Typically, in such problems one has to deal with Fredholm integral equations of the first kind, or more general

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

with a compact linear operator \bar{K} , 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 approximate solution u of \bar{u} from

$$f = Ku \tag{1}$$

is wanted. However, the computation of u by a direct inversion of K is not reasonable since (1) is ill-posed. In this case, regularization techniques are required to enforce stability during the inversion process and to compute useful reconstructions.

A commonly used idea to realize regularization techniques with statistical motivation is the Bayesian model, using the posterior probability density $p(u|f)$, given according to Bayes formula

$$p(u|f) \sim p(f|u)p(u) . \tag{2}$$

The computationally interesting Bayesian approach is the maximum a-posteriori probability (MAP) estimation, which consists of computing an estimate u of the unknown object by maximizing the a-posteriori probability density $p(u|f)$. If the measurements f are available, the density $p(u|f)$ is denoted as the a-posteriori likelihood function, just depending on u . The Bayesian approach (2) has the advantage that it allows for incorporating additional prior information on u via the a-priori probability density $p(u)$ into the reconstruction process. The most frequently used a-priori densities are Gibbs functions [39, 40],

$$p(u) \sim e^{-\alpha R(u)} , \tag{3}$$

where α denotes a positive parameter and R a convex energy functional. Typical examples for probability densities $p(f|u)$ in (2) are exponentially distributed raw data f . In the canonical case of additive white Gaussian noise with expected value 0 and variance σ^2 one finds

$$p(f|u) \sim e^{-\frac{1}{2\sigma^2} \|Ku - f\|_{L^2(\Sigma)}^2} ,$$

and the minimization of the negative log-likelihood function leads to the classical Tikhonov regularization methods [16], based on minimizing a functional of the form

$$\min_{u \geq 0} \frac{1}{2} \|Ku - f\|_{L^2(\Sigma)}^2 + \alpha R(u) . \tag{4}$$

The first term, so-called data fidelity term, penalizes the deviation from equation (1), while $R(u)$ is a convex regularization term penalizing deviations from a certain ideal structure (smoothness) of the solution. If we choose K as the identity operator and the total variation (TV) functional as the regularization energy, i.e. $R(u) := |u|_{BV(\Omega)}$ as in (6), we obtain the well known ROF model [70] for image denoising. An additional positivity constraint as in (4) is essential for common applications, since the unknown functions u usually represent densities or intensity information.

In the applications mentioned above the measured data are stochastic due to the radioactive decay of tracers in PET imaging and to laser scanning techniques in fluorescence microscopy. The random variables of the measured data in those applications are not Gaussian but Poisson distributed [80] with expected values given by $(Ku)_i$,

$$p(f|u) = \prod_i \frac{(Ku)_i^{f_i}}{f_i!} e^{-(Ku)_i} .$$

In this work, we will concentrate on MAP estimates for inverse problems with Poisson distributed data. The MAP estimation via the negative log-likelihood function (2) asymptotically leads to the following variational problem [16],

$$\min_{u \geq 0} \int_{\Sigma} (Ku - f \log Ku) d\mu + \alpha R(u) . \quad (5)$$

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 the two probability measures f and Ku . A particular complication of (5) compared to (4) is the strong nonlinearity in the data fidelity term and resulting issues in the computation of minimizers.

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 strategies receive 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 PET or fluorescence microscopy, which cannot be taken into account by direct methods. Thus, in this work we deal with extensions of the expectation-maximization (EM) or Richardson-Lucy algorithm [28, 55, 69], which is a popular iterative reconstruction method to compute (5) in the absence of regularization ($R \equiv 0$) with incomplete Poisson data f [76]. However, the EM algorithm is difficult to be generalized to regularized cases. The robust and accurate solutions of this problem for appropriate models of R and its analysis are the main contributions of this paper.

The specific choice of the regularization functional R in (5) is crucial for 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 simplicity in analysis and computation. However, such regularization approaches always lead to blurring of 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 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, which has been derived as a denoising technique in [70] and has been generalized to various other imaging tasks subsequently. The exact definition of TV [1], used in this paper, is

$$R(u) := |u|_{BV(\Omega)} = \sup_{\substack{g \in C_0^\infty(\Omega, R^d) \\ \|g\|_\infty \leq 1}} \int_{\Omega} u \operatorname{div} g, \quad (6)$$

which is formally (true if u is sufficiently regular)

$$|u|_{BV(\Omega)} = \int_{\Omega} |\nabla u|. \quad (7)$$

The space of functions with bounded total variation is denoted by $BV(\Omega)$. For further properties and details of BV functions, we refer to [1, 36]. The motivation for using TV is the effective suppression of noise and the realization of almost homogeneous regions with sharp edges. These features are attractive for PET and nanoscopic imaging if 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 (5), for instance in case of Tikhonov regularization [10], diffusion regularization [11] or L^1 regularization functional [56]. However, most works deal with TV regularization functionals (6) or (7), e.g. in application to PET [50, 64], deconvolution problems [12, 29, 37, 75], or denoising problems with identity operator K [52], but still with some restrictions. These limitations can be traced back to the strong computational difficulties in the minimization of (5) with TV regularization and can be separated in two main problems:

- The methods use the exact definition of total variation in (6), but require an inversion of the operator K^*K , where K^* is the adjoint operator of K . For example, the authors in [37] and [75] 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 or discrete cosine transform. Additionally, in contrast to the PIDSplit+ algorithm in [75], the PIDAL algorithm in [37] ensures that Ku is nonnegative and not that the final solution u is nonnegative, which however is essential in (5).
- To overcome the non differentiability of the TV regularization functional, the other class of methods uses an approximation of (7) by differentiable functionals,

$$|u|_{BV(\Omega)}^\epsilon = \int_{\Omega} \sqrt{|\nabla u|^2 + \epsilon}, \quad \epsilon > 0,$$

and creates blurring effects in reconstructed images. In [7], 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 suggested in [29, 50, 64] are realized as elementary modifications of the EM algorithm 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 these algorithms are even less robust.

In this work, we propose a robust algorithm for the TV regularized Poisson likelihood estimation problem (5) without approximation of TV, i.e. we use (6) respectively a dual version. This enables us to realize cartoon reconstructions with sharp edges. For this purpose, we use a forward-backward (FB) splitting approach [54, 65, 79] for this framework, that can be realized by alternating a classical EM reconstruction step and solving a modified version of the ROF problem analogous to the projected gradient descent algorithm of Chambolle in [22]. The first advantage of our approach is that the EM algorithm does not need any inversion of the operator K and hence is applicable for an arbitrary operator, e.g. the Radon or X-ray transform [62]. 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, this strategy 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. Additionally, we study the existence, uniqueness and stability of the minimization problem, prove positivity preservation of the algorithm and provide a convergence analysis for a damped FB splitting strategy.

It is well known that images reconstructed with TV-based methods suffer from a loss of contrast. Hence, we suggest to extend the TV regularized Poisson likelihood estimation problem (5), and with it also the FB-EM-TV algorithm, to an iterative regularization strategy using Bregman iterations, which incorporates simultaneous contrast correction. The contrast improvement is realized via inverse scale space methods and Bregman distance iterations, introduced in [63, 20, 21]. Related to these methods, an iterative contrast correction can be implemented as a sequence of modified Poisson likelihood estimation problems (5) with

$$R(u) := D_{|\cdot|_{BV(\Omega)}}^{\tilde{p}}(u, \tilde{u}), \quad \tilde{p} \in \partial |\tilde{u}|_{BV(\Omega)},$$

as the regularization functional. The Bregman distance based on TV, $D_{|\cdot|_{BV(\Omega)}}$, penalizes deviations from piecewise constant functions and does not affect the position of image edges. However, the Bregman iteration facilitates contrast improving intensity changes and enables improved reconstructions.

The paper is organized as follows. In Chapter 2 we will recall a mathematical model for inverse problems with Poisson noise. Starting from a statistical view of the image reconstruction in form of a maximum a-posteriori probability (MAP) estimation based on works of [16], we will proceed to a continuous representation in terms of multidimensional variational problems. An important point in this context is the realization of a-priori knowledge via regularization functionals. As a simple special case, we will derive the well known EM or Richardson-Lucy algorithm with positivity constraints. In Chapter 3, the EM algorithm will be combined with total variation regularization. We will deduce a robust FB-EM-TV algorithm, realized as a two step iteration scheme, and we will provide suitable stopping criterions. The method will be reinterpreted as a modified forward-backward splitting algorithm known from convex optimization. Subsequently, we will present the Bregman-FB-EM-TV method, whereby a loss of contrast in FB-EM-TV will be enhanced by iterative regularization with Bregman distances. In Chapter 4 we will study the FB-EM-TV model from an analytical point of view. After proving the well-posedness of the minimization problem

in terms of existence, uniqueness and stability, we will provide a convergence analysis and positivity preserving properties of the proposed FB-EM-TV algorithm. Thus, we will offer damping conditions to guarantee convergence of the forward-backward splitting algorithm. The numerical realization of ROF related problems, appearing in the second half step of our splitting strategy, is studied in Chapter 5. In the last chapter the performance of our techniques will be illustrated by synthetical and real 2D and 3D reconstructions in high-resolution fluorescence microscopy and positron emission tomography in medical imaging.

2. Mathematical modeling and EM algorithm

2.1. Mathematical model for data acquisition

This section provides an overview of mathematical modeling essential for a reasonable formulation of inverse problems with Poisson noise. In the following we will just concentrate on the relevant aspects of the model construction and refer to the work of Bertero et al. [16] for a detailed discussion. An imaging system consists in general of two structural elements:

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

Hence, we assume 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. So, we consider the detected value f_i as a realization of a random variable F_i .

Additionally, a modeling of the imaging apparatus is necessary 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 concentrate on problems where the transformation is a linear operator and the data acquisition process can be described by a linear operator equation of the form

$$\bar{f} = \bar{K}\bar{u} . \tag{8}$$

Here, $\bar{K} : L^1(\Omega) \rightarrow L^1(\Sigma)$ is a linear and compact operator (thus with a nonclosed range) that additionally preserves positivity. A typical example of (8) is a Fredholm integral equation of the first kind with

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

where \bar{k} is a nonnegative kernel of the operator \bar{K} . In (8), the function \bar{u} describes the desired exact properties of the object and \bar{f} denotes the exact signals before detection. Problem statements of the type above can be found in numerous real-life applications, such as positron emission tomography (PET) in medical imaging [62, 76, 83], fluorescence microscopy [46, 29], astronomy [16] or radar imaging [47, 62].

The modeling of the data acquisition, in the manner as described above, transfers the problem of object reconstruction into the solution of a linear inverse problem of the form (8). 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, such that only an approximate solution u of \bar{u} can be computed from the equation

$$f = Ku . \tag{9}$$

The operator $K : L^1(\Omega) \rightarrow \mathbb{R}^N$ here is a semi-discrete operator based on \bar{K} , which, in contrary to \bar{K} , transforms the desired properties u to the discrete raw data. To complete the modeling of the image reconstruction problem, we additionally have to take the noise in the measurements f into account. Thus we need a model for the probability density of the noise. In this work we concentrate on a specific non-Gaussian noise, namely the so-called Poisson noise. This type of noise appears for example in PET due to radioactive decay of tracers and counting of photon coincidences [80, 83], or in optical nanoscopy due to photon counts by laser sampling of an object [61, 29]. In such cases, every F_i corresponds to a Poisson random variable with an expectation value given by $(Ku)_i$, i.e.

$$F_i \text{ is Poisson distributed with parameter } (Ku)_i . \tag{10}$$

In the following, $p(f; u)$ denotes the conditional probability density of data f for a given image u . Additionally, we make the assumption that the random variables F_i are pairwise independent and identically distributed, i.e. we can write

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

This assumption is in general reasonable since each random variable can be assigned to a specific detector element. Combined with (10), this property leads to the following conditional probability density

$$p(f; u) = \prod_{i=1}^N \frac{(Ku)_i^{f_i}}{f_i!} e^{-(Ku)_i} . \tag{11}$$

Hence, a complete model for the process of data generation and data acquisition is available, if the operator K and the conditional probability density $p(f; u)$ are known.

2.2. Statistical problem formulation of image reconstruction

Due to the compactness of the operator \bar{K} is (8) an ill-posed problem [33, 43]. Note that the problem (9) is not ill-posed in a strong sense, because the operator K has a finite range. Nonetheless, the problem is highly ill-conditioned since K approximates \bar{K} , hence still 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

$$\frac{1}{s} \|Ku - f\|_{L^s(\Sigma)}^s + \alpha R(u) , \quad \alpha > 0 , \quad s \in [1, \infty) . \tag{12}$$

The first term, so-called data fidelity term, penalizes the deviation from equality in (9). The second term is a regularization functional, typically convex, which introduces a-priori information about the expected solution. 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. However, from the viewpoint of statistical modeling, the functionals in (12) are inappropriate for problems with Poisson distributed data, since they result from the assumption of exponential distributed raw data $f = K\bar{u} + \eta$, where η is a vector valued random variable with statistically independent and identically distributed components. Typical examples are that η is Laplace distributed ($s = 1$) or Gaussian distributed ($s = 2$) [16].

In the following, we provide a statistical problem formulation of image reconstruction in the case of Poisson noisy raw data. A commonly used idea to realize regularization techniques with statistical motivation is the Bayesian model. For this purpose, we assume that the desired solution u is a realization of a random variable U itself. The density $p(f; u)$ can be interpreted as a conditional probability of f , if the random variable U takes the value u , i.e. we can write $p(f; u) = p(f|U = u)$ and simply use $p(f|u)$. With this notation, the Bayes formula yields the a-posteriori probability density of u for a given value f of F ,

$$p(u|f) = \frac{p(f|u)p(u)}{p(f)}. \tag{13}$$

Inserting the given measurements f , the density $p(u|f)$ is denoted as the a-posteriori likelihood function, which depends on u only. Now, to determine an approximation to the unknown object \bar{u} , we use the maximum a-posteriori probability (MAP) estimator which maximizes the likelihood function, i.e.

$$u_{MAP} \in \underset{\substack{u \in L^1(\Omega) \\ u \geq 0 \text{ a.e.}}}{\text{arg max}} p(u|f). \tag{14}$$

The positivity constraint on the solution is needed, since in typical applications the functions represent densities or intensity information.

The main advantage of the Bayesian approach (13) is the possibility to incorporate additional prior information into the reconstruction process. Via the a-priori probability density $p(u)$ statistical properties of the desired object u can be embedded. The most frequently used a-priori densities are Gibbs functions [39, 40], in analogy to statistical mechanics,

$$p(u) \sim e^{-\alpha R(u)}, \tag{15}$$

where α is a positive parameter and $R : L^1(\Omega) \rightarrow \mathbb{R} \cup \{+\infty\}$ a convex energy functional.

For a detailed specification of the likelihood function in (14), we proceed with the assumption that a model for the process of data acquisition and data generation, in the manner of the last section, is available. For this reason, we plug the probability density for Poisson noise (11) and the Gibbs a-priori density (15) in the definition of the likelihood function (13) and obtain the following negative log-likelihood function

$$-\log p(u|f) = \sum_{i=1}^N \left((Ku)_i - f_i \log(Ku)_i \right) + \alpha R(u) \tag{16}$$

in which the additive terms independent of u are neglected. At this point we will pass over from a discrete to a continuous representation of data, which corresponds to the way events on detectors are measured. Hence, we assume that any element g in the discrete data space \mathbb{R}^N can be interpreted as a sampling of a function in $L^1(\Sigma)$, which we denote for the sake of convenience with g again. Then, with the indicator function

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

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 we can rewrite the MAP estimate in (14) as the following continuous variational problem,

$$u_{MAP} \in \underset{\substack{u \in L^1(\Omega) \\ u \geq 0 \text{ a.e.}}}{\arg \min} \int_{\Sigma} (Ku - f \log Ku) \, d\mu + \alpha R(u) \quad (17)$$

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

A particular complication of (17) compared to (12) is the strong nonlinearity in the data fidelity term and resulting issues in the computation of minimizers. Finally, with respect to problem (17), we point out that the functional R in the Gibbs a-priori density (15) is related to a regularization functional in the context of inverse problems (cf. (12)). Due to the problem formulation of the image reconstruction via Bayes' theorem one refers to Bayesian regularization in this context.

2.3. EM algorithm

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 leading to a constrained minimization problem (17). In this section we will give a review on a popular reconstruction algorithm for this problem, the so-called expectation-maximization (EM) algorithm [28, 62, 76], 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 [69, 55]. The EM algorithm is an iterative procedure to maximize the likelihood function $p(u|f)$ in problems with incomplete data and will form a basis for our algorithms introduced later. Here, we disregard the prior knowledge first of all and assume that any object u has the same relevance, i.e. the Gibbs a-priori density $p(u)$ in (15) is constant. For simplicity we normalize $p(u)$ such that $R \equiv 0$. Hence, the problem in (17) reduces to the following variational problem with a positivity constraint,

$$\min_{\substack{u \in L^1(\Omega) \\ u \geq 0 \text{ a.e.}}} \int_{\Sigma} (Ku - f \log Ku) \, d\mu . \quad (18)$$

To derive the algorithm, we consider the first order optimality condition of the constrained minimization problem (18). Formally, the Karush-Kuhn-Tucker (KKT) conditions [48, Thm. 2.1.4] provide the existence of a Lagrange multiplier $\lambda \geq 0$, such that the stationary points of the functional in (18) need to fulfill the equations

$$\begin{aligned} 0 &= K^* \mathbf{1}_{\Sigma} - K^* \left(\frac{f}{Ku} \right) - \lambda , \\ 0 &= \lambda u , \end{aligned} \quad (19)$$

where K^* is the adjoint operator of K and $\mathbf{1}_{\Sigma} \in (L^1(\Sigma))^* = L^\infty(\Sigma)$ is the characteristic function on Σ . Since the optimization problem (18) is convex (cf. the comments on equation (22)), every function u fulfilling the equations (19) is a global

minimum of (18). Multiplying the first equation in (19) by u , the Lagrange multiplier λ can be eliminated by the second equation and division by $K^*\mathbf{1}_\Sigma$ leads to a simple iteration scheme,

$$u_{k+1} = \frac{u_k}{K^*\mathbf{1}_\Sigma} K^* \left(\frac{f}{Ku_k} \right), \quad (20)$$

which preserves positivity if the operator K preserves positivity and the initialization u_0 is positive. This iteration scheme is the well known EM algorithm, respectively Richardson-Lucy algorithm. In [76], Shepp and Vardi showed that this iteration is a closed example of the EM algorithm proposed by Dempster, Laird and Rubin in [28], 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 (14), i.e. the solution of (18), can be found in literature [62, 68, 80, 49]. Besides, it is known that the speed of convergence of iteration (20) 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 and 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 $R \equiv 0$.

As described in [68], the EM iterates show the following typical behavior for ill-posed problems. 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 controlled by using appropriate stopping rules to obtain reasonable results. In [68], it is shown that certain stopping rules indeed allow stable approximations. Another possibility to considerably improve reconstruction results are regularization techniques. In the following chapter we will discuss techniques with edge preserving properties.

3. TV regularization methods with Poisson noise

3.1. FB-EM-TV algorithm

The EM or Richardson-Lucy algorithm, discussed in the previous section, is currently one of the standard iterative reconstruction method for most inverse problems with incomplete Poisson data based on the linear equation (9). However, by setting $R \equiv 0$, 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 SNR, like lower tracer dose rate or tracer with short radioactive half life in case of PET examinations, the multiplicative fixed point iteration (20) provides 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 strategy to improve the reconstructions from the EM algorithm is to use a regularized approach. In the classical EM algorithm, the negative log-

likelihood functional (18) is minimized. In the regularized EM approach, we modify the functional by adding a weighted total variation (TV) term [70],

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0 \text{ a.e.}}} \int_{\Sigma} (Ku - f \log Ku) \, d\mu + \alpha |u|_{BV(\Omega)} \, , \quad \alpha > 0 \, . \quad (21)$$

This variational problem is exactly (17) with TV regularization, i.e. from the statistical point of view in Section 2.2 we use the a-priori probability density $p(u)$ with $R(u) = |u|_{BV(\Omega)}$ in (15). This means that images with smaller total variation (higher prior probability) are preferred in the minimization (21). 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. Obviously, such an approach cannot be used for studying certain properties inside the structures in an object (which is anyway unrealistic in case of low SNR), but it is well suited for segmenting different structures and analyzing them quantitatively.

TV regularization is a popular and important concept in several fields of mathematical image processing. It has been derived as a denoising technique in [70] and generalized to various other imaging tasks subsequently. The space of functions with bounded total variation is denoted by $BV(\Omega)$ and is a Banach space equipped with the norm

$$\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + |u|_{BV(\Omega)} \, .$$

$BV(\Omega)$ is a popular function space in image processing since it can represent discontinuous functions (related to the edges in an image) which are even preferred during the minimization of TV. For further properties and details on functions with bounded variation we refer to [1, 36, 41].

For designing the proposed alternating algorithm, we consider the first order optimality condition of the minimization problem (21). Due to the total variation, this variational problem is not differentiable in the usual sense. However, we can extend the data fidelity term to a convex functional without changing the stationary points, namely

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0 \text{ a.e.}}} D_{KL}(f, Ku) + \alpha |u|_{BV(\Omega)} \quad (22)$$

with the Kullback-Leibler (KL) functional D_{KL} [67] defined by

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

such that the minimization problem (22) becomes convex, see Chapter 4. For such problems powerful methods from convex analysis are available and we consider for a convex functional $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$, X Banach space, a generalized derivative called the subdifferential [32] at a point u defined by

$$\partial J(u) := \{ p \in X^* : J(\tilde{u}) - J(u) - \langle p, \tilde{u} - u \rangle \geq 0 \, , \quad \forall \tilde{u} \in X \} \, , \quad (23)$$

where X^* denotes the dual space of X . The single elements of $\partial J(u)$ are called subgradients of J at u .

For the use of subdifferential calculus on the functional in (22), note that due to the definition of the functional $D_{KL}(f, K \cdot)$ on $L^1(\Omega)$, its subgradients are elements of $L^\infty(\Omega)$. The subgradients of $|\cdot|_{BV(\Omega)}$ are in the larger function space $(BV(\Omega))^*$

though, since $BV(\Omega) \subset L^1(\Omega)$. Nevertheless, we can extend TV to a convex functional on $L^1(\Omega)$ by setting

$$|u|_{BV(\Omega)} := \infty \quad \text{if} \quad u \in L^1(\Omega) \setminus BV(\Omega),$$

such that during the minimization process in (22) solutions from the smaller space $BV(\Omega)$ will still be preferred. Hence, due to the continuity of the KL functional and [32, p. 26, Prop. 5.6], we obtain the identity

$$\partial \left(D_{KL}(f, Ku) + \alpha |u|_{BV(\Omega)} \right) = \partial_u D_{KL}(f, Ku) + \alpha \partial |u|_{BV(\Omega)}$$

in $L^\infty(\Omega) \subset (BV(\Omega))^*$ for any $f \in L^1(\Omega)$. Finally, since the subdifferential ∂_u of the KL functional D_{KL} are singletons, the first optimality condition of (22) for a positive solution u is given by

$$K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{Ku} \right) + \alpha p = 0, \quad p \in \partial |u|_{BV(\Omega)}, \quad (24)$$

where K^* denotes the adjoint of K . Formally, this condition is a nonlinear integrodifferential equation

$$K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{Ku} \right) - \alpha \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) = 0.$$

The simplest iteration scheme to compute a solution of the variational problem (21), respectively (22), is a gradient-type method, which however is not robust in case of TV and severe step size restrictions are needed since the subgradient p of TV is treated explicitly. A better idea is to use an iteration scheme which evaluates the nonlocal term (including the operator K) in (24) at the previous iterate u_k and the local term (including the subgradient of TV) at the new iterate u_{k+1} , i.e.

$$\mathbf{1}_\Omega - \frac{1}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{Ku_k} \right) + \alpha \frac{1}{K^* \mathbf{1}_\Sigma} p_{k+1} = 0, \quad p_{k+1} \in \partial |u_{k+1}|_{BV(\Omega)}, \quad (25)$$

with an additional division of (24) by $K^* \mathbf{1}_\Sigma$. In this iteration, the new iterate u_{k+1} appears only as a point of reference for the subdifferential of $|\cdot|_{BV(\Omega)}$. This is a considerable drawback since u_{k+1} cannot be determined from (25) due to the missing of an one-to-one relation between subgradients (dual variables) and primal variable u . In addition, such iteration schemes cannot guarantee preservation of positivity. Hence, we obtain an improved method if we approximate also the constant term $\mathbf{1}_\Omega$ in (25) by $\frac{u_{k+1}}{u_k}$ such that u_{k+1} appears directly, i.e.

$$u_{k+1} - \frac{u_k}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{Ku_k} \right) + \alpha \frac{u_k}{K^* \mathbf{1}_\Sigma} p_{k+1} = 0, \quad p_{k+1} \in \partial |u_{k+1}|_{BV(\Omega)}. \quad (26)$$

In order to verify that the iteration scheme (26) actually preserves positivity, we proceed analogous to the EM algorithm in Section 2.3. Due to the nonnegativity constraint in (21), the KKT conditions formally provide the existence of a Lagrange multiplier $\lambda \geq 0$, such that the stationary points of (22) need to fulfill

$$\begin{aligned} 0 &\in K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{Ku} \right) + \alpha \partial |u|_{BV(\Omega)} - \lambda, \\ 0 &= \lambda u. \end{aligned} \quad (27)$$

By multiplying the first equation in (27) 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

$$u - \frac{u}{K^*\mathbf{1}_\Sigma} K^* \left(\frac{f}{Ku} \right) + \alpha \frac{u}{K^*\mathbf{1}_\Sigma} p = 0, \quad p \in \partial |u|_{BV(\Omega)}, \quad (28)$$

which is just the optimality condition (24) multiplied by u , i.e. this multiplication corresponds to the nonnegativity constraint in (21). Now, the iteration (26) is just a semi-implicit approach to the fixed point equation (28). In Section 4.3, we will prove that this iteration method 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 (26) is just a single EM step in (20). Consequently, the method (26) solving the variational problem (21) can be realized as a nested two step iteration

$$\begin{aligned} u_{k+\frac{1}{2}} &= \frac{u_k}{K^*\mathbf{1}_\Sigma} K^* \left(\frac{f}{Ku_k} \right), & \text{(EM step)} \\ u_{k+1} &= u_{k+\frac{1}{2}} - \alpha \frac{u_k}{K^*\mathbf{1}_\Sigma} p_{k+1}, \quad p_{k+1} \in \partial |u_{k+1}|_{BV(\Omega)}. & \text{(TV step)} \end{aligned} \quad (29)$$

Thus, we alternate an EM reconstruction step with a TV correction step to compute a solution of (21). In Section 3.3 we will see that this iteration scheme can be interpreted as a modified forward-backward (FB) splitting strategy and thus denote it as FB-EM-TV algorithm. The complex second half step from $u_{k+\frac{1}{2}}$ to u_{k+1} in (29) can be realized by solving the convex variational problem

$$u_{k+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{K^*\mathbf{1}_\Sigma (u - u_{k+\frac{1}{2}})^2}{u_k} + \alpha |u|_{BV(\Omega)} \right\}. \quad (30)$$

Inspecting the first order optimality condition confirms the equivalence of this minimization with the TV correction step in (29). It is interesting to note that the problem (30) is just a modified version of the Rudin-Osher-Fatemi (ROF) model, with the difference of a weight $\frac{K^*\mathbf{1}_\Sigma}{u_k}$ in the fidelity term. This analogy creates the opportunity to carry over efficient numerical schemes known for the ROF model. In the numerical realization in Chapter 5, we offer an algorithm analogous to the projected gradient descent algorithm of Chambolle in [22]. In this way, the weighted ROF problem with the exact definition of TV can be solved by using duality, obtaining an accurate, robust, and efficient algorithm.

3.2. Damped FB-EM-TV algorithm

The alternating structure of the proposed iteration (29) has the particular advantage that we might control the interaction between reconstruction and denoising via a simple adaption of the TV correction step. A possibility is a damped TV correction step, namely

$$u_{k+1} = (1 - \omega_k) u_k + \omega_k u_{k+\frac{1}{2}} - \omega_k \alpha \frac{u_k}{K^*\mathbf{1}_\Sigma} p_{k+1}, \quad \omega_k \in (0, 1], \quad (31)$$

which relates the current EM iterate $u_{k+\frac{1}{2}}$ with the previous TV denoised iterate u_k via a convex combination by using a damping parameter ω_k . The damped half step

(31) can be realized in analogy to (30), namely by minimizing the following variational problem

$$u_{k+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u - \tilde{u}_{k+\frac{1}{2}})^2}{u_k} + \omega_k \alpha |u|_{BV(\Omega)} \right\} \quad (32)$$

with

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

This aspect of damping is not only motivated by numerical results, see Section 5, but also required to attain a monotone descent of the objective functional in (21) and (22) during the minimization process (see Section 4.4). Finally, for $\omega_k = 1$, the iteration (31) simplifies to the original TV denoising step in (29). 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 will prove the convergence of the proposed two step iteration with respect to the damped regularization step (31) in Theorem 4.14. Additionally, we will present an explicit bound on ω_k for the convergence in the special case of denoising problems (see Corollary 4.15), i.e. K being the identity operator.

3.3. (Damped) FB-EM-TV algorithm in context of operator splitting

In the previous Sections 3.1 and 3.2, we introduced the FB-EM-TV reconstruction method as a two step algorithm (29) with an additional damping modification (31). 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. [73, 17, 25, 26]. Most of the papers dealing with convex splitting strategies go back to early works of Douglas and Rachford [30] and other authors in [54] and [79].

The optimality condition (24) of our underlying variational problem (21), respectively (22), can be interpreted as a decomposition problem ($C = A + B$), regarding the convex Kullback-Leibler functional and the convex TV regularization term (respectively their subdifferentials). Hence, we consider the stationary equation

$$0 \in C(u) := \underbrace{K^* \mathbf{1}_{\Sigma} - K^* \left(\frac{f}{Ku} \right)}_{=: A(u)} + \underbrace{\alpha \partial |u|_{BV(\Omega)}}_{=: B(u)}, \quad (33)$$

with two maximal monotone operators A and B . Hence, the damped two step iteration (29) with the modified regularization step (31) and $\omega_k \in (0, 1]$ reads as follows,

$$\begin{aligned} \frac{K^* \mathbf{1}_{\Sigma} (u_{k+\frac{1}{2}} - u_k)}{u_k} + A u_k &= 0 \\ \frac{K^* \mathbf{1}_{\Sigma} (u_{k+1} - \omega_k u_{k+\frac{1}{2}} - (1 - \omega_k) u_k)}{u_k} + \omega_k B u_{k+1} &= 0 \end{aligned} \quad (34)$$

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

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

with

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

Compared to the undamped FB-EM-TV strategy (29), in the case of the damped iteration scheme, 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

$$\begin{aligned} u_{k+1} &= \left(I + \frac{\omega_k u_k}{K^* \mathbf{1}_\Sigma} B \right)^{-1} \left(I - \frac{\omega_k u_k}{K^* \mathbf{1}_\Sigma} A \right) u_k \\ &= (L_k + B)^{-1} (L_k - A) u_k \end{aligned} \tag{35}$$

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 [54] and Passty [65]. In our case, we will see in Theorem 4.14 that the key to proving the convergence of the FB-EM-TV 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. [54] or [74], 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. [37] or [75]. There, the authors use an augmented Lagrangian approach and a split Bregman method, which are equivalent [35] and correspond to the Douglas-Rachford splitting strategy applied to the dual problem of (21) [74]. 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.

3.4. Stopping rules and pseudocode for the (damped) FB-EM-TV algorithm

Moreover, we need to define appropriate stopping rules in order to guarantee the accuracy of the proposed (damped) FB-EM-TV algorithm. In addition to a maximal number of iterations, the error in the optimality condition (24) 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,

$$\langle u, v \rangle_w := \int_{\Omega} u v w \, d\lambda \quad \text{and} \quad \|u\|_{2,w} := \sqrt{\langle u, u \rangle_w} , \tag{36}$$

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

$$\text{opt}_{k+1} := \left\| K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{K u_{k+1}} \right) + \alpha p_{k+1} \right\|_{2, u_{k+1}}^2 . \tag{37}$$

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 (24), but also in the convergence of the sequence of primal functions $\{u_k\}$ and the sequence of subgradients $\{p_k\}$ with $p_k \in \partial |u_k|_{BV(\Omega)}$. Hence, in order to establish appropriate stopping rules

for these iterates, we consider the damped TV correction step (31) with the EM reconstruction step in (29),

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

By combining this iteration scheme with the optimality condition (28) evaluated at u_k , which must be fulfilled in the case of convergence, we obtain an optimality statement for the sequences $\{p_k\}$ and $\{u_k\}$,

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

With the aid of the weighted norm (36), we now have additional stopping criteria for the FB-EM-TV algorithm, which guarantee the accuracy of the primal functions $\{u_k\}$ and the subgradients $\{p_k\}$, namely

$$\begin{aligned} u_{opt_{k+1}} &:= \left\| \frac{K^* \mathbf{1}_\Sigma (u_{k+1} - u_k)}{\omega_k u_k} \right\|_{2, u_{k+1}}^2, \\ p_{opt_{k+1}} &:= \|\alpha (p_{k+1} - p_k)\|_{2, u_{k+1}}^2. \end{aligned} \tag{38}$$

We finally mention that the stopping criteria (37) and (38) are well defined, since we can prove that each iterate u_k of the damped FB-EM-TV splitting strategy is strictly positive, see Lemma 4.12.

Based on the observations in previous sections we can use Algorithm 1 to solve the TV regularized Poisson likelihood estimation problem (21).

Algorithm 1 (Damped) FB-EM-TV 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** ▷ (37), (38)
 - i) Compute $u_{k+\frac{1}{2}}$ via EM step in (29).
 - ii) Set $\omega_k = \omega$.
 - iii) Compute u_{k+1} via modified ROF model (32). ▷ Section 5
 - iv) $k \leftarrow k + 1$
 - end while**
 - return** u_k
-

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, see e.g. [53, 78, 82]. Most of them are based on the discrepancy principle and Chi-square distributions, generalized cross validation methods or unbiased predictive risk estimates. 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. [9] and the references therein or [15].

3.5. Bregman-FB-EM-TV algorithm

The presented FB-EM-TV algorithm (29) solves the estimation problem (21) and provides cartoon-like reconstructions with sharp edges. However, the realization of TV correction steps via the weighted ROF model (30) has the drawback that the reconstructed images suffer from contrast reduction [58, 63]. Thus, we propose to extend the TV regularized Poisson likelihood estimation problem (21), and therewith also the (damped) FB-EM-TV, by an iterative regularization strategy to a simultaneous contrast correction. More precisely, we perform a contrast enhancement by inverse scale space methods based on Bregman distance iteration. These techniques have been derived by Osher et al. in [63], with a detailed analysis for Gaussian-type problems (4), and have been generalized to time-continuity [21], L^p -norm data fitting terms [20] and nonlinear inverse problems [5]. Following these methods, an iterative refinement in our case is realized by a sequence of modified Poisson likelihood estimation problems based on (21).

The inverse scale space methods concerning TV, derived in [63], follow the concept of iterative regularization by Bregman distance [18]. In the case of the Poisson model, the method initially starts with a simple FB-EM-TV algorithm, i.e. it consists in computing a minimizer u^1 of (21), respectively (22). Then, updates are determined subsequently by considering variational problems with a shifted TV term, namely

$$u^{l+1} \in \arg \min_{\substack{u \in BV(\Omega) \\ u \geq 0 \text{ a.e.}}} \left\{ D_{KL}(f, Ku) + \alpha \left(|u|_{BV(\Omega)} - \langle p^l, u \rangle \right) \right\}, \quad p^l \in \partial |u^l|_{BV(\Omega)}. \quad (39)$$

The mentioned Bregman distance with respect to $|\cdot|_{BV(\Omega)}$ is defined via

$$D_{|\cdot|_{BV(\Omega)}}^{\tilde{p}}(u, \tilde{u}) := |u|_{BV(\Omega)} - |\tilde{u}|_{BV(\Omega)} - \langle \tilde{p}, u - \tilde{u} \rangle, \quad \tilde{p} \in \partial |\tilde{u}|_{BV(\Omega)},$$

where $\langle \cdot, \cdot \rangle$ denotes the standard duality product. The introduction of this definition allows for characterizing the sequence of modified variational problems (39) by adding of constant terms as

$$u^{l+1} \in \arg \min_{\substack{u \in BV(\Omega) \\ u \geq 0 \text{ a.e.}}} \left\{ D_{KL}(f, Ku) + \alpha D_{|\cdot|_{BV(\Omega)}}^{p^l}(u, u^l) \right\}, \quad p^l \in \partial |u^l|_{BV(\Omega)}. \quad (40)$$

Note that the first iterate u^1 can also be realized by the variational problem (39) or (40), if the function u^0 will be chosen constant and we set $p^0 := 0 \in \partial |u^0|_{BV(\Omega)}$.

The Bregman distance $D_{|\cdot|_{BV(\Omega)}}^{\tilde{p}}$ does not represent a distance in the common (metric) sense, since it is not symmetric in general and the triangle inequality does not hold. However, the formulation in (40) has the advantage over (39), that $D_{|\cdot|_{BV(\Omega)}}^{\tilde{p}}$ is a distance measure in the following sense

$$D_{|\cdot|_{BV(\Omega)}}^{\tilde{p}}(u, \tilde{u}) \geq 0 \quad \text{and} \quad D_{|\cdot|_{BV(\Omega)}}^{\tilde{p}}(u, \tilde{u}) = 0 \quad \text{for} \quad u = \tilde{u}.$$

Besides, the Bregman distance is convex in the first argument because $|\cdot|_{BV(\Omega)}$ is convex. In general, i.e. for any convex functional J the Bregman distance can be interpreted as the difference between J in u and the Taylor linearization of J around \tilde{u} if, in addition, J is continuously differentiable (see e.g. [20]).

From the statistical point of view in Section 2.2, the Bregman distance regularized variational problem (40) uses an adapted a-priori probability density $p(u)$ (15) in the Bayesian model formulation (13). Namely, instead of a zero centered a-priori probability $R(u) = |u|_{BV(\Omega)}$ as in the case of the FB-EM-TV algorithm, here we

consider in every Bregman refinement step a new a-priori probability which is related to a shifted total variation, i.e we use the following Gibbs function (15)

$$p(u) \sim e^{-\alpha D_{|\cdot|_{BV(\Omega)}}^l(u, u^l)}.$$

This means that images with smaller total variation and a close distance to the maximum likelihood estimator u^l of the previous FB-EM-TV problem are preferred in the minimization of (40).

Before deriving a two step iteration corresponding to (29), we will motivate the contrast enhancement by iterative regularization in (40). The regularization with TV in (22) prefers functions with only few oscillations, ideally piecewise constant functions. As a consequence, the reconstruction results suffer from a loss of contrast. The iterative Bregman regularization has the advantage that, with u^l as an approximation to the possible solution, additional information is available. The variational problem (40) can be interpreted as follows: search for a solution that matches the Poisson distributed data after applying K and simultaneous minimizes the residual of the Taylor approximation of $|\cdot|_{BV(\Omega)}$ around u^l . This form of regularization hardly changes the position of gradients with respect to the last iterate u^l , but that an increase of intensities is permitted. This leads to a noticeable contrast enhancement.

For designing a two step iteration analogous to the FB-EM-TV algorithm, we consider the first order optimality condition for the variational problem (40). Due to the convexity of the Bregman distance in the first argument, we can determine the subdifferential of (40). Analogous to the derivation of the FB-EM-TV iteration, we can split the subdifferential of a sum of functions to a sum of subdifferentials due to the continuity of the Kullback-Leibler functional. Now, since the subdifferential of the KL functional can be expressed formally by Fréchet derivatives analogous to (24) and

$$\partial(-\langle p^l, u \rangle) = \{-p^l\}$$

holds, the first order optimality condition of (40) for a positive solution u^{l+1} is given by

$$0 \in K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{K u^{l+1}} \right) + \alpha \left(\partial |u^{l+1}|_{BV(\Omega)} - p^l \right), \quad p^l \in \partial |u^l|_{BV(\Omega)}. \quad (41)$$

For u^0 constant and $p^0 := 0 \in \partial |u^0|_{BV(\Omega)}$, this condition yields a well defined update formula of the iterates p^l , namely

$$p^{l+1} := p^l - \frac{1}{\alpha} \left(K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{K u^{l+1}} \right) \right) \in \partial |u^{l+1}|_{BV(\Omega)}. \quad (42)$$

Now, analogous to the FB-EM-TV algorithm, we can apply the idea of the nested two step iteration (29) in every refinement step, $l = 0, 1, \dots$. Thus, for the solution of (40), the condition (41) yields a strategy consisting of an EM reconstruction step

$$u_{k+\frac{1}{2}}^{l+1} = \frac{u_k^{l+1}}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u_k^{l+1}} \right), \quad (43)$$

followed by solving the adapted weighted ROF problem

$$u_{k+1}^{l+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{K^* \mathbf{1}_\Sigma (u - u_{k+\frac{1}{2}}^{l+1})^2}{u_k^{l+1}} + \alpha (J(u) - \langle p^l, u \rangle) \right\}. \quad (44)$$

Following [63], we provide an opportunity to transfer the shift term $\langle p^l, u \rangle$ to the data fidelity term. This approach facilitates the implementation of contrast enhancement with the Bregman distance via a slight modification of the FB-EM-TV algorithm. With the scaling $K^* \mathbf{1}_\Sigma v^l := \alpha p^l$ and (41) we obtain the following update formula for the iterates v^l ,

$$v^{l+1} = v^l - \left(\mathbf{1}_\Omega - \frac{1}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u^{l+1}} \right) \right), \quad v^0 = 0. \quad (45)$$

Using this scaled update, we can rewrite the second step (44) to

$$u_{k+1}^{l+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{K^* \mathbf{1}_\Sigma \left((u - u_{k+\frac{1}{2}}^{l+1})^2 - 2u u_k^{l+1} v^l \right)}{u_k^{l+1}} + \alpha |u|_{BV(\Omega)} \right\}.$$

Note that in the equation

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

the last two terms on the right-hand side are independent of u and hence the variational problem (44) simplifies to

$$u_{k+1}^{l+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{K^* \mathbf{1}_\Sigma \left(u - (u_{k+\frac{1}{2}}^{l+1} + u_k^{l+1} v^l) \right)^2}{u_k^{l+1}} + \alpha |u|_{BV(\Omega)} \right\}, \quad (46)$$

i.e. the second step (44) can be realized by a minor modification of the TV regularization step introduced in (30).

In Section 3.2 we additionally introduced a damped variant of the FB-EM-TV algorithm. This damping strategy can be also realized in each Bregman refinement step, namely the TV regularization step (46) simply needs to be adapted to

$$u_{k+1}^{l+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{K^* \mathbf{1}_\Sigma \left(u - \tilde{u}_{k+\frac{1}{2}}^{l+1} \right)^2}{u_k^{l+1}} + \omega_k^{l+1} \alpha |u|_{BV(\Omega)} \right\} \quad (47)$$

with

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

3.6. Stopping rules and pseudocode for the Bregman-FB-EM-TV algorithm

As usual for iterative methods, the described refinement strategy via the iterative Bregman distance regularization 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 by the inverse scale space strategy. In the case of Gaussian noise, the authors in [63] and [21] suggested to use the so-called generalized discrepancy principle (cf. [33] and [66] 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.

$$l_* = \max\{ l \in \mathbb{N} : \|K u^l - f\|_{L^2(\Sigma)} \geq \tau \delta \}, \quad \tau > 1.$$

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-TV iteration loop. For this purpose, we can proceed analogously to the discussion of stopping rules in the case of the FB-EM-TV algorithm in Section 3.4. In addition to a maximum number of iterations, we consider the error in the optimality condition (41) as a basic stopping criterion using the weighted norm $\|\cdot\|_{2,w}$ (36), i.e.

$$\begin{aligned} opt_{k+1}^{l+1} &:= \left\| K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{Ku_{k+1}^{l+1}} \right) + \alpha p_{k+1}^{l+1} - \alpha p^l \right\|_{2, u_{k+1}^{l+1}}^2 \\ &= \left\| K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{Ku_{k+1}^{l+1}} \right) + \alpha p_{k+1}^{l+1} - K^* \mathbf{1}_\Sigma v^l \right\|_{2, u_{k+1}^{l+1}}^2. \end{aligned} \quad (48)$$

Analogous to (38), we also obtain additional stopping criteria to guarantee the accuracy of the primal function sequence $\{u_k^{l+1}\}$ and the subgradient sequence $\{p_k^{l+1}\}$ with $p_k^{l+1} \in \partial |u_k^{l+1}|_{BV(\Omega)}$, namely

$$\begin{aligned} u_{opt_{k+1}^{l+1}} &:= \left\| \frac{K^* \mathbf{1}_\Sigma (u_{k+1}^{l+1} - u_k^{l+1})}{\omega_k^{l+1} u_k^{l+1}} \right\|_{2, u_{k+1}^{l+1}}^2, \\ p_{opt_{k+1}^{l+1}} &:= \left\| \alpha (p_{k+1}^{l+1} - p_k^{l+1}) \right\|_{2, u_{k+1}^{l+1}}^2. \end{aligned} \quad (49)$$

Summarizing the observations above, we can make use of Algorithm 2 to solve the stepwise refinement (39) of the TV regularized Poisson likelihood estimation problem (21).

Remark.

- Note that the update variable v in (45) has an interesting interpretation as an error function with reference to the optimality condition of the unregularized Poisson log-likelihood functional (18). 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 (18),

$$K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{Ku} \right) = 0.$$

Hence, caused by the TV regularization step (46), we can expect that the iterative regularization based on the Bregman distance leads to a stepwise contrast enhancement. 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-TV step (30), the Bregman refinement strategy (46) uses an adapted "noisy" function in the data fidelity term, where the intensities of

Algorithm 2 Bregman-FB-EM-TV Algorithm

1. **Parameters:** $f, \alpha > 0, \omega \in (0,1], \maxBregIts \in \mathbb{N}, \delta > 0, \tau > 1,$
 $\maxEMIts > \mathbb{N}, \text{tol} > 0$

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**
($opt_k^{l+1} \geq \text{tol}$ **or** $u_{opt_k^{l+1}} \geq \text{tol}$ **or**
 $p_{opt_k^{l+1}} \geq \text{tol}$)) **do** \triangleright (48), (49)

i) Compute $u_{k+\frac{1}{2}}^{l+1}$ via EM step in (43).

ii) Set $\omega_k^{l+1} = \omega$.

iii) Compute u_{k+1}^{l+1} via modified ROF model (47). \triangleright Section 5

iv) $k \leftarrow k + 1$

end while

b) Compute update v^{l+1} via (45).

c) Set $u_0^{l+2} = u_k^{l+1}$.

d) $l \leftarrow l + 1$

end while

return u_0^{l+1}

the EM solution $u_{k+\frac{1}{2}}^{l+1}$ are increased by a weighted error function v^l . Following additionally [58] or [81], the elements of the dual space of $BV(\Omega)$, in our case $p^l = \frac{K^* \mathbf{1}_\Omega}{\alpha} v^l \in \partial |u^{l+1}|_{BV(\Omega)} \subset (BV(\Omega))^*$, can be characterized as textures respectively strongly oscillating patterns. Based on this interpretation, it makes sense to consider v^l as the current error function of the unregularized Poisson log-likelihood functional (18).

- As an alternative to the approach in Section 3.5, a dual inverse scale space strategy based on Bregman distance iterations can be used to obtain simultaneous contrast correction, see [19]. However, both inverse scale space methods compute very similar iterates and we could not observe 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 [19]), which are not possible for the primal approach in Section 3.5 so far.
- Motivated by the work in [63], one may also consider the modeling of an iterative reconstruction refinement inside the FB-EM-TV algorithm. On the basis of the two step iteration proposed in (29) and (30), this strategy would lead to a TV regularization step which can be realized by a sequence of modified ROF problems based on (30). More precisely, for any fixed index k , the iterate u_{k+1} is determined

via a sequence of the following minimization problem,

$$u_{k+1}^{l+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u - u_{k+\frac{1}{2}})^2}{u_k} + \alpha \left(|u|_{BV(\Omega)} - \langle p^l, u \rangle \right) \right\} \quad (50)$$

with $p^l \in \partial |u_{k+1}^l|_{BV(\Omega)}$, a constant initialization u_{k+1}^0 and $p^0 := 0 \in \partial |u_{k+1}^0|_{BV(\Omega)}$. Analogous to Section 3.5, 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 (50) can be rewritten similar to (46), namely

$$u_{k+1}^{l+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} \left(u - (u_{k+\frac{1}{2}} + v^l) \right)^2}{u_k} + \alpha |u|_{BV(\Omega)} \right\}$$

with the update formula

$$v^{l+1} = v^l + (u_{k+\frac{1}{2}} - u_{k+1}^{l+1}), \quad v^0 = 0.$$

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 3.5.

4. Analysis

In this chapter we carry out a mathematical analysis of the TV regularized Poisson based variational model (21). More precisely, we prove that the problem is well-posed, that the FB-EM-TV algorithm preserves the positivity of the solution and that the proposed damped FB-EM-TV iteration scheme has a stable convergence behavior. Note that the same results can also be extended to the general case of regularized Poisson likelihood estimation problem with an arbitrary convex regularization functional, including those which are non differentiable in the classical sense (see [71]).

4.1. Assumptions, definitions and preliminary results

At the beginning of this section, we will repeat the properties of the operator K , having been assumed in the previous sections, and will introduce other necessary assumptions, which will be used in the following analysis.

As introduced in Section 2.1, we consider the forward operator K as a semi discrete operator based on $\bar{K} : L^1(\Omega) \rightarrow L^1(\Sigma)$, which transforms, in contrary to \bar{K} , a function from $L^1(\Omega)$ to the discrete data space \mathbb{R}^N . Nevertheless, to be able to present a unified theory with respect to the continuous problem formulation (21), we passed over from a discrete to a continuous representation of the raw data using a point measure μ in Section 2.2. There, we assumed that any element g in the discrete data space \mathbb{R}^N can be interpreted as a sampling of a function in $L^1_{\mu}(\Sigma)$, which we denote for the sake of convenience with g again and where $L^1_{\mu}(\Sigma)$ denotes the space of Lebesgue measurable 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 Lebesgue measure and we have $L^1_{\mu}(\Sigma) = L^1(\Sigma)$. Finally, for the assumptions

below, note that the operator \bar{K} is linear, compact and additionally obtains positivity. Hence, based on the observations above, we make the following assumptions.

Assumption 4.1.

- (i) The operator $K : L^1(\Omega) \rightarrow L^1_\mu(\Sigma)$ is linear and bounded.
- (ii) 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 $u = 0$.
- (iii) If $u \in L^1(\Omega)$ satisfies $c_1 \leq u \leq c_2$ a.e. for some positive constants $c_1, c_2 > 0$, then there exist $c_3, c_4 > 0$ such that $c_3 \leq Ku \leq c_4$ a.e. on Σ .

Remark. At the first glance, the Assumption 4.1 (iii) is restrictive, but 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 application, e.g. in geophysics and potential theory or in deconvolution problems as fluorescence microscopy or astronomy. Another interesting example of operators, which fulfill the Assumption 4.1 (iii), is the X-ray transform which assigns the integral values along all straight lines to a given function. This transform coincides in two dimensions with the well-known Radon transform and is strongly applied in medical imaging. The Assumption 4.1 (iii) 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.

Next we will give a definition of the Kullback-Leibler functional in order to simplify the following analysis of the regularized Poisson likelihood estimation problem (21).

Definition 4.2 (Kullback-Leibler Functional). *The Kullback-Leibler (KL) functional is a function $D_{KL} : L^1(\Sigma) \times L^1(\Sigma) \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$ with $\Sigma \subset \mathbb{R}^m$ bounded and measurable, given by*

$$D_{KL}(\varphi, \psi) = \int_{\Sigma} \left(\varphi \log \left(\frac{\varphi}{\psi} \right) - \varphi + \psi \right) d\nu \quad \text{for all } \varphi, \psi \geq 0 \text{ a.e.}, \quad (51)$$

where ν is a measure. Note that, using the convention $0 \log 0 = 0$, the integrand in (51) is nonnegative and vanishes if and only if $\varphi = \psi$.

Remark. In the literature there exist further notations for the KL functional, like cross-entropy, information for discrimination or Kullback’s I-divergence, cf. e.g. [27, 31, 67]. The functional (51) 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 (51) is a Bregman distance or divergence with respect to the Boltzmann-Shannon entropy [67].

In the next Lemmas, we will recall a collection of basic results about the KL functional and the total variation $|\cdot|_{BV(\Omega)}$ functional from [67] and [1], which will be used in the following analysis. For further information to the both terms, we refer to [67, 31] and [1, 2, 3, 36, 41].

Lemma 4.3 (Properties of KL Functional). *Let K satisfy Assumption 4.1 (i) and (ii). 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 K , the function $(\varphi, u) \mapsto D_{KL}(\varphi, Ku)$ is also convex.
- (ii) For any fixed nonnegative $\varphi \in L^1(\Sigma)$, the function $u \mapsto D_{KL}(\varphi, Ku)$ is lower semicontinuous with respect to the weak topology of $L^1(\Sigma)$.
- (iii) For any nonnegative function φ and ψ in $L^1(\Sigma)$, one has

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

Proof. See [67, Lemma 3.3 - 3.4]. □

Corollary 4.4. *If $\{\varphi_n\}$ and $\{\psi_n\}$ are bounded sequences in $L^1(\Sigma)$, then*

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

Lemma 4.5 (Properties of TV Functional). *The following statements hold:*

- (i) $|\cdot|_{BV(\Omega)}$ is convex on $BV(\Omega)$.
- (ii) $|\cdot|_{BV(\Omega)}$ is lower semicontinuous with respect to the weak topology of $L^1(\Omega)$.
- (iii) Any uniformly bounded sequence $\{u_n\}$ in $BV(\Omega)$ is relatively compact in $L^1(\Omega)$.

Proof. See [1, Theorem 2.3 - 2.5]. □

4.2. Well-posedness of the minimization problem

In the following we verify existence, uniqueness, and stability of the minimization problem (21). In order to use the known properties of the KL functional from Lemma 4.3 for the analysis of (21), we add the term $f \log f - f$ to the data fidelity term. Because this expression is independent of the desired function u , the stationary points of the minimization problem are not affected (if they exist) and (21) is equivalent to

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0 \text{ a.e.}}} F(u) := D_{KL}(f, Ku) + \alpha |u|_{BV(\Omega)} , \quad \alpha > 0 , \quad (52)$$

where D_{KL} is the Kullback-Leibler functional as in Definition 4.2. For the following analysis, the precompactness result from Lemma 4.5 (iii) is of fundamental importance. Hence, in order to use this property we introduce the following definition.

Definition 4.6 (BV-Coercivity). *A functional F defined on $L^1(\Omega)$ is BV-coercive (cf. [48], Def. IV.3.2.6), if the sub-level sets of F are bounded in the $\|\cdot\|_{BV(\Omega)}$ norm, i.e. for all $r \in \mathbb{R}_{\geq 0}$ the set $\{u \in L^1(\Omega) : F(u) \leq r\}$ is uniformly bounded in the BV norm; or equivalent*

$$F(u) \rightarrow +\infty \quad \text{whenever} \quad \|u\|_{BV(\Omega)} \rightarrow +\infty . \quad (53)$$

Lemma 4.7 (BV-Coercivity of the Minimization Functional). *Let K satisfy Assumption 4.1 (i) and (ii). Moreover, assume that $\alpha > 0$, $f \in L^1_\mu(\Sigma)$ is nonnegative and that the operator K does not annihilate constant functions. Since K is linear, the latter condition is equivalent to*

$$K\mathbf{1}_\Omega \neq 0 , \quad (54)$$

where $\mathbf{1}_\Omega$ denotes the characteristic function on Ω . Then, the functional F defined in (52) is BV-coercive.

Remark. According to the definition of the function space of functions with bounded (total) variation, $BV(\Omega) \subset L^1(\Omega)$ is valid and we can extend the admissible solution set of the minimization problem (52) from $BV(\Omega)$ to $L^1(\Omega)$. To this end, we continue the total variation to a functional on $L^1(\Omega)$ by setting $|u|_{BV(\Omega)} = +\infty$ if $u \in L^1(\Omega) \setminus BV(\Omega)$, where furthermore solutions from $BV(\Omega)$ are preferred during minimization.

Proof of Lemma 4.7. For the proof of BV-coercivity, we derive an estimate of the form

$$\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + |u|_{BV(\Omega)} \leq c_1 (F(u))^2 + c_2 F(u) + c_3, \quad (55)$$

with constants $c_1 \geq 0$, $c_2 > 0$ and $c_3 \geq 0$. Then, the desired coercivity property (53) follows directly from the positivity of the functional F for all $u \in L^1(\Omega)$ 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

$$u = w + v, \quad (56)$$

where

$$w = \left(\frac{\int_{\Omega} u \, dx}{|\Omega|} \right) \mathbf{1}_{\Omega} \quad \text{and} \quad v := u - w \quad \text{with} \quad \int_{\Omega} v \, dx = 0. \quad (57)$$

First, we estimate $|v|_{BV(\Omega)}$ and $\|v\|_{L^1(\Omega)}$. Because constant functions have no variation, the positivity of the KL functional yields

$$\alpha |v|_{BV(\Omega)} \leq \alpha |u|_{BV(\Omega)} \leq F(u) \quad \Rightarrow \quad |v|_{BV(\Omega)} \leq \frac{1}{\alpha} F(u).$$

This together with the Poincaré-Wirtinger inequality (see e.g. [3, Sect. 2.5.1]) yields an estimate of the L^1 norm,

$$\|v\|_{L^1(\Omega)} \leq C_1 |v|_{BV(\Omega)} \leq C_1 \frac{1}{\alpha} F(u), \quad (58)$$

where $C_1 > 0$ is a constant that depends on $\Omega \subset \mathbb{R}^d$ and d only. Now, using the decomposition (56) and the estimates for $|v|_{BV(\Omega)}$ and $\|v\|_{L^1(\Omega)}$, the problem (55) reduces to the estimation of the L^1 norm of constant functions, since

$$\begin{aligned} \|u\|_{BV(\Omega)} &\leq \|w\|_{L^1(\Omega)} + \|v\|_{L^1(\Omega)} + |v|_{BV(\Omega)} \\ &\leq \|w\|_{L^1(\Omega)} + (C_1 + 1) \frac{1}{\alpha} F(u). \end{aligned} \quad (59)$$

To estimate $\|w\|_{L^1(\Omega)}$, we consider the L^1_{μ} distance between $Ku = Kw + Kv$ and f . With Lemma 4.3 (iii) we obtain an upper bound,

$$\begin{aligned} \|(Kv - f) + Kw\|_{L^1_{\mu}(\Sigma)}^2 &\leq \left(\frac{2}{3} \|f\|_{L^1_{\mu}(\Sigma)} + \frac{4}{3} \|Kv + Kw\|_{L^1_{\mu}(\Sigma)} \right) D_{KL}(f, Ku) \\ &\leq \left(\frac{2}{3} \|f\|_{L^1_{\mu}(\Sigma)} + \frac{4}{3} \|Kv\|_{L^1_{\mu}(\Sigma)} + \frac{4}{3} \|Kw\|_{L^1_{\mu}(\Sigma)} \right) F(u), \end{aligned}$$

and as a lower bound we obtain,

$$\begin{aligned} \|(Kv - f) + Kw\|_{L^1_\mu(\Sigma)}^2 &\geq \left(\|Kv - f\|_{L^1_\mu(\Sigma)} - \|Kw\|_{L^1_\mu(\Sigma)} \right)^2 \\ &\geq \|Kw\|_{L^1_\mu(\Sigma)} \left(\|Kw\|_{L^1_\mu(\Sigma)} - 2\|Kv - f\|_{L^1_\mu(\Sigma)} \right). \end{aligned}$$

Combining (58) with both inequalities yields

$$\begin{aligned} \|Kw\|_{L^1_\mu(\Sigma)} &\left(\|Kw\|_{L^1_\mu(\Sigma)} - 2 \left(\|K\| C_1 \frac{1}{\alpha} F(u) + \|f\|_{L^1_\mu(\Sigma)} \right) \right) \\ &\leq \left(\frac{2}{3} \|f\|_{L^1_\mu(\Sigma)} + \frac{4}{3} \|K\| C_1 \frac{1}{\alpha} F(u) + \frac{4}{3} \|Kw\|_{L^1_\mu(\Sigma)} \right) F(u). \end{aligned} \quad (60)$$

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 (54) on K . Thus, there exists a constant $C_2 > 0$ with

$$C_2 = \frac{\int_\Sigma |K \mathbf{1}_\Omega| \, d\mu}{|\Omega|} \quad \text{and} \quad \|Kw\|_{L^1_\mu(\Sigma)} = C_2 \|w\|_{L^1(\Omega)}. \quad (61)$$

This identity used in the inequality (60) yields

$$\begin{aligned} C_2 \|w\|_{L^1(\Omega)} &\left(C_2 \|w\|_{L^1(\Omega)} - 2 \left(\|K\| C_1 \frac{1}{\alpha} F(u) + \|f\|_{L^1_\mu(\Sigma)} \right) - \frac{4}{3} F(u) \right) \\ &\leq \left(\frac{2}{3} \|f\|_{L^1_\mu(\Sigma)} + \frac{4}{3} \|K\| C_1 \frac{1}{\alpha} F(u) \right) F(u). \end{aligned} \quad (62)$$

To receive an estimate of the form (55), we distinguish between two cases:

Case 1: If

$$C_2 \|w\|_{L^1(\Omega)} - 2 \left(\|K\| C_1 \frac{1}{\alpha} F(u) + \|f\|_{L^1_\mu(\Sigma)} \right) - \frac{4}{3} F(u) \geq 1, \quad (63)$$

then we conclude from (62) that

$$\|w\|_{L^1(\Omega)} \leq \frac{1}{C_2} \left(\frac{2}{3} \|f\|_{L^1_\mu(\Sigma)} + \frac{4}{3} \|K\| C_1 \frac{1}{\alpha} F(u) \right) F(u),$$

and obtain with (59),

$$\|u\|_{BV(\Omega)} \leq \frac{4C_1 \|K\|}{3C_2 \alpha} (F(u))^2 + \left(\frac{2}{3C_2} \|f\|_{L^1_\mu(\Sigma)} + \frac{C_1 + 1}{\alpha} \right) F(u). \quad (64)$$

Case 2: If the condition (63) does not hold, i.e.

$$\|w\|_{L^1(\Omega)} < \frac{1}{C_2} \left(1 + 2 \left(\|K\| C_1 \frac{1}{\alpha} F(u) + \|f\|_{L^1_\mu(\Sigma)} \right) + \frac{4}{3} F(u) \right),$$

then we find from (59) that

$$\|u\|_{BV(\Omega)} \leq \left(\frac{2\|K\| C_1 \frac{1}{\alpha} + \frac{4}{3}}{C_2} + \frac{C_1 + 1}{\alpha} \right) F(u) + \frac{1 + 2\|f\|_{L^1_\mu(\Sigma)}}{C_2}. \quad (65)$$

Now, with Assumption 4.1 (i) we have that $\|K\| < \infty$. Moreover, since $f \in L^1_\mu(\Sigma)$, we obtain from (64) and (65) the desired coercivity property (55). \square

Theorem 4.8 (Existence of Minimizers). *Let K satisfy Assumption 4.1 (i) and (ii). Moreover, assume that $\alpha > 0$, $f \in L^1_\mu(\Sigma)$ is nonnegative and that the operator K satisfies (54). Then, the functional F defined in (52) has a minimizer in $BV(\Omega)$.*

Proof. We use the direct method of the calculus of variations, see e.g. [3, Sect. 2.1.2]: Let $\{u_n\} \subset BV(\Omega)$, $u_n \geq 0$ a.e., be a minimizing sequence of the functional F , i.e.

$$\lim_{n \rightarrow \infty} F(u_n) = \inf_{u \in BV(\Omega)} F(u) =: F_{min} < \infty .$$

With the assumptions on the operator K , Lemma 4.7 implies that the functional F is BV-coercive and with it also that all elements of the sequence $\{u_n\}$ are uniformly bounded in the BV norm. Hence, as a consequence of the precompactness result from Lemma 4.5 (iii), there exists a subsequence $\{u_{n_j}\}$ which converges to some $\tilde{u} \in L^1(\Omega)$. Actually, the function \tilde{u} lies in $BV(\Omega)$, since $|\cdot|_{BV(\Omega)}$ is lower semicontinuous (see Lemma 4.5 (ii)) and the sequence $\{u_n\}$ is uniformly bounded in $BV(\Omega)$. Simultaneously, caused by Lemma 4.3 (ii), also the objective functional F is lower semicontinuous and implies the inequality

$$F(\tilde{u}) \leq \liminf_{j \rightarrow \infty} F(u_{n_j}) = F_{min} ,$$

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 straightforward to see that the negative logarithm function is strictly convex and consequently also the function $u \mapsto D_{KL}(f, Ku)$, if $f \in L^1_\mu(\Sigma)$ 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 4.1 (i)). Since the regularization term is assumed convex (see Lemma 4.5) we can immediately conclude the following result.

Theorem 4.9 (Uniqueness of Minimizers). *Let K satisfy Assumption 4.1 (i) and (ii). Assume that K is an injective operator and $f \in L^1_\mu(\Sigma)$ fulfills $\inf_\Sigma f > 0$. Then, the function $u \mapsto D_{KL}(f, Ku)$ and also the functional F from (52) is strictly convex. In particular, the minimizer of F is unique in $BV(\Omega)$.*

After existence and uniqueness of minimizers we will show the stability of the TV regularized Poisson estimation problem (52) with respect to a certain kind of data perturbations in the following. In Section 2.1 we already described that the given measurements in practice are discrete and can be interpreted in our framework as averages of a function $f \in L^1(\Sigma)$. The open question certainly is the suitable choice of the function f . Moreover, the physically limited discrete construction of the detectors 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 (66).

Theorem 4.10 (Stability with Respect to Perturbations in Measurements). *Let K satisfy Assumption 4.1. Fix $\alpha > 0$ and assume that the functions $f_n \in L^1_\mu(\Sigma)$, $n \in \mathbb{N}$, are nonnegative approximations of a data function $f \in L^1_\mu(\Sigma)$ in the form*

$$\lim_{n \rightarrow \infty} D_{KL}(f_n, f) = 0 . \tag{66}$$

Moreover, let

$$u_n \in \underset{\substack{u \in BV(\Omega) \\ u \geq 0 \text{ a.e.}}}{\arg \min} \left\{ F_n(v) := D_{KL}(f_n, Kv) + \alpha |v|_{BV(\Omega)} \right\}, \quad n \in \mathbb{N}, \quad (67)$$

and let u be a solution of the regularized problem (52) corresponding to the data function f . Additionally, we assume that the operator K does not annihilate constant functions and that $\log f$ and $\log Ku$ belong to the function space $L_\mu^\infty(\Sigma)$, i.e. there exist positive constants c_1, \dots, c_4 such that

$$0 < c_1 \leq f \leq c_2 \quad \text{and} \quad 0 < c_3 \leq Ku \leq c_4 \quad \text{a.e. on } \Sigma. \quad (68)$$

Moreover, we suppose that the sequence $\{f_n\}$ is uniformly bounded in the $L_\mu^1(\Sigma)$ -norm, i.e. it exists a positive constant c_5 such that

$$\|f_n\|_{L_\mu^1(\Sigma)} \leq c_5, \quad \forall n \in \mathbb{N}. \quad (69)$$

Then, the problem (52) is stable with respect to the perturbations in the data, i.e. the sequence $\{u_n\}$ has a convergent subsequence and every convergent subsequence converges in the L^1 norm to a minimizer of the functional F in (52).

Proof. For the existence of a convergent subsequence of $\{u_n\}$, we will use the precompactness result from Lemma 4.5 (iii). To this end, we have to show the uniform boundedness of the sequence $\{F_n(u_n)\}$ and the uniform BV-coercivity of the functionals F_n , see the definition below.

We show the uniform boundedness of the sequence $\{F_n(u_n)\}$ first. Let $\alpha > 0$ be a fixed regularization parameter. For any $n \in \mathbb{N}$, the definition of u_n as a minimizer of the objective functional F_n in (67) implies that

$$F_n(u_n) = D_{KL}(f_n, Ku_n) + \alpha |u_n|_{BV(\Omega)} \leq D_{KL}(f_n, Ku) + \alpha |u|_{BV(\Omega)}. \quad (70)$$

Hence, the sequence $\{F_n(u_n)\}$ is bounded, if the sequence $\{D_{KL}(f_n, Ku)\}$ on the right-hand side of (70) is bounded. To show this, we use the uniform boundedness of sequence $\{f_n\}$ in the $L_\mu^1(\Sigma)$ -norm (69). Hence, condition (66) and the result in Corollary 4.4 yield the strong convergence of $\{f_n\}$ to f in $L_\mu^1(\Sigma)$, i.e. we have

$$\lim_{n \rightarrow \infty} \|f - f_n\|_{L_\mu^1(\Sigma)} = 0. \quad (71)$$

Thus, the condition (68) implies together with the inequality

$$\begin{aligned} \left| D_{KL}(f_n, Ku) - D_{KL}(f, Ku) - D_{KL}(f_n, f) \right| &= \left| \int_\Sigma (\log Ku - \log f)(f - f_n) d\mu \right| \\ &\leq \underbrace{\|\log Ku - \log f\|_{L_\mu^\infty(\Sigma)}}_{< \infty} \underbrace{\|f - f_n\|_{L_\mu^1(\Sigma)}}_{\stackrel{(71)}{0}}, \end{aligned}$$

the following convergence,

$$\lim_{n \rightarrow \infty} D_{KL}(f_n, Ku) = D_{KL}(f, Ku). \quad (72)$$

Because u is a minimizer of the regularized problem (52) corresponding to the data function f , the expressions $D_{KL}(f, Ku)$ and $|u|_{BV(\Omega)}$ are bounded, and thus also the sequence $\{D_{KL}(f_n, Ku)\}$ is bounded, since convergent to $D_{KL}(f, Ku)$. Together with (70) this fact yields the uniform boundedness of the sequence $\{F_n(u_n)\}$.

Next, we prove that the functionals F_n are uniform BV-coercive, i.e. for any sequence $\{u_n\}$ in $L^1(\Omega)$ with $u_n \geq 0$ a.e.,

$$F_n(u_n) \rightarrow +\infty \quad \text{whenever} \quad \|u_n\|_{BV(\Omega)} \rightarrow +\infty .$$

For the proof we put $u_n = w_n + v_n$ as in (56) and (57), and repeat the proof of Lemma 4.7 with u_n and F_n instead of u and F . Since the operator K does not annihilate constant functions, we obtain $\|Kw_n\|_{L^1_\mu(\Sigma)} = C_2 \|w_n\|_{L^1(\Omega)}$ with C_2 as in (61) and hence analogous to (62),

$$\begin{aligned} C_2 \|w_n\|_{L^1(\Omega)} \left(C_2 \|w_n\|_{L^1(\Omega)} - 2 \left(\|K\| C_1 \frac{1}{\alpha} F_n(u_n) + \|f_n\|_{L^1_\mu(\Sigma)} \right) - \frac{4}{3} F_n(u_n) \right) \\ \leq \left(\frac{2}{3} \|f_n\|_{L^1_\mu(\Sigma)} + \frac{4}{3} \|K\| C_1 \frac{1}{\alpha} F_n(u_n) \right) F_n(u_n) . \end{aligned}$$

Since the sequence $\{f_n\}$ converges strongly to f in $L^1_\mu(\Sigma)$ (71), it is also bounded in the $L^1_\mu(\Sigma)$ -norm. The upper bound on each $\|f_n\|_{L^1_\mu(\Sigma)}$ and the boundedness of the operator norm of K yield the uniform BV-coercivity of the functionals F_n analogous to both cases in the proof of Lemma 4.7.

The uniform BV-coercivity implies together with the boundedness of the sequence $\{F_n(u_n)\}$ that the sequence $\{u_n\}$ is uniformly bounded in the BV norm. Then, the precompactness result from Lemma 4.5 (iii) ensures the existence of a subsequence $\{u_{n_j}\}$ which converges strongly to some $\tilde{u} \in L^1(\Omega)$. Actually, the function \tilde{u} lies in $BV(\Omega)$, since $|\cdot|_{BV(\Omega)}$ is lower semicontinuous with respect to the weak topology of $L^1(\Omega)$ (see Lemma 4.5 (ii)), i.e. we have

$$|\tilde{u}|_{BV(\Omega)} \leq \liminf_{j \rightarrow \infty} |u_{n_j}|_{BV(\Omega)} < \infty .$$

Now let $\{u_{n_j}\}$ be an arbitrary subsequence of $\{u_n\}$ which converges to some $\tilde{u} \in L^1(\Omega)$ with respect to the L^1 -norm. The boundedness of the operator K (see Assumption 4.1 (i)) implies the strong convergence of the sequence $\{Ku_{n_j}\}$ to $K\tilde{u}$ in $L^1_\mu(\Sigma)$, 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 $L^1_\mu(\Sigma)$ (71). 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 4.1 (ii)), we can apply Fatou's Lemma to the sequence $\{f_{n_j} \log(f_{n_j} / Ku_{n_j}) - f_{n_j} + Ku_{n_j}\}$ and obtain

$$D_{KL}(f, K\tilde{u}) \leq \liminf_{j \rightarrow \infty} D_{KL}(f_{n_j}, Ku_{n_j}) . \quad (73)$$

Due to the lower semicontinuity of the functional $|\cdot|_{BV(\Omega)}$ (see Lemma 4.5 (ii)) and due to (70), (72) and (73), we now obtain the following inequality,

$$\begin{aligned} D_{KL}(f, K\tilde{u}) + \alpha |\tilde{u}|_{BV(\Omega)} &\stackrel{(73)}{\leq} \liminf_{j \rightarrow \infty} D_{KL}(f_{n_j}, Ku_{n_j}) + \alpha \liminf_{j \rightarrow \infty} |u_{n_j}|_{BV(\Omega)} \\ &\leq \liminf_{j \rightarrow \infty} \left(D_{KL}(f_{n_j}, Ku_{n_j}) + \alpha |u_{n_j}|_{BV(\Omega)} \right) \\ &\leq \limsup_{j \rightarrow \infty} \left(D_{KL}(f_{n_j}, Ku_{n_j}) + \alpha |u_{n_j}|_{BV(\Omega)} \right) \\ &\stackrel{(70)}{\leq} \limsup_{j \rightarrow \infty} \left(D_{KL}(f_{n_j}, Ku) + \alpha |u|_{BV(\Omega)} \right) \end{aligned}$$

$$\stackrel{(72)}{=} D_{KL}(f, Ku) + \alpha |u|_{BV(\Omega)} ,$$

which means that \tilde{u} is a minimizer of the functional F in (52). □

Remark.

- For the proof of stability, condition (68) is required, which assumes that the functions $\log f$ and $\log Ku$ belong to $L_\mu^\infty(\Sigma)$, where u is a regularized solution of the minimization problem (52). 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 (68) is not simple to justify and requires a more precise analysis of the variational problem (52). Due to Assumption 4.1 (iii), it suffices to prove that u is bounded and bounded away from zero. For instance, the authors in [67] show that this condition on u is available, if we replace the TV functional in (52) by the KL functional $D_{KL}(\cdot, u^*)$ as regularization energy, where u^* denotes the a-priori estimation of the solution and satisfies 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, in the case of total variation regularization we did not yet succeed to prove a similar property. Nevertheless, note that in Section 4.3 we can show at least that the iterate sequence $\{u_k\}$ of the FB-EM-TV splitting algorithm (29) has the boundedness and the boundedness away from zero property, assuming that the data function f belongs to $L_\mu^\infty(\Sigma)$ with $\inf_\Sigma f > 0$ and the initialization function u_0 is strictly positive. For this reason, we think that condition (68) is an acceptable assumption.
- Analogous to the reasoning above, the convergence of subsequences $\{u_{n_j}\}$ in Theorem 4.10 can also be proved in the $L^p(\Omega)$ -norm with $1 \leq p < d/(d-1)$, $\Omega \subset \mathbb{R}^d$, since any uniformly bounded sequence $\{u_n\}$ in $BV(\Omega)$ is actually relatively compact in $L^p(\Omega)$ for $1 \leq p < d/(d-1)$, see [1, Thm. 2.5].
- As in Theorem 4.10 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 (68), 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 (52). Therefore, this condition is severely restrictive for the possible perturbations of the operator K .
- We finally mention that some stability estimates for the TV regularized Poisson likelihood estimation problem (52) have been also derived in [8], 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.

4.3. Positivity preservation of the FB-EM-TV algorithm

In the following we consider a particular property of the iteration sequence $\{u_k\}$ obtained by the FB-EM-TV splitting approach (29) and its damped modification (31), namely the positivity preservation of this iteration scheme. Given a strictly positive $u_k \in BV(\Omega)$ 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 (29), if the data function f is strictly positive and the operator K fulfills the positivity preservation property in Assumption 4.1 (iii). Consequently, an existence and uniqueness proof for the regularization half step (30) and its damped variant (32), analogous to the classical results for the ROF model, delivers also the existence of $u_{k+1} \in BV(\Omega)$. 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 (30) 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 (21).

Now, to clarify the positivity preservation of the (damped) FB-EM-TV iteration scheme, we present a maximum principle for the following weighted ROF problem,

$$\min_{u \in BV(\Omega)} J(u) := \frac{1}{2} \int_{\Omega} \frac{(u - q)^2}{h} dx + \beta |u|_{BV(\Omega)}, \quad \beta > 0, \quad (74)$$

which represents the more general form of the regularization half step (30) and its damped modification (32) in the forward-backward splitting strategy.

Lemma 4.11 (Maximum Principle for the Weighted ROF Problem). *Let $\tilde{u} \in BV(\Omega)$ be a minimizer of the variational problem (74), where the function q belongs to $L^\infty(\Omega)$ with $\inf_{\Omega} q > 0$ and let the weighting function h be strictly positive. Then the following maximum principle holds*

$$0 < \inf_{\Omega} q \leq \inf_{\Omega} \tilde{u} \leq \sup_{\Omega} \tilde{u} \leq \sup_{\Omega} q. \quad (75)$$

Proof. Let \tilde{u} be a minimizer of the functional J defined in (74). For the proof of the maximum principle, we show that there exists a function v with

$$0 < \inf_{\Omega} q \leq \inf_{\Omega} v \leq \sup_{\Omega} v \leq \sup_{\Omega} q \quad (76)$$

and

$$J(v) \leq J(\tilde{u}). \quad (77)$$

Then, the desired boundedness property (75) follows directly from the strict convexity of the functional J in (74), 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.

$$v := \min\{\max\{\tilde{u}, \inf_{\Omega} q\}, \sup_{\Omega} q\}.$$

With this definition, the property (76) is directly guaranteed. To show (77), we use

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

and estimate $|v|_{BV(\Omega)}$ by $|\tilde{u}|_{BV(\Omega)}$ first. Since the function v has (due to its definition) no variation on $\Omega \setminus M$, we obtain

$$|v|_{BV(\Omega)} = |v|_{BV(M)} = |\tilde{u}|_{BV(M)} \leq |\tilde{u}|_{BV(\Omega)}. \quad (78)$$

Moreover, we see that the data fidelity terms of J in (74) with respect to v and \tilde{u} coincide on M , due to the definition of the function v . In case of $x \in \Omega \setminus M$, we distinguish between two cases:

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

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

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

$$\begin{aligned} 0 &\leq -v(x) + q(x) = -\inf q + q(x) \leq -\tilde{u}(x) + q(x) \\ &\Rightarrow (v(x) - q(x))^2 \leq (\tilde{u}(x) - q(x))^2. \end{aligned}$$

Finally, we obtain

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

and property (77) is fulfilled due to the strict positivity of the weighting function h and (78). \square

Lemma 4.12 (Positivity of (damped) FB-EM-TV Algorithm). *Let $\{\omega_k\}$ be a given sequence of damping parameters with $\omega_k \in (0, 1]$ for all $k \geq 0$ and $u_0 > 0$. Additionally, we assume that $f \in L^\infty_\mu(\Sigma)$ with $\inf_\Sigma f > 0$ and that the operator K satisfies the positivity preservation property in Assumption 4.1 (iii). Then, each half step of the (damped) FB-EM-TV 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 (29) is strictly positive. Because the TV correction step in (29) can be realized via the weighted ROF problem (30), the maximum principle in Lemma 4.11 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 (31) via the variational problem (32), 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 4.11. \square

Finally we consider also the positivity preservation of the Poisson denoising strategy, which can be obtained from the (damped) FB-EM-TV algorithm by using the identity operator K . In this case the FB-EM-TV splitting strategy (29) with the damped modification (31) result in the following iteration scheme (note that the EM reconstruction step vanishes in the denoising case, i.e. it holds $u_{k+\frac{1}{2}} = f$),

$$u_{k+1} = (1 - \omega_k) u_k + \omega_k f - \omega_k \alpha u_k p_{k+1}, \quad p_{k+1} \in \partial |u_{k+1}|_{BV(\Omega)}, \quad (79)$$

with $\omega_k \in (0, 1]$, in order to denoise an image corrupted by Poisson noise. Analogous to the FB-EM-TV algorithm, this iteration step can be realized by solving a weighted ROF problem of the form (cf. (30) and (32)),

$$u_{k+1} \in \arg \min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{(u - (\omega_k f + (1 - \omega_k) u_k))^2}{u_k} + \omega_k \alpha |u|_{BV(\Omega)} \right\}. \quad (80)$$

Although this denoising iteration is a special case of the damped FB-EM-TV algorithm, we study its properties here explicitly, because it will later simplify the convergence criteria of the Poisson denoising method.

Lemma 4.13 (Maximum Principle and Positivity of the Poisson Denoising Scheme). *Let $\{\omega_k\}$ be a sequence of damping parameters with $\omega_k \in (0, 1]$ for all $k \geq 0$, the data function f lies in $L_{\mu}^{\infty}(\Omega)$ with $\inf_{\Omega} f > 0$ and the initialization function u_0 fulfills*

$$0 < \inf_{\Omega} f \leq \inf_{\Omega} u_0 \leq \sup_{\Omega} u_0 \leq \sup_{\Omega} f. \quad (81)$$

Moreover, let $\{u_k\}$ be a sequence of iterates generated by the damped Poisson denoising scheme (80). Then, the following maximum principle holds,

$$0 < \inf_{\Omega} f \leq \inf_{\Omega} u_k \leq \sup_{\Omega} u_k \leq \sup_{\Omega} f, \quad \forall k \geq 0. \quad (82)$$

Simultaneously, this result guarantees also that each step of the damped Poisson denoising method (79) and with it also the solution is strictly positive.

Proof. We prove the assertion by induction. For $k = 0$, the condition (82) is fulfilled due to (81). For a general $k \geq 0$, Lemma 4.11 offers a maximum principle for the Poisson denosing model (80) using $q := \omega_k f + (1 - \omega_k) u_k$ and $h := u_k$, i.e. we have

$$\begin{aligned} 0 < \inf_{\Omega} \{\omega_k f + (1 - \omega_k) u_k\} &\leq \inf_{\Omega} u_{k+1} \\ &\leq \sup_{\Omega} u_{k+1} \leq \sup_{\Omega} \{\omega_k f + (1 - \omega_k) u_k\}. \end{aligned} \quad (83)$$

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 (83) and the induction hypothesis the desired maximum principle (82). \square

Remark. The assumption (81) on the initialization function u_0 is fulfilled in general, since u_0 will be usually chosen as a positive and constant function or as the given noisy image f itself.

4.4. Convergence of the damped FB-EM-TV algorithm

In Section 3.3 we interpreted the (damped) FB-EM-TV reconstruction method 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 [79] and Gabay [38]. For the proposed algorithm (35),

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

Gabay provided in [38] a proof of 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 [79] 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,

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

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 (21) with respect to the iterates u_k of the FB-EM-TV algorithm. In the following theorem we will establish the convergence of the damped FB-EM-TV splitting algorithm under appropriate assumptions on the damping parameters ω_k .

Theorem 4.14 (Convergence of Damped FB-EM-TV Algorithm). *Let K satisfy Assumption 4.1 and do not annihilate constant functions. Moreover, let (u_k) be a sequence of iterates obtained by the damped FB-EM-TV algorithm (34) and let the data function $f \in L^\infty_\mu(\Sigma)$ fulfill $\inf_\Sigma f > 0$. Now, if there exists a sequence of corresponding damping parameters $\{\omega_k\}$, $\omega_k \in (0, 1]$, satisfying the inequality*

$$\omega_k \leq \frac{\int_\Omega \frac{K^* \mathbf{1}_\Sigma (u_{k+1} - u_k)^2}{u_k} d\lambda}{\sup_{v \in [u_k, u_{k+1}]} \frac{1}{2} \int_\Sigma \frac{f(Ku_{k+1} - Ku_k)^2}{(Kv)^2} d\mu} (1 - \epsilon) , \quad \epsilon \in (0, 1) , \quad (84)$$

then the objective functional F defined in (52) is decreasing during the iteration. If, in addition, 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 $k \geq 0$,*

$$0 < c_1 \leq K^* \mathbf{1}_\Sigma , \quad 0 < c_2 \leq \omega_k , \quad 0 < c_3 \leq u_k , \quad (85)$$

then the sequence of iterates $\{u_k\}$ has a convergent subsequence in the weak topology on $BV(\Omega)$ and in the strong topology on $L^1(\Omega)$ and every such a convergent subsequence converges to a minimizer of the functional F .*

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 TV regularization functional at the limit u of the primal iterates

$\{u_k\}$, i.e. $p \in \partial |u|_{BV(\Omega)}$. In the last step, we show that the limit u actually 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 $\{\omega_k\}$, which guarantees for all $k \geq 0$ a descent of the form

$$F(u_{k+1}) + \underbrace{\frac{\epsilon}{\omega_k} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u_{k+1} - u_k)^2}{u_k} d\lambda}_{\geq 0} \leq F(u_k), \quad \epsilon > 0. \quad (86)$$

This condition ensures actually a descent of F , since the second term on the left-hand side of (86) is positive due to $\omega_k > 0$ and due to the strict positivity of the iterates u_k (see Lemma 4.12). To show (86), we start with the damped TV regularization step (31), multiply it with $u_{k+1} - u_k$ and integrate the result over the domain Ω . Thus, for $p_{k+1} \in \partial |u_{k+1}|_{BV(\Omega)}$, we obtain

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

Due to the definition of subgradients in (23), we now have that

$$\langle p_{k+1}, u_{k+1} - u_k \rangle \geq |u_{k+1}|_{BV(\Omega)} - |u_k|_{BV(\Omega)}$$

and thus

$$\begin{aligned} \alpha |u_{k+1}|_{BV(\Omega)} - \alpha |u_k|_{BV(\Omega)} + \frac{1}{\omega_k} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u_{k+1} - u_k)^2}{u_k} d\lambda \\ \leq - \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u_k - u_{k+\frac{1}{2}})(u_{k+1} - u_k)}{u_k} d\lambda. \end{aligned}$$

Adding the difference $D_{KL}(f, Ku_{k+1}) - D_{KL}(f, Ku_k)$ on both sides of this inequality and considering the definitions of the KL functional D_{KL} in (51) and the objective functional F in (52) yields

$$\begin{aligned} F(u_{k+1}) - F(u_k) + \frac{1}{\omega_k} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u_{k+1} - u_k)^2}{u_k} d\lambda \\ \leq \int_{\Sigma} \left(f \log \left(\frac{f}{Ku_{k+1}} \right) + Ku_{k+1} - f \log \left(\frac{f}{Ku_k} \right) - Ku_k \right) d\mu \\ \quad - \int_{\Omega} \left(K^* \mathbf{1}_{\Sigma} (u_{k+1} - u_k) - \frac{K^* \mathbf{1}_{\Sigma} u_{k+\frac{1}{2}}}{u_k} (u_{k+1} - u_k) \right) d\lambda \quad (87) \\ = \int_{\Sigma} \left(f \log \left(\frac{f}{Ku_{k+1}} \right) - f \log \left(\frac{f}{Ku_k} \right) \right) d\mu \\ \quad + \int_{\Omega} \left(K^* \left(\frac{f}{Ku_k} \right) (u_{k+1} - u_k) \right) d\lambda. \end{aligned}$$

The last equality in (87) holds, since $u_{k+\frac{1}{2}}$ is given by the EM reconstruction step in (29) and K^* is the adjoint operator of K , i.e. we have

$$\int_{\Sigma} Ku \, d\mu = \langle \mathbf{1}_{\Sigma}, Ku \rangle_{\Sigma} = \langle K^* \mathbf{1}_{\Sigma}, u \rangle_{\Omega} = \int_{\Omega} K^* \mathbf{1}_{\Sigma} u \, d\lambda.$$

Now, we characterize the right-hand side of (87) using an auxiliary functional

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

and consider the directional derivatives of G . That is, we define for any $w_1 \in L^1(\Omega)$ a function $\phi_{w_1}(t) := G(u + tw_1)$ and see that the directional derivative $G'(u; w_1)$ of G at u in the direction w_1 is given by

$$\begin{aligned} G'(u; w_1) &= \phi'_{w_1}(t) \Big|_{t=0} = \int_{\Sigma} \frac{\partial}{\partial t} \left(f \log \left(\frac{f}{Ku + tKw_1} \right) \right) d\mu \Big|_{t=0} \\ &= \left\langle -\frac{f}{Ku}, Kw_1 \right\rangle_{\Sigma} = \left\langle -K^* \left(\frac{f}{Ku} \right), w_1 \right\rangle_{\Omega}. \end{aligned}$$

Interpreting the right-hand side of inequality (87) formally as a Taylor linearization of G yields

$$\begin{aligned} F(u_{k+1}) - F(u_k) &+ \frac{1}{\omega_k} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u_{k+1} - u_k)^2}{u_k} d\lambda \\ &\leq G(u_{k+1}) - G(u_k) - G'(u_k; u_{k+1} - u_k) \\ &= \frac{1}{2} G''(v; u_{k+1} - u_k, u_{k+1} - u_k), \quad v \in [u_k, u_{k+1}], \\ &\leq \sup_{v \in [u_k, u_{k+1}]} \frac{1}{2} G''(v; u_{k+1} - u_k, u_{k+1} - u_k). \end{aligned} \tag{88}$$

We can now compute the second directional derivative $G''(u; w_1, w_2)$ of G using the function $\phi_{w_2}(t) := G(u + tw_2; w_1)$ for any $w_2 \in L^1(\Omega)$,

$$\begin{aligned} G''(u; w_1, w_2) &= \phi'_{w_2}(t) \Big|_{t=0} = - \int_{\Sigma} \frac{\partial}{\partial t} \left(\frac{f}{Ku + tKw_2} Kw_1 \right) d\mu \Big|_{t=0} \\ &= \int_{\Sigma} \frac{f Kw_2 Kw_1}{(Ku)^2} d\mu. \end{aligned}$$

Plugging the computed derivative $G''(u; w_1, w_2)$ in the inequality (88), we obtain

$$\begin{aligned} F(u_{k+1}) - F(u_k) &+ \frac{1}{\omega_k} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u_{k+1} - u_k)^2}{u_k} d\lambda \\ &\leq \sup_{v \in [u_k, u_{k+1}]} \frac{1}{2} \int_{\Sigma} \frac{f (Ku_{k+1} - Ku_k)^2}{(Kv)^2} d\mu. \end{aligned} \tag{89}$$

Finally, we split the third term on the left-hand side of (89) with $\epsilon \in (0, 1)$,

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

and obtain the desired condition (86), i.e. a descent of the objective functional F , if

$$\sup_{v \in [u_k, u_{k+1}]} \frac{1}{2} \int_{\Sigma} \frac{f(Ku_{k+1} - Ku_k)^2}{(Kv)^2} d\mu \leq \frac{1 - \epsilon}{\omega_k} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u_{k+1} - u_k)^2}{u_k} d\lambda. \quad (90)$$

By solving (90) for w_k , we obtain the required condition (84) for the damping parameters $\{\omega_k\}$ in order to have a descent of the objective functional F . Additionally, by a suitable choice of ϵ in (84), 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 the functional F , we need a convergent subsequence of the primal iterates $\{u_k\}$. Since the operator K does not annihilate constant functions, the functional F is BV-coercive according to Lemma 4.7 and we obtain from (55),

$$\|u_k\|_{BV(\Omega)} \leq c_4 (F(u_k))^2 + c_5 F(u_k) + c_6 \leq c_4 (F(u_0))^2 + c_5 F(u_0) + c_6,$$

for all $k \geq 0$ with constants $c_4 \geq 0$, $c_5 > 0$ and $c_6 \geq 0$. The latter inequality holds due to the positivity of F and due to the monotone decrease of the sequence $\{F(u_k)\}$ with the corresponding choice of the damping parameters $\{\omega_k\}$ in (84). For this reason, the sequence $\{u_k\}$ is uniformly bounded in the BV norm and the Banach-Alaoglu theorem (see e.g. [57, Thm. 2.6.18]) yields the compactness in the weak* topology on $BV(\Omega)$, which implies the existence of a subsequence $\{u_{k_l}\}$ with

$$u_{k_l} \rightharpoonup^* u \quad \text{in } BV(\Omega).$$

Moreover, the definition of the weak* topology on $BV(\Omega)$ in [2, Def. 3.11] also implies the strong convergence of the subsequence $\{u_{k_l}\}$ in $L^1(\Omega)$,

$$u_{k_l} \rightarrow u \quad \text{in } L^1(\Omega).$$

Subsequently, we can also consider the sequence $\{u_{k+1}\}$ and can choose with the same argumentation further subsequences, again denoted by k_l , such that

$$\begin{aligned} u_{k_l+1} &\rightharpoonup^* \tilde{u} && \text{in } BV(\Omega), \\ u_{k_l+1} &\rightarrow \tilde{u} && \text{in } L^1(\Omega). \end{aligned}$$

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 (86) recursively and obtain the following estimate,

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

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

$$\lim_{k \rightarrow \infty} \int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u_{k+1} - u_k)^2}{\omega_k u_k} d\lambda = 0. \quad (91)$$

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

$$\|u_{k+1} - u_k\|_{L^1(\Omega)}^2 \leq \underbrace{\int_{\Omega} \frac{\omega_k u_k}{K^* \mathbf{1}_{\Sigma}} d\lambda}_{\stackrel{(85)}{\leq} c_1 \|u_k\|_{L^1(\Omega)}} \underbrace{\int_{\Omega} \frac{K^* \mathbf{1}_{\Sigma} (u_{k+1} - u_k)^2}{\omega_k u_k} d\lambda}_{\stackrel{(91)}{\rightarrow} 0}. \quad (92)$$

The first term on the right-hand side of (92) 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 (85) and the sequence $\{u_k\}$ is uniformly bounded in the BV norm. Moreover, since the second term on the right-hand side of (92) converges to zero (cf. (91)), we obtain from (92) that

$$u_{k+1} - u_k \rightarrow 0 \quad \text{in } L^1(\Omega) . \quad (93)$$

Hence, the uniqueness of the limit 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. $p_k \in \partial |u_k|_{BV(\Omega)}$. To this end, we use the general property that the subdifferentials of a convex one-homogeneous functional $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$, X Banach space, can be characterized by

$$\partial J(u) = \{ p \in X^* : \langle p, u \rangle = J(u), \quad \langle p, v \rangle \leq J(v) \quad \forall v \in X \} .$$

In the case of TV, we see that for each subgradient p_k the dual norm is bounded by

$$\|p_k\| = \sup_{\|v\|_{BV(\Omega)} = 1} \langle p_k, v \rangle \leq \sup_{\|v\|_{BV(\Omega)} = 1} |v|_{BV(\Omega)} \leq \sup_{\|v\|_{BV(\Omega)} = 1} \|v\|_{BV(\Omega)} = 1 .$$

Hence, the sequence $\{p_k\}$ is uniformly bounded in the $(BV(\Omega))^*$ -norm and the Banach-Alaoglu theorem yields the compactness in the weak* topology on $(BV(\Omega))^*$, which implies the existence of a subsequence, again denoted by k_l , such that

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

Fourth step: Show that $p \in \partial |u|_{BV(\Omega)}$

Now we have the weak* convergence of sequences $\{u_{k_l}\}$ and $\{u_{k_l+1}\}$ in $BV(\Omega)$ and the weak* convergence of $\{p_{k_l+1}\}$ in $(BV(\Omega))^*$. Next, we will show that the limit p of the dual iterates is a subgradient of $|\cdot|_{BV(\Omega)}$ at the limit u of the primal iterates, i.e. $p \in \partial |u|_{BV(\Omega)}$. Hence we have to prove (see the definition of the subgradients in (23)) that

$$|u|_{BV(\Omega)} + \langle p, v - u \rangle \leq |v|_{BV(\Omega)} , \quad \forall v \in BV(\Omega) . \quad (94)$$

For this purpose, let $p_{k_l+1} \in \partial |u_{k_l+1}|_{BV(\Omega)}$, then the definition of the subgradient of $|\cdot|_{BV(\Omega)}$ (23) yields

$$|u_{k_l+1}|_{BV(\Omega)} + \langle p_{k_l+1}, v - u_{k_l+1} \rangle \leq |v|_{BV(\Omega)} , \quad \forall v \in BV(\Omega) . \quad (95)$$

Since $|\cdot|_{BV(\Omega)}$ is lower semicontinuous (see Lemma 4.5 (ii)), we can estimate the BV seminorm at u from above,

$$|u|_{BV(\Omega)} \leq \liminf_{l \rightarrow \infty} |u_{k_l+1}|_{BV(\Omega)} \leq |u_{k_l+1}|_{BV(\Omega)} ,$$

and (95) delivers

$$|u|_{BV(\Omega)} + \langle p_{k_l+1}, v - u_{k_l+1} \rangle \leq |v|_{BV(\Omega)} , \quad \forall v \in BV(\Omega) . \quad (96)$$

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

$$\langle p_{k_l+1}, v \rangle \rightarrow \langle p, v \rangle , \quad \forall v \in BV(\Omega) .$$

Hence, to prove $p \in \partial|u|_{BV(\Omega)}$, it remains to show with respect to (96) and (94) that

$$\langle p_{k_l+1}, u_{k_l+1} \rangle \rightarrow \langle p, u \rangle. \quad (97)$$

For this purpose we consider the complete iteration scheme of the damped FB-EM-TV algorithm (31) with $u_{k+\frac{1}{2}}$ in (29),

$$u_{k_l+1} - (1 - \omega_{k_l}) u_{k_l} - \omega_{k_l} \left(\frac{u_{k_l}}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u_{k_l}} \right) \right) + \omega_{k_l} \alpha \frac{u_{k_l}}{K^* \mathbf{1}_\Sigma} p_{k_l+1} = 0, \quad (98)$$

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

$$\begin{aligned} -\alpha \langle p_{k_l+1}, u_{k_l+1} \rangle &= \int_{\Omega} \frac{K^* \mathbf{1}_\Sigma (u_{k_l+1} - u_{k_l}) u_{k_l+1}}{\omega_{k_l} u_{k_l}} d\lambda + \left\langle \mathbf{1}_\Sigma - \frac{f}{K u_{k_l}}, K u_{k_l+1} \right\rangle \\ &= \underbrace{\int_{\Omega} \frac{K^* \mathbf{1}_\Sigma (u_{k_l+1} - u_{k_l})^2}{\omega_{k_l} u_{k_l}} d\lambda}_{\xrightarrow{(91)} 0} + \left\langle \mathbf{1}_\Sigma - \frac{f}{K u_{k_l}}, K u_{k_l+1} \right\rangle \\ &\quad + \underbrace{\int_{\Omega} \frac{K^* \mathbf{1}_\Sigma (u_{k_l+1} - u_{k_l}) u_{k_l}}{\omega_{k_l} u_{k_l}} d\lambda}_{\xrightarrow{(93)} 0} \end{aligned} \quad (99)$$

The second term on the right-hand side of (99) 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 ω_k (85) and the boundedness preservation of K in Assumption 4.1 (iii)) and due to the $L^1(\Omega)$ -norm convergence (93). Now, using the boundedness of the operator K for the convergence of $\frac{f}{K u_{k_l}}$, we obtain

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

and thus can deduce from (99) that

$$-\alpha \langle p_{k_l+1}, u_{k_l+1} \rangle \rightarrow \int_{\Omega} \left(K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{K u} \right) \right) u d\lambda \stackrel{(101)}{=} -\alpha \langle p, u \rangle.$$

Hence, we can conclude (97) and therewith $p \in \partial|u|_{BV(\Omega)}$.

Fifth step: Convergence to a minimizer of the objective functional

Now, let $\{u_{k_l}\}$ and $\{u_{k_l+1}\}$ be arbitrary convergent subsequences of the primal iteration sequence $\{u_k\}$, which converge to some $u \in BV(\Omega)$ in the weak* topology on $BV(\Omega)$ and in the strong topology on $L^1(\Omega)$. 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 convergence to some $p \in (BV(\Omega))^*$ such that $p \in \partial|u|_{BV(\Omega)}$. To verify the convergence of the damped FB-EM-TV splitting algorithm, it remains to show that u is a minimizer of the functional F . For this purpose, we consider the complete iteration scheme of the damped FB-EM-TV algorithm (98) with reference to the convergent subsequences and show their weak* convergence to the optimality

condition (24) of the variational problem (22). Note that it actually suffices to prove only the convergence to (24) and not to (28), since the function u is positive due to the strict positivity assumption on the iterates u_k for all $k \geq 0$ in (85). An equivalent formulation to equation (98) reads as follows

$$\frac{u_{k_l+1} - u_{k_l}}{\omega_{k_l} u_{k_l}} + \mathbf{1}_\Omega - \frac{1}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u_{k_l}} \right) + \frac{\alpha}{K^* \mathbf{1}_\Sigma} p_{k_l+1} = 0. \quad (100)$$

The convergence can be verified in the following way. Due to the boundedness away from zero assumptions in (85), we can use result (91) in order to deduced the following convergence,

$$c_1 c_2 c_3 \int_\Omega \frac{(u_{k+1} - u_k)^2}{\omega_k^2 u_k^2} d\lambda \leq \int_\Omega \frac{K^* \mathbf{1}_\Sigma (u_{k+1} - u_k)^2}{\omega_k^2 u_k^2} \omega_k u_k d\lambda \xrightarrow{(91)} 0.$$

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

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

Therefore, if we pass over to the weak* limit of the subsequences in (100) using the boundedness of the operator K in Assumption 4.1 (i) for the convergence of $\frac{f}{K u_{k_l}}$, we obtain that both limit functions u and p of the sequences $\{u_{k_l}\}$ and $\{p_{k_l+1}\}$ fulfill the optimality condition (24) of the variational problem (22),

$$\mathbf{1}_\Omega - \frac{1}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u} \right) + \frac{\alpha}{K^* \mathbf{1}_\Sigma} p = 0. \quad (101)$$

This means that the subsequence $\{u_{k_l}\}$ converges in the weak* topology on $BV(\Omega)$ and in the strong topology on $L^1(\Omega)$ to a minimizer of the functional F . \square

Remark.

- We note that inequality (90) in the proof above at the same time motivates the mathematical necessity of a damping in the FB-EM-TV splitting strategy. In the undamped case, i.e. $\omega_k = 1$, the term on the right-hand side of (90) is maximal for $\epsilon \rightarrow 0^+$, due to the strict positivity of $K^* \mathbf{1}_\Sigma$ and u_k for all $k \geq 0$ in (85). In general, one cannot say whether this term is greater than the supremum on the left-hand side of (90) 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 (90) in order to guarantee a descent of the objective functional F .
- Analogous to the proof above, the strong convergence of the sequence $\{u_k\}$ to a minimizer of the functional F in Theorem 4.14 can also be proved in the L^p norm with $1 \leq p < d/(d-1)$, since any uniformly bounded sequence $\{u_k\}$ in $BV(\Omega)$ is actually relatively compact in $L^p(\Omega)$ for $1 \leq p < d/(d-1)$ (see [1, Thm. 2.5]). Therefore, since the subsequence $\{u_{k_l}\}$ is furthermore uniformly bounded in the BV norm, there exists a subsequence $\{u_{k_{l_m}}\}$ with

$$u_{k_{l_m}} \rightarrow \tilde{u} \quad \text{in } L^p(\Omega) \quad \text{with } 1 \leq p < d/(d-1).$$

With the uniqueness of the limit and the definition of the weak* topology on $BV(\Omega)$, we obtain

$$\begin{aligned} u_{k_{l_m}} &\rightharpoonup^* u && \text{in } BV(\Omega) , \\ u_{k_{l_m}} &\rightarrow u && \text{in } L^1(\Omega) . \end{aligned}$$

Due to the uniqueness of the limit, i.e. $\tilde{u} = u$, we can pass over in the proof from $\{u_{k_l}\}$ to $\{u_{k_{l_m}}\}$.

- The assumptions on boundedness away from zero in (85) are reasonable from our point of view. In the 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 is fulfilled due to the strict positivity of each half step of the (damped) FB-EM-TV splitting method (see Lemma 4.12).
- Inspired by the relaxed EM reconstruction strategy proposed in [62, Chap. 5.3.2], another possibility of influencing convergence arises in the FB-EM-TV strategy by adding a relaxation parameter $\nu > 0$ to the EM fixed point iteration in the form,

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

Corresponding, one can obtain a reasonable TV denoising step in the FB-EM-TV splitting idea via

$$u_{k+1} = \left(u_{k+\frac{1}{2}}^\nu - \alpha u_k^\nu p_{k+1} \right)^\nu, \quad p_{k+1} \in \partial |u_{k+1}|_{BV(\Omega)}, \quad (\text{relaxed TV step}),$$

with the relaxed EM step $u_{k+\frac{1}{2}}$ above. The relaxed terms in the TV denoising step are necessary to fit the basic variational problem (21) and its optimality condition (24). Due to the computational challenge of the relaxed TV denoising step, which would require again novel methods, a comparison of this strategy with our damping strategy proposed in Section 3.2 would go beyond the scope of this paper.

In practice, determining the damping parameters ω_k via the general condition in (84) 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 (79), i.e. for the identity operator K .

Corollary 4.15 (Convergence of the Damped Poisson Denoising Scheme). *Let $\{u_k\}$ be a sequence of iterates generated by the damped Poisson denoising scheme (80) and let 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 (84) in Theorem 4.14 on the damping parameters simplifies to*

$$\omega_k \leq \frac{2(\inf_\Omega f)^2}{(\sup_\Omega f)^2} (1 - \epsilon), \quad \epsilon \in (0, 1) . \quad (102)$$

Proof. In the special case of the identity operator K , the maximum principle of the damped Poisson denoising scheme from Lemma 4.13 is the main idea for simplifying

the desired condition (84) on the damping parameters. For this sake, we consider the inequality (90), which guarantees a monotone descent of the objective functional if

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

Our goal is now to find an estimate for the coefficients $\frac{f u_k}{2v^2}$. 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 (80), we can use the maximum principle from Lemma 4.13 and obtain an estimate for the coefficients,

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

which should be less or equal $\frac{1-\epsilon}{\omega_k}$. Thus, choosing ω_k according to the estimate (102) guarantees a monotone descent of the objective functional. \square

5. Numerical realization of the weighted ROF problem

To solve the TV regularized Poisson likelihood estimation problem (21), we proposed the (Bregman-)FB-EM-TV algorithm as a nested two step iteration strategy in Sections 3.1 and 3.5. However, we left open the question of the numerical realization of the complex TV correction half step (30), (32), (46) and (47) contained in both iteration strategies. 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 problem (80). The most general form of all the schemes above is

$$\min_{u \in BV(\Omega)} \frac{1}{2} \int_{\Omega} \frac{(u - q)^2}{h} + \beta |u|_{BV(\Omega)}, \quad \beta > 0, \quad (103)$$

with an appropriate setting of the "noise" function q , the weight function h and the regularization parameter β . The choice of all these parameters with respect to the desired restoration method is summarized in Table 1.

Table 1. Overview for the setting of the functions q , h and parameter β in (103) with respect to the different algorithms proposed in Chapter 3.

<i>Algorithm</i>	q	h	β
Poisson Denoising (80)	f	u_k	α
Damped Poisson Denoising (80)	$\omega_k f + (1 - \omega_k) u_k$	u_k	$\omega_k \alpha$
FB-EM-TV Algorithm (30)	$u_{k+\frac{1}{2}}$	$\frac{u_k}{K^* \mathbf{1}_{\Sigma}}$	α
Damped FB-EM-TV Algorithm (32)	$\omega_k u_{k+\frac{1}{2}} + (1 - \omega_k) u_k$	$\frac{u_k}{K^* \mathbf{1}_{\Sigma}}$	$\omega_k \alpha$
Bregman-FB-EM-TV Algorithm (46)	$u_{k+\frac{1}{2}}^{l+1} + u_k^{l+1} v^l$	$\frac{u_k^{l+1}}{K^* \mathbf{1}_{\Sigma}}$	α
Damped Bregman-FB-EM-TV Algorithm (47)	$\omega_k^{l+1} u_{k+\frac{1}{2}}^{l+1} + \omega_k^{l+1} u_k^{l+1} v^l + (1 - \omega_k^{l+1}) u_k^{l+1}$	$\frac{u_k^{l+1}}{K^* \mathbf{1}_{\Sigma}}$	$\omega_k^{l+1} \alpha$

We see that the variational problem (103) is just a modified version of the well known ROF model, with an additional weight h in the data fidelity term. This analogy

creates the opportunity to carry over the different numerical schemes known for the ROF model, e.g. we refer to [24, 4, 23] and the references therein, where most of these computational schemes can be adapted to the weighted modification (103). Here, we use the exact dual TV approach (6) for the minimization of (103), which does not need any smoothing of the total variation. Then, our approach is analogous to the projected gradient descent algorithm of Chambolle in [22], which characterizes the subgradients of TV as divergences of vector fields with supremum norm less or equal one. Using this method, the weighted ROF problem (103) can be solved efficiently, obtaining an accurate and robust algorithm. Note that another efficient numerical scheme to solve the weighted ROF problem (103) has been proposed in [71], which is similar to the alternating split Bregman algorithm of Goldstein and Osher in [42]. In [71] a slightly modified augmented Lagrangian approach of the alternating split Bregman algorithm has been used in order to handle the weight in the data fidelity term better.

In the following, we establish an iterative algorithm to compute the solution of the variational problem (103) using a modified variant of the projected gradient descent algorithm of Chambolle [22]. To this end, the formulation (103) 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),

$$\inf_{u \in BV(\Omega)} \sup_{\substack{g \in C_0^\infty(\Omega, R^d) \\ \|g\|_\infty \leq 1}} L(u, g) := \frac{1}{2} \int_{\Omega} \frac{(u - q)^2}{h} + \beta \int_{\Omega} u \operatorname{div} g . \quad (104)$$

Formally, the infimum regarding u and the supremum regarding p can be swapped. In the case of the standard ROF model, i.e if the weight h in (103) is missing, this property is proved in [59], which can be carried over to the weighted variant (103) with minimal modifications. Moreover, a more precise analysis of this property for general saddle point problems is available in [32, p. 175, Prop. 2.3]. After exchanging inf and sup, the primal optimality condition for the saddle point problem (104) is given by

$$\frac{\partial}{\partial u} L(u, g) = 0 \quad \Leftrightarrow \quad u = q - \beta h \operatorname{div} g . \quad (105)$$

Hence, if an optimal dual variable \tilde{g} is available, the condition (105) can be used to obtain a solution of (104) and (103), i.e. the primal solution u is given by

$$u = q - \beta h \operatorname{div} \tilde{g} . \quad (106)$$

For the computation of \tilde{g} , we substitute (105) into (104) and obtain a purely dual problem which depends on g only. With terms that are constant with respect to optimization variable and hence do not change the supremum, and under the substitution of maximization by minimization of the negative functional, we obtain

$$\begin{aligned} \tilde{g} = \arg \min_{g \in C_0^\infty(\Omega, R^d)} & \int_{\Omega} \frac{(\beta h \operatorname{div} g - q)^2}{h} , \\ \text{s.t. } & |g(x)|_{\ell^2}^2 - 1 \leq 0 , \quad \forall x \in \Omega , \end{aligned} \quad (107)$$

where $|\cdot|_{\ell^2}$ is the Euclidean vector norm. For the choice of this vector norm, compare the remark at the end of this chapter. The constraint for the dual variable g in (107) is a consequence of the exact (dual) definition of total variation (6).

Since the dual problem (107) is a (weighted) quadratic optimization problem with a nonlinear inequality constraint, we use the Karush-Kuhn-Tucker (KKT) conditions

(cf. e.g. [48, Thm. 2.1.4]) to compute the optimal dual variable \tilde{g} . Hence, the KKT conditions yield the existence of a Lagrange multiplier $\lambda(x) \geq 0$ a.e. on Ω , such that

$$-\nabla(\beta h \operatorname{div} g - q)(x) + \lambda(x)g(x) = 0, \quad \forall x \in \Omega, \quad (108)$$

and

$$\lambda(x) (|g(x)|_{\ell^2}^2 - 1) = 0, \quad \forall x \in \Omega. \quad (109)$$

Now, the multiplier λ can be specified explicitly from the complementarity condition (109), which implies that for any $x \in \Omega$,

$$\lambda(x) > 0 \quad \text{and} \quad |g(x)|_{\ell^2} = 1 \quad \text{or} \quad \lambda(x) = 0.$$

Thus, in any case we obtain from (108),

$$\lambda(x) = |\lambda(x)g(x)|_{\ell^2} = |\nabla(\beta h \operatorname{div} g - q)(x)|_{\ell^2}, \quad \forall x \in \Omega,$$

and can write (108) as a fixed point equation for g , obtaining the following iteration sequence,

$$g^{n+1}(x) = \frac{g^n(x) + \tau (\nabla(\beta h \operatorname{div} g^n - q)(x))}{1 + \tau |\nabla(\beta h \operatorname{div} g^n - q)(x)|_{\ell^2}}, \quad \forall x \in \Omega. \quad (110)$$

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 [22, Thm. 3.1] can be transferred to the weighted ROF problem (103). The proof based on the Banach fixed point theorem and required the condition

$$0 < \tau \leq (4d\gamma \|h\|_{L^\infty(\Omega)})^{-1}, \quad (111)$$

in order to obtain a contraction constant less one, where $4d$ is the upper bound of the discrete divergence operator. Hence, we can guarantee the convergence of (110) to a optimal solution, if the damping parameter τ satisfies the condition (111). Note that the weight h can be interpreted as an adaptive regularization, since the regularization parameter β is weighted in (110) by the function h .

Remark. Finally, we point out that the definition of the total variation in (6) 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. For a detailed analysis, we refer e.g. to [58, 34, 14, 77].

6. Numerical results in PET and fluorescence microscopy

In this section we will illustrate the performance of the proposed numerical schemes and will test the theoretical results by 2D and 3D reconstructions on synthetic and real data in positron emission tomography and fluorescence microscopy.

6.1. Fluorescence microscopy

In recent years revolutionary imaging techniques have been developed in light microscopy with enormous importance for biology, material sciences, and medicine. The technology of light microscopy has been considered to be exhausted for a couple of decades, since the resolution is basically limited by Abbe's law for diffraction of light. By developing stimulated emission depletion (STED)- and 4Pi-microscopy [45] now resolutions are achieved that are way beyond this diffraction barrier [51, 44]. STED-microscopy [84] takes an interesting laser sampling approach, which in principle would even allow molecular resolutions. Fluorescent dyes are stimulated by a small laser spot and are directly quenched by an additional interfering laser spot. Since this depletion spot vanishes at one very small point in the middle, fluorescence of the simulating spot is only detected at this tiny position. Hence, data with previously unknown resolution can be measured. However, by reaching the diffraction limit of light, measurements suffer from blurring effects and in addition suffer from Poisson noise due to laser sampling.

In the case of optical nanoscopy the linear, compact operator \bar{K} describes a convolution operator with a kernel $\bar{k} \in C(\Omega \subset \mathbb{R}^d)$,

$$(\bar{K}u)(x) = (\bar{k} * u)(x) := \int_{\Omega} \bar{k}(x-y) u(y) dy. \quad (112)$$

The kernel is often referred to as the point spread function (PSF), whose Fourier transform is called object transfer function. From a computational point of view, it is important to say that the convolution operator in the proposed algorithms can be computed efficiently by FFT following the Fourier convolution theorem,

$$\bar{k} * u = \mathcal{F}^{-1}(\mathcal{F}(\bar{k})\mathcal{F}(u)).$$

To get an impression of images suffering from blurring effects and Poisson noise, we refer to Figure 1. Exemplary, we can see a synthetic data set in Figure 1(a) regarding a special 4Pi convolution kernel, which is illustrated in Figure 1(d) and 1(e). Compared to standard convolution kernels, e.g. of Gaussian type, the latter one bears an additional challenge since it varies considerably in structure. This leads to side lobe effects in the object structure of the measured data as we can see in Figure 1(d) and 1(e). In practice, this type of convolution can be found for instance in 4Pi microscopy [45], since two laser beams interfere in the focus. Under certain circumstances, convolution kernels can also be locally varying, such that blind deconvolution strategies are in need. In this section we assume a 4Pi convolution kernel of the form

$$\bar{k}(x, y) \sim \cos^4\left(\left(\frac{2\pi}{\lambda}\right)y\right) e^{-\left(\frac{x}{\sigma_x}\right)^2 - \left(\frac{y}{\sigma_y}\right)^2}, \quad (113)$$

with the standard deviations σ_x and σ_y , and where λ denotes the refractive index characterizing the doubling properties.

6.2. 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. [83, 80, 6]). 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 depict 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 be 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 [62, Sect. 3.2]. The X-ray transform maps a function on \mathbb{R}^d into the set of its line integrals [62, 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 [62, Sect. 2.1]. If $\theta \in S^{d-1}$ and $s \in \mathbb{R}$, then the Radon transform can be defined by

$$(\bar{K}u)(\theta, s) = \int_{x \cdot \theta = s} u(x) dx = \int_{\theta^\perp} u(s\theta + y) dy ,$$

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 .

6.3. 2D synthetic results in fluorescence microscopy

At the beginning we will illustrate our proposed techniques using synthetic 2D fluorescence microscopy data $f \in \mathbb{R}^{200 \times 200}$ (see Figure 1(c)) simulated for a simple object $\bar{u} \in \mathbb{R}^{200 \times 200}$ (see Figure 1(a)). The data are obtained via a 4Pi convolution kernel (113) presented in Figure 1(d) and 1(e), where additionally Poisson noise is simulated.

In Figure 2 we present EM reconstructions for different numbers of iterations following algorithm (20) with data f illustrated in Figure 1(c). We can observe that early stopping in Figure 2(a) leads to a natural regularization, however with blurring effects and undesired side lobes in the whole object. A higher number of iterations leads to sharper results, as in Figure 2(b), however the reconstructions suffer more and more from the undesired "checkerboard effect", as in Figure 2(c). In Figure 2(d) we additionally display the expected monotone descent of the objective functional in (18) for 100 EM iterations. Finally, we present in Figure 2(e) the typical behavior of EM iterates for ill-posed problems as described in [68]. Namely, the (metric) distance, here Kullback-Leibler, between the iterates and the exact solution decreases initially before it increases as the noise is amplified during the iteration process. The minimal distance in Figure 2(e) is reached approximately after 50 iterations.

In Figure 3 we illustrate reconstruction results obtained with the FB-EM-TV algorithm using different regularization parameters α . In comparison to the EM reconstructions in Figure 2, the regularized EM algorithm deconvolves the given data without remaining side lobes and reduces noise and oscillations very well.

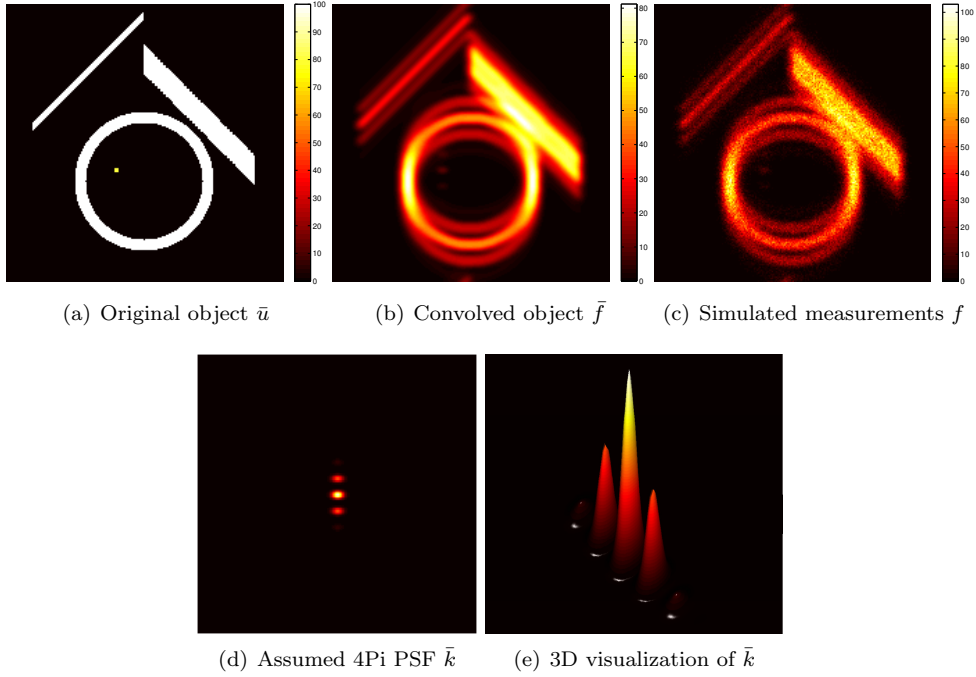


Figure 1. Synthetic 2D fluorescence microscopy data using 4Pi PSF and Poisson noise. (a) Exact object (ground truth) $\bar{u} \in \mathbb{R}^{200 \times 200}$. (b) Convolved object $\bar{f} = \bar{K}\bar{u}$ (112) using 4Pi PSF shown in (d). (c) Simulated measurements f , where \bar{f} in (b) is perturbed by Poisson noise. (d) Assumed 4Pi microscopy PSF with parameters $\lambda = 0.12$, $\sigma_x = 0.02$, $\sigma_y = 0.07$ in (113). (e) 3D visualization of the 4Pi PSF in (d).

Additionally, the FB-EM-TV algorithm successfully reconstructs the main geometrical configurations of the desired object in Figure 1(a), despite the low SNR of the given data in Figure 1(c). The reconstruction in Figure 3(a) is slightly under-smoothed, whereas in Figure 3(c) the computed image is over-smoothed. A visually reasonable reconstruction is illustrated in Figure 3(b). Moreover, different statistical results for the FB-EM-TV reconstruction in Figure 3(b) are plotted in Figure 4. As expected, we can observe a decreasing behavior of the stopping rules proposed in Section 3.4, the objective functional values in (22) and Kullback-Leibler distances to the given measurements f and exact image \bar{u} .

In Section 4.4, in particular in Theorem 4.14, we found out that a damping strategy is theoretical required in the FB-EM-TV algorithm in order to attain a monotone descent of the objective functional in (22) during the minimization process and hence to guarantee convergence of the splitting scheme. However, in numerical tests we could observe that the damping strategy is only needed in the case of high values of the regularization parameter α . For instance we compare the behavior of the objective functional values for different damping parameters in the case of $\alpha = 10$ in Figure 5. Without damping (i.e. $\omega_k = 1$ for all $k \geq 0$) we obtain oscillations in Figure 5(a). These oscillations decrease in the case of a small damping (i.e. in the case of ω_k less one) as plotted in Figure 5(b) and vanish if the damping is strong

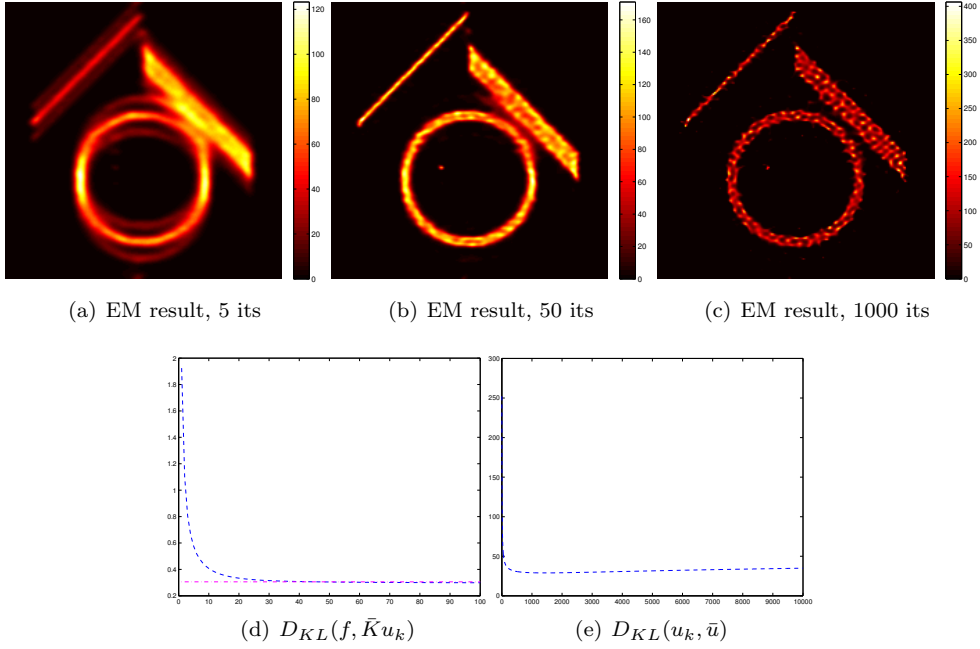


Figure 2. Synthetic 2D fluorescence microscopy data from Figure 1: EM reconstructions. (a)-(c) Reconstruction results obtained with the EM algorithm (20) and stopped at different iteration numbers. (d) Kullback-Leibler distances D_{KL} between given measurements f and convolved EM iterates $\bar{K}u_k$ for 100 iterations (blue dashed line), as well as between f and exact convolved object $\bar{K}\bar{u}$ (magenta dash-dot line). (e) Kullback-Leibler distance between EM iterates u_k and exact object \bar{u} for 10000 iterations.

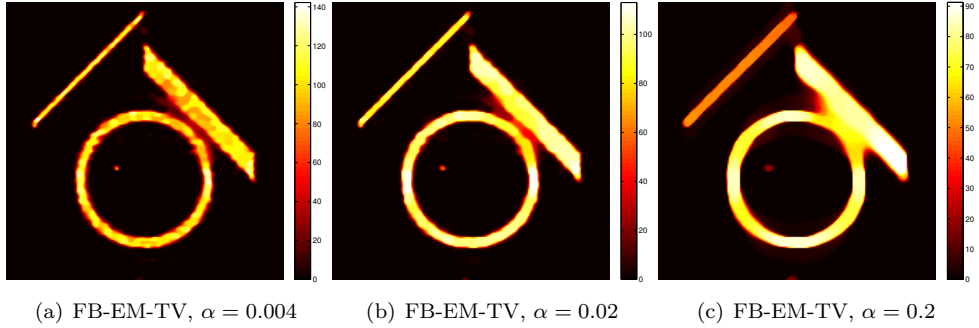


Figure 3. Synthetic 2D fluorescence microscopy data from Figure 1: FB-EM-TV reconstructions. (a)-(c) Reconstruction results obtained with the FB-EM-TV splitting algorithm (29) using different regularization parameters α .

enough, such that a monotone descent in the objective functional can be achieved for $\omega_k = 0.05$ for all $k \geq 0$ in Figure 5(c).

Although the desired object in Figure 1(a) can be reconstructed quite well with the FB-EM-TV algorithm (see Figure 3(b)), we can observe a natural loss of contrast

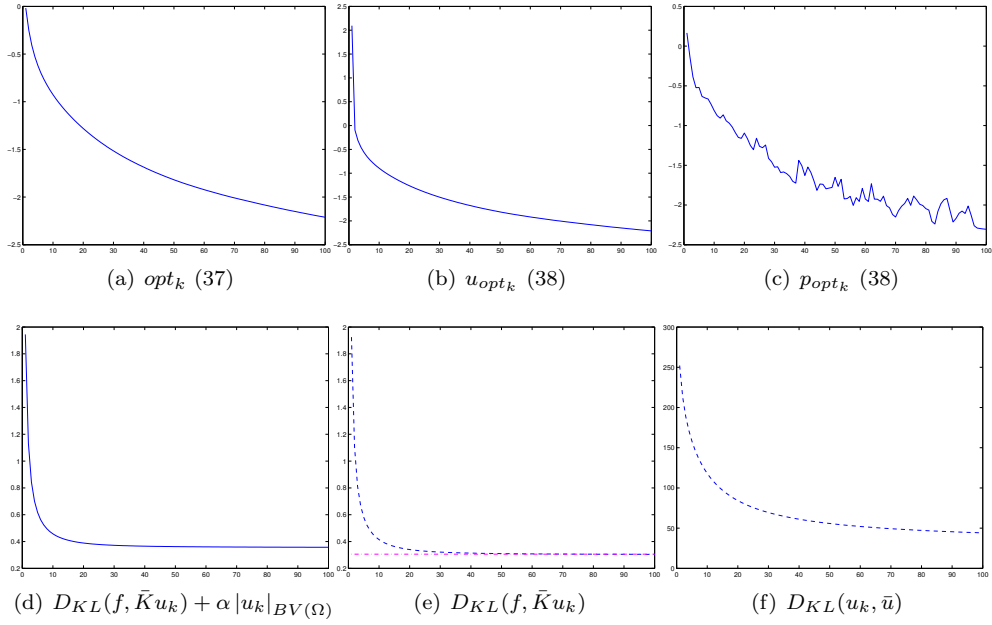


Figure 4. Synthetic 2D fluorescence microscopy data from Figure 1: different statistics for the result in Figure 3(b) with 100 FB-EM-TV iterations. (a)-(c) Stopping rules proposed in Section 3.4. (d) Values of the objective functional in (22). (e) Kullback-Leibler distances D_{KL} between given measurements f and convolved FB-EM-TV iterates $\bar{K}u_k$ (blue dashed line), as well as between f and exact convolved object $\bar{K}\bar{u}$ (magenta dash-dot line). (f) Kullback-Leibler distances between FB-EM-TV iterates u_k and exact object \bar{u} .

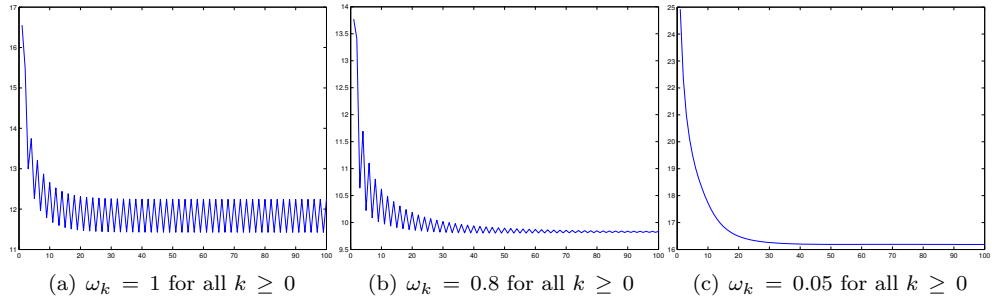


Figure 5. Synthetic 2D fluorescence microscopy data from Figure 1: influence of damping parameters in the FB-EM-TV algorithm with respect to the behavior of the objective functional values. (a)-(c) Iterations vs. values of the objective functional in (22) for different damping parameter values ω_k for $\alpha = 10$.

as mentioned in Section 3.5. This means that some parts of the test object cannot be separated sufficiently. To overcome this problem we proposed to use inverse scale space methods based on Bregman distance iteration (see Section 3.5). In Figure 6 we present reconstruction results for different refinement steps of the Bregman-FB-EM-TV algorithm proposed in Section 3.5. Corresponding to the characteristic of inverse

scale space methods, we observe that the results will be improved with increasing Bregman iteration number with respect to the contrast enhancement, as we can see in the maximal intensity of reconstructions in Figure 6. Moreover, in Figure 6(d) we plot the expected monotone descent of the objective functional in (39) for 4 outer Bregman refinement steps with always 100 inner FB-EM-TV iteration steps. In particular, we can observe the occurring jumps in the functional values which correspond to the contrast refinement effects at each Bregman step. Finally, we present in Figure 6(e) the decreasing behavior of the Kullback-Leibler distances to the exact image \bar{u} for the complete Bregman-FB-EM-TV iteration sequence.

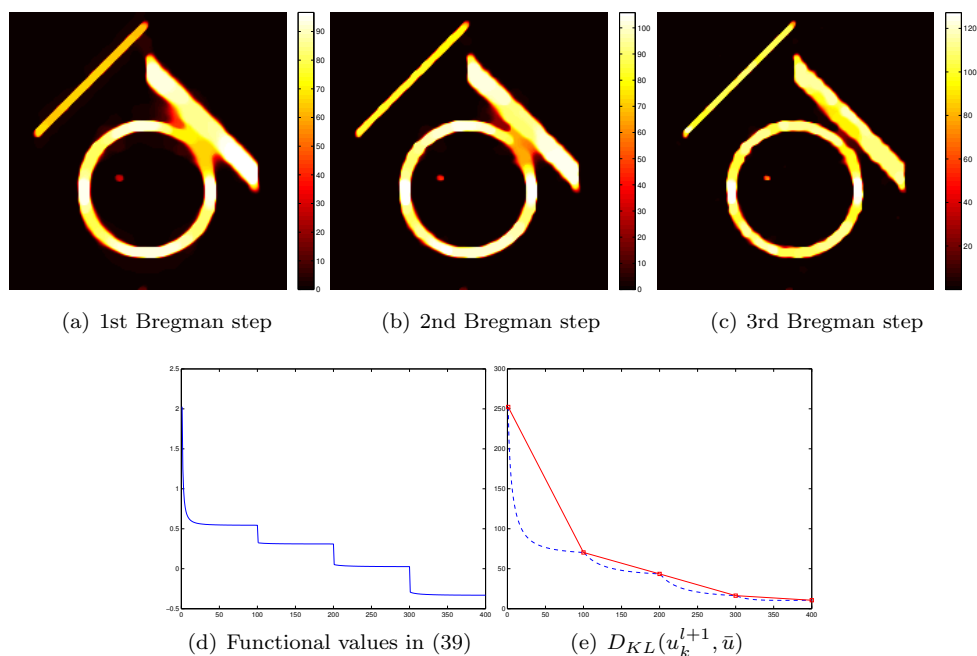


Figure 6. Synthetic 2D fluorescence microscopy data from Figure 1: Bregman-FB-EM-TV reconstructions. (a)-(c) Reconstruction results at different contrast enhancement steps of the Bregman-FB-EM-TV algorithm proposed in Section 3.5. (d) Values of the objective functional in (39) for 4 outer Bregman refinement steps with always 100 inner FB-EM-TV iteration steps. (e) Kullback-Leibler distances between Bregman-FB-EM-TV iterates u_k^{l+1} and exact object \bar{u} for 4 outer Bregman refinement steps with always 100 inner FB-EM-TV iteration steps (blue dashed line), as well as between the final results of each Bregman refinement step and exact object \bar{u} (red line).

6.4. 2D real data results in fluorescence microscopy

In Figure 7 we illustrate the proposed techniques by reconstructing syntaxin clusters [84], a membrane integrated protein participating in exocytosis. Here, the contrast enhancing property of the Bregman-FB-EM-TV algorithm in Figure 7(d) is observable as well, compared to the FB-EM-TV result in Figure 7(c). In particular, it is possible to preserve fine structures in the image.

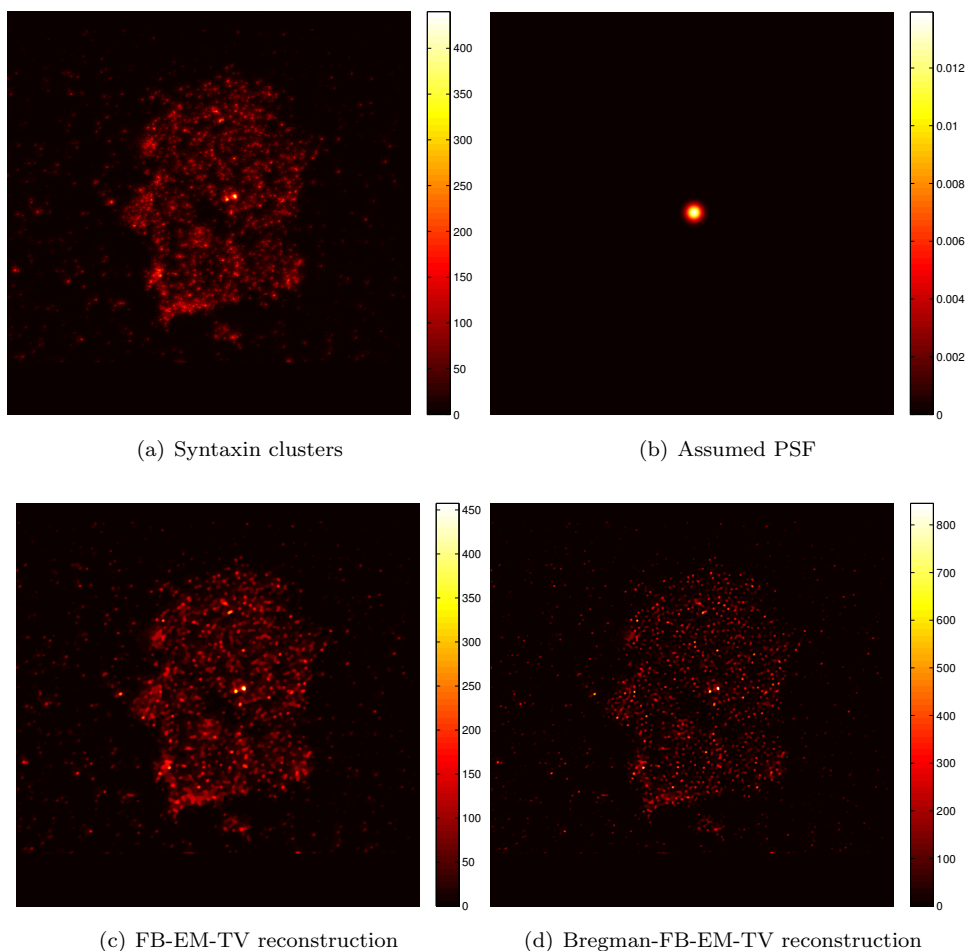


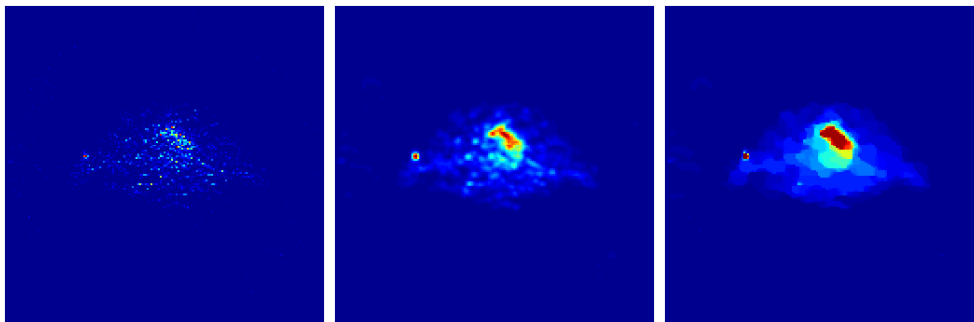
Figure 7. Immunofluorescence CW-STED microscopy measurements: reconstruction results obtained with the (Bregman-)FB-EM-TV algorithms. (a) CW-STED micrograph of protein syntaxin on a membrane sheet of a fixed mammalian (PC12) cell [84], image size 1000×1000 . (b) Assumed point spread function (PSF) for the reconstruction process. (c) Reconstruction with the FB-EM-TV algorithm (29). (d) Reconstruction with the Bregman-FB-EM-TV algorithm proposed in Section 3.5, here 3rd Bregman refinement step.

6.5. 2D real data results in PET

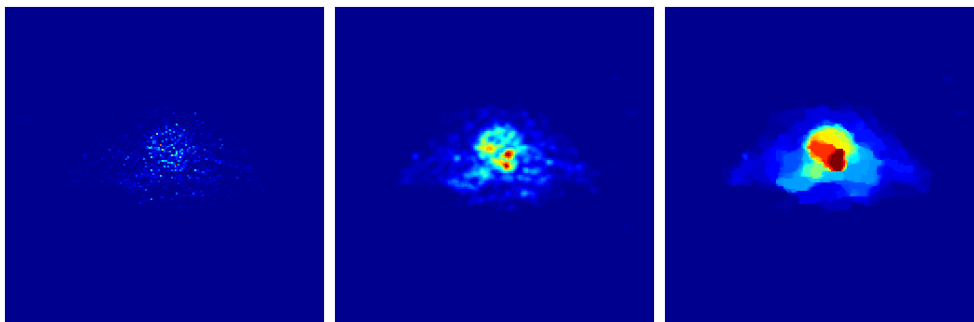
In Figure 8 we illustrate the performance of the FB-EM-TV algorithm by evaluation of cardiac $H_2^{15}O$ measurements obtained with positron emission tomography. This tracer is used in nuclear medicine for the quantification of myocardial blood flow [72]. However, this quantification needs a segmentation of myocardial tissue, left and right ventricle [72, 13], 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 Figure 8(a), whereby the tracer intensity in the left ventricle is presented in Figure 8(b), using measurements

25 seconds and 45 seconds after tracer injection in the blood circulation respectively. To illustrate the SNR problem, we present reconstructions with the classical EM algorithm (20) in Figure 8 (left). As expected, the results suffer from unsatisfactory quality and are impossible to interpret. Hence, we take EM reconstructions with Gaussian smoothing (Figure 8 (middle)) as references. The results in Figure 8 (right) show the reconstructions with the proposed FB-EM-TV algorithm (29). 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 [13].



(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)

Figure 8. Cardiac H_2^{15}O PET measurements: tracer intensity 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.

6.6. 3D real data results in PET

In this section we present 3D reconstruction results using cardiac ^{18}F -FDG measurements obtained with PET. The measurements and the corresponding 3D EM algorithm for the reconstruction process were provided by K. Schäfers and T. Kösters (EIMI, WWU Münster). For the sake of completeness, we note that the size of the sinograms (data domain) is $192 \times 192 \times 47$ and the image size is $175 \times 175 \times 47$. In the following we use a different numerical scheme as in Section 5 in order to solve the weighted ROF problem (103), namely an augmented Lagrangian method described in

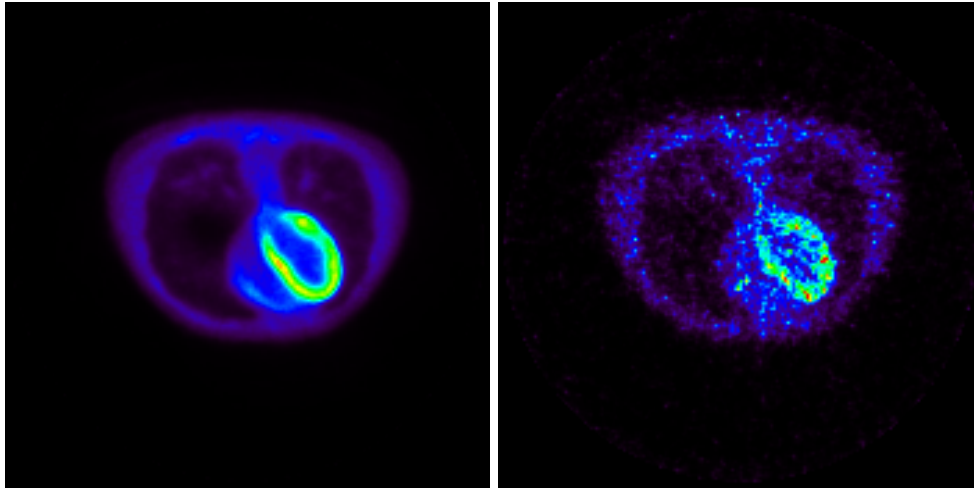
[71] which is a modified version of the well known alternating split Bregman algorithm. We also note that the inversion of a Laplace operator equation is required in this approach, which can be solved efficiently using the discrete cosine transform. To realize this step, we use a MATLAB implementation of the multidimensional (inverse) discrete cosine transform of A. Myronenko [60].

The ^{18}F -FDG tracer is an important radiopharmaceutical in nuclear medicine and is used for measuring glucose 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 Figure 9 (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 Figure 9 (right).

In Figure 10 we show reconstruction results obtained from the FB-EM-TV algorithm (left) and its extension via Bregman distance regularization (right) using measurements after 5 seconds acquisition time of the data. Thereby, 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 Figure 11, where we present scaled versions of both reconstructions in order to allow a quantitative comparison. In Figure 11, the reconstructions from Figure 10 are scaled to the maximum intensity of the EM result in Figure 9 (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.

7. Summary and conclusions

In this paper we have derived novel image reconstruction methods for inverse problems with data corrupted by Poisson noise. In particular, we concentrated on deblurring problems in optical nanoscopy and reconstruction problems in positron emission tomography. Motivated by a statistical modeling of such reconstruction problems we developed an accurate, robust, and reasonably fast FB-EM-TV splitting method. By combining the classical EM algorithm with a simultaneous TV regularization we can reconstruct cartoon-like images with sharp edges, which yield a reasonable basis for quantitative investigations. The main advantages of our splitting strategy is that it enables a high flexibility, can be performed well for large regularization parameters and is also favourably applicable for problems with a low SNR. Moreover, we could show the positivity preservation of the proposed algorithm and provided a detailed analysis of the TV regularized Poisson likelihood estimation problem including a proof of the well-posedness. By interpreting the alternating FB-EM-TV algorithm in terms of convex splitting strategies, we were able to give a proof of convergence. Finally, to overcome the problem of contrast reduction caused by TV regularization, we extended the Poisson likelihood estimation problem and the FB-EM-TV algorithm to an iterative regularization strategy using inverse scale space methods based on Bregman distance iteration. The results presented in this paper significantly improve the ones previously obtained with total variation regularization and rise to hope for further practical use of this approach, in particular in PET and optical nanoscopy.



(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



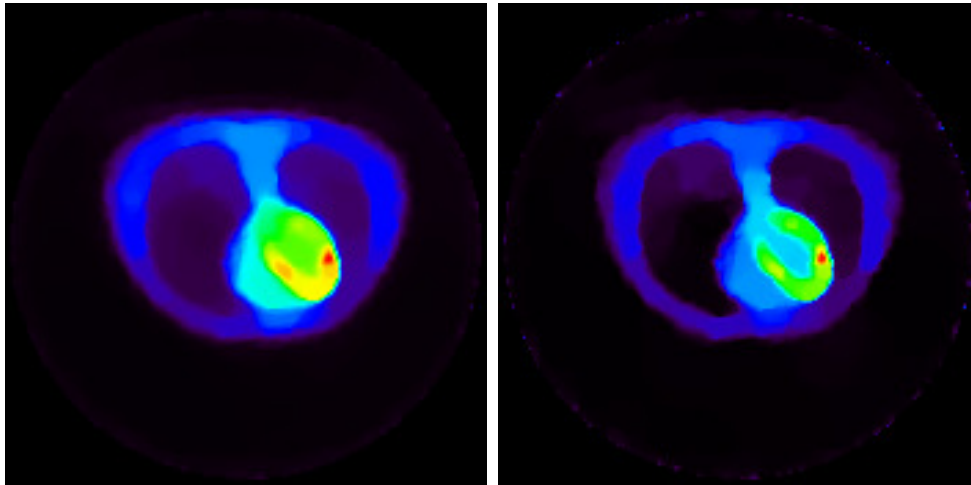
(c) Sagittal view: 20 minutes (left) and 5 seconds (right) data acquisition time

Figure 9. Cardiac ^{18}F -FDG 3D PET measurements: tracer intensity results obtained with the EM algorithm (20) 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.

We finally mention that our approach, namely the FB-EM-TV splitting algorithm as well as the Bregman distance iteration, can be carried out in a completely analogous way for arbitrary convex regularization functionals (see [71]), which might become of further practical importance in the future.

Acknowledgments

This work has been supported by the German Ministry of Education and Research (BMBF) through the project *INVERS: Deconvolution problems with sparsity constraints in nanoscopy and mass spectrometry*. This research was performed when the first author was with the Mathematical Imaging Group at WWU Münster. C. Brune acknowledges further support by the Deutsche Telekom Foundation. M.



(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)



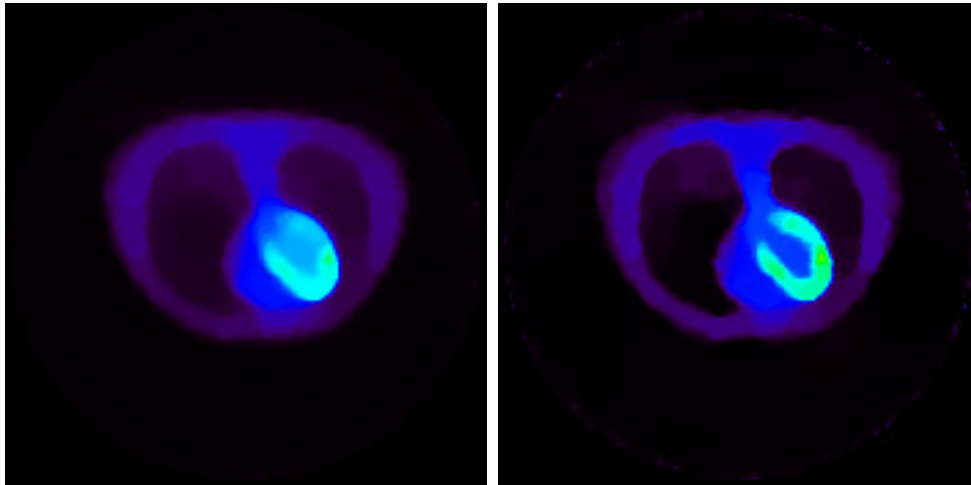
(c) Sagittal view: FB-EM-TV reconstruction (left) and Bregman-FB-EM-TV result (right)

Figure 10. 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 (29), 20 iterations. **Right:** Reconstruction with the Bregman-FB-EM-TV algorithm proposed in Section 3.5 at 6th refinement step.

Burger has been further supported by the German Science Foundation DFG through the SFB 656 *Molecular Cardiovascular Imaging* and the project *Regularization with singular energies*. The authors thank Klaus Schäfers (European Institute for Molecular Imaging (EIMI), WWU Münster) for providing real data in PET. Furthermore, the authors thank Katrin Willig and Andreas Schönle (both MPI Biophysical Chemistry, Göttingen) for providing real data in optical nanoscopy.

References

- [1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10(6):1217–1229, 1994.
- [2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press, 2000.
- [3] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential*



(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)

Figure 11. 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 Figure 10 are scaled to the maximum intensity of ground truth in Figure 9 (left).

Equations and the Calculus of Variations, volume 147 of *Applied Mathematical Sciences*. Springer, 2002.

[4] J.-F. Aujol. Some first-order algorithms for total variation based image restoration. *J. Math. Imaging Vis.*, 34(3):307–327, 2009.

[5] M. Bachmayr. Iterative total variation methods for nonlinear inverse problems. Master’s thesis, Johannes Kepler University, Linz, 2007.

[6] D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey, editors. *Positron Emission Tomography: Basic Sciences*. Springer, 2005.

[7] J. M. Bardsley. An efficient computational method for total variation-penalized Poisson likelihood estimation. *Inverse Problems and Imaging*, 2(2):167–185, 2008.

[8] J. M. Bardsley. A theoretical framework for the regularization of Poisson likelihood estimation problems. *Inverse Problems and Imaging*, 4(1):11–17, 2010.

[9] J. M. Bardsley and J. Goldes. Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation. *Inverse Problems*, 25(9):095005, 2009.

[10] J. M. Bardsley and N. Laoboul. Tikhonov regularized Poisson likelihood estimation: theoretical justification and a computational method. *Inverse Problems in Science and Engineering*, 16(2):199–215, 2008.

- [11] J. M. Bardsley and N. Laoboul. An analysis of regularization by diffusion for ill-posed Poisson likelihood estimations. *Inverse Problems in Science and Engineering*, 17(4):537–550, 2009.
- [12] J. M. Bardsley and A. Luttmann. Total variation-penalized Poisson likelihood estimation for ill-posed problems. *Adv. Comput. Math.*, 31:35–59, 2009.
- [13] M. Benning, T. Kösters, F. Wübbeling, K. Schäfers, and M. Burger. A nonlinear variational method for improved quantification of myocardial blood flow using dynamic H₂¹⁵O PET. In *Nuclear Science Symposium Conference Record*, pages 4472–4477, 2008.
- [14] B. Berkels, M. Burger, M. Droske, O. Nemitz, and M. Rumpf. Cartoon extraction based on anisotropic image classification. In *Vision, Modeling, and Visualization Proceedings*, pages 293–300, 2006.
- [15] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni. A discrepancy principle for Poisson data. *Inverse Problems*, 26(10):105004, 2010.
- [16] M. Bertero, H. Lanteri, and L. Zanni. Iterative image reconstruction: a point of view. In Y. Censor, M. Jiang, and A. Louis, editors, *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)*, volume 7 of *Publications of the Scuola Normale, CRM series*, pages 37–63, 2008.
- [17] K. Bredies. A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space. *Inverse Problems*, 25(1):015005, 2009.
- [18] 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.*, 7:200–217, 1967.
- [19] C. Brune, A. Sawatzky, and M. Burger. Primal and dual Bregman methods with application to optical nanoscopy. *Int. J. Comput. Vis.*, 92(2):211–229, 2011.
- [20] M. Burger, K. Frick, S. Osher, and O. Scherzer. Inverse total variation flow. *Multiscale Model. Simul.*, 6(2):366–395, 2007.
- [21] M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. *Comm. Math. Sci.*, 4(1):179–212, 2006.
- [22] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20:89–97, 2004.
- [23] A. Chambolle. Total variation minimization and a class of binary MRF models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 3757 of *LNCS*, pages 136–152. Springer, 2005.
- [24] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock. *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, chapter An Introduction to Total Variation for Image Analysis, pages 263–340. De Gruyter, 2010.
- [25] P. Combettes and V. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4:1168–1200, 2005.
- [26] P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6):065014, 2008.
- [27] I. Csiszar. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Statist.*, 19(4):2032–2066, 1991.
- [28] 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(1):1–38, 1977.
- [29] 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. Technical Report 5272, Institut National de Recherche en Informatique et en Automatique, 2004.
- [30] 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(2):421–439, 1956.
- [31] P. P. B. Eggermont. Maximum entropy regularization for Fredholm integral equations of the first kind. *SIAM J. Math. Anal.*, 24(6):1557–1576, 1993.
- [32] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*, volume 1 of *Studies in Mathematics and Its Applications*. North-Holland Publishing Company, 1976.
- [33] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Kluwer Academic Publisher, 2000.
- [34] S. Esedoglu and S. J. Osher. Decomposition of images by the anisotropic Rudin-Osher-Fatemi model. *Comm. Pure Appl. Math.*, 57(12):1609–1626, 2004.
- [35] 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.
- [36] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in

- Advanced Mathematics. CRC Press, 1992.
- [37] 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.
- [38] D. Gabay. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, volume 15 of *Studies in Mathematics and its Applications*, chapter Applications of the method of multipliers to variational inequalities, pages 299–331. Elsevier Science Publishers B.V., Amsterdam, 1983.
- [39] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *J. Appl. Stat.*, 20(5):25–62, 1993.
- [40] S. Geman and D. E. McClure. Bayesian image analysis: an application to single photon emission tomography. *Statistical Computation Section, American Statistical Association*, pages 12–18, 1985.
- [41] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*, volume 80 of *Monographs in Mathematics*. Birkhäuser, 1984.
- [42] T. Goldstein and S. Osher. The split Bregman method for L^1 -regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–343, 2009.
- [43] C. W. Groetsch. *Inverse Problems in the Mathematical Sciences*. Vieweg Verlag, 1993.
- [44] S. Hell and A. Schönle. Nanoscale resolution in far-field fluorescence microscopy. In P. W. Hawkes and J. C. H. Spence, editors, *Science of Microscopy*. Springer, 2006.
- [45] S. Hell and E. H. K. Stelzer. Properties of a 4Pi confocal fluorescence microscope. *J. Opt. Soc. Am. A*, 9(12):2159–2166, 1992.
- [46] S. W. Hell. Toward fluorescence nanoscopy. *Nature Biotechnology*, 21(11):1347–1355, 2003.
- [47] F. M. Henderson and A. J. Lewis. *Principles and Applications of Imaging Radar: Manual of Remote Sensing*, volume 2. Wiley and Sons, 1998.
- [48] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*, volume 305 of *Grundlehren der mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences)*. Springer Verlag, 1993.
- [49] A. N. Iusem. Convergence analysis for a multiplicatively relaxed EM algorithm. *Mathematical Methods in the Applied Sciences*, 14(8):573–593, 1991.
- [50] E. Jonsson, S. C. Huang, and T. Chan. Total variation regularization in positron emission tomography. CAM Report 98-48, UCLA, 1998.
- [51] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *PNAS*, 97(15):8206–8210, 2000.
- [52] T. Le, R. Chartrand, and T. J. Asaki. A variational approach to reconstructing images corrupted by Poisson noise. *J. Math. Imaging Vis.*, 27(3):257–263, 2007.
- [53] H. Liao, F. Li, and M. K. Ng. Selection of regularization parameter in total variation image restoration. *J. Opt. Soc. Am. A*, 26(11):2311–2320, 2009.
- [54] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.
- [55] L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astronomical Journal*, 79:745–754, 1974.
- [56] A. Luttmann. A theoretical analysis of L^1 regularized Poisson likelihood estimation. *Inverse Prob. Sci. Eng.*, 18(2):251–264, 2010.
- [57] R. E. Megginson. *An Introduction to Banach Space Theory*, volume 183 of *Graduate Texts in Mathematics*. Springer, 1998.
- [58] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, volume 22 of *University Lecture Series*. American Mathematical Society, Boston, MA, USA, 2001.
- [59] J. Müller. Parallel total variation minimization. Master’s thesis, Institute for Computational and Applied Mathematics, University of Münster, 2008.
- [60] A. Myronenko, January 2011. <https://sites.google.com/site/myronenko/software>.
- [61] 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.*, 123:236–247, 1998.
- [62] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. SIAM Monographs on Mathematical Modeling and Computation, 2001.
- [63] 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(2):460–489, 2005.
- [64] V. Y. Panin, G. L. Zeng, and G. T. Gullberg. Total variation regulated EM algorithm [SPECT reconstruction]. *IEEE Trans. Nucl. Sci.*, 46(6):2202–2210, 1999.
- [65] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in hilbert spaces.

- J. Math. Anal. Appl.*, 72:383–390, 1979.
- [66] R. Plato. On the discrepancy principle for iterative and parametric methods to solve linear ill-posed equations. *Numer. Math.*, 75(1):99–120, 1996.
- [67] E. Resmerita and R. S. Anderssen. Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems. *Math. Meth. Appl. Sci.*, 30:1527–1544, 2007.
- [68] E. Resmerita, H. W. Engl, and A. N. Iusem. The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis. *Inverse Problems*, 23(6):2575–2588, 2007.
- [69] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62(1):55–59, 1972.
- [70] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60:259–268, 1992.
- [71] A. Sawatzky. *(Nonlocal) Total Variation in Medical Imaging*. PhD thesis, University of Münster, 2011. CAM Report 11-47, UCLA.
- [72] 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(8):1031–1040, 2002.
- [73] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In *Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision*, LNCS 5567, pages 464–476. Springer, 2009.
- [74] S. Setzer. *Splitting Methods in Image Processing*. PhD thesis, University of Mannheim, 2009.
- [75] S. Setzer, G. Steidl, and T. Teuber. Deblurring Poissonian images by split Bregman techniques. *J. Vis. Commun. Image R.*, 21(3):193–199, 2010.
- [76] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, 1(2):113–122, 1982.
- [77] 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, pages 477–489. Springer, 2009.
- [78] 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(1):273–303, 2006.
- [79] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control Optim.*, 29(1):119–138, 1991.
- [80] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *J. Am. Stat. Assoc.*, 80:8–20, 1985.
- [81] L. A. Vese and S. J. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *J. Sci. Comput.*, 19:553–572, 2003.
- [82] C. R. Vogel. *Computational Methods for Inverse Problems*. Frontiers in Applied Mathematics. SIAM, 2002.
- [83] M. N. Wernick and J. N. Aarsvold, editors. *Emission Tomography: The Fundamentals of PET and SPECT*. Elsevier Academic Press, 2004.
- [84] K. I. Willig, B. Harke, R. Medda, and S. W. Hell. STED microscopy with continuous wave beams. *Nature Meth.*, 4(11):915–918, 2007.