Primal and Dual Bregman Methods with Application to Optical Nanoscopy

Christoph Brune · Alex Sawatzky · Martin Burger

Abstract Measurements in nanoscopic imaging suffer from blurring effects modeled with different point spread functions (PSF). Some apparatus even have PSFs that are locally dependent on phase shifts. Additionally, raw data are affected by Poisson noise resulting from laser sampling and "photon counts" in fluorescence microscopy. In these applications standard reconstruction methods (EM, filtered backprojection) deliver unsatisfactory and noisy results. Starting from a statistical modeling in terms of a MAP likelihood estimation we combine the iterative EM algorithm with total variation (TV) regularization techniques to make an efficient use of a-priori information. Typically, TV-based methods deliver reconstructed cartoon images suffering from contrast reduction. We propose extensions to EM-TV, based on Bregman iterations and primal and dual inverse scale space methods, in order to obtain improved imaging results by simultaneous contrast enhancement. Besides further generalizations of the primal and dual scale space methods in terms of general, convex variational regularization methods, we provide error estimates and convergence rates for exact and noisy data. We illustrate the performance of our techniques on synthetic and experimental biological data.

Westfälische Wilhelms-Universität Münster,
Institut für Numerische und Angewandte Mathematik,
Einsteinstr. 62, D-48149 Münster, Germany,
E-mail: {christoph.brune,alex.sawatzky,martin.burger}@wwu.de,
home page: `http://imaging.uni-muenster.de`

1 Introduction

Image reconstruction is a fundamental problem in many fields of applied sciences, e.g. nanoscopic imaging, medical imaging or astronomy. Fluorescence microscopy for example is an important imaging technique for the investigation of biological (live-) cells, up to nano-scale. In this case image reconstruction arises in form of deconvolution problems. Undesirable blurring effects can be ascribed to a diffraction of light.

Mathematically, image reconstruction in such applications can often be formulated as the computation of a function $\tilde{u} \in \mathcal{U}(\Omega)$ from the operator equation

$$K\tilde{u} = g . \tag{1}$$

Typically, this is a Fredholm integral equation of the first kind with given *exact data* $g \in \mathcal{V}(\Sigma)$ and the desired exact solution $\tilde{u}$.

Here, K denotes a linear and compact operator $K : \mathcal{U}(\Omega) \to \mathcal{V}(\Sigma)$ and $\mathcal{U}(\Omega)$ and $\mathcal{V}(\Sigma)$ are Banach spaces of functions on bounded and compact sets Ω respectively Σ. In the case of nanoscopic imaging $\tilde{u}$ is a convolution operator

$$(Ku)(x) = (k * u)(x) = \int_{\Omega} k(x-y)u(y)\,dy ,$$

where k is a convolution kernel, describing the blurring effects caused by a nanoscopic apparatus.

Since K cannot be inverted continuously (due to the compactness of the forward operator), most inverse problems are ill-posed. Furthermore, in real-life applications the exact data g are usually not available, but a noisy version f instead. Hence, we need to compute approximations to the ill-posed problem

$$Ku = f \tag{2}$$

instead of (1), with $u \in \mathcal{U}(\Omega)$ and $f \in \mathcal{V}(\Sigma)$.

Since direct inversion of K is not suitable, regularization techniques are needed to produce reasonable reconstructions. A frequently used way to realize the latter is the Bayesian model, whose aim is the computation of an estimate u of the unknown object by maximizing the a-posteriori probability density $p(u|f)$ with measurements f. The latter is given according to Bayes formula

$$p(u|f) = \frac{p(f|u)\,p(u)}{p(f)} \ . \tag{3}$$

This approach is called maximum a-posteriori probability (MAP) estimation. If the measurements $\tilde{f}$ are given, we describe the density $p(u|f)$ as the a-posteriori likelihood function which depends on u only. The Bayesian approach (3) has the advantage that it allows to incorporate additional information about u via the prior probability density $p(u)$ into the reconstruction process. The most frequently used prior densities are Gibbs functions

$$p(u) \sim e^{-\alpha\,J(u)} \ , \tag{4}$$

where α is a positive parameter and J a convex regularization energy $J : \mathcal{W}(\Omega) \to \mathbb{R} \cup \{\infty\}$. Since stochastic modeling is often done in discrete terms based on the modeling of random variables, we introduce a semi-discrete, linear and compact operator

$$\overline{K} : \mathcal{U}(\Omega) \to \mathcal{D}(\Sigma) \, , \tag{5}$$

with a finite-dimensional range $\mathcal{D}(\Sigma)$, to be able to derive corresponding continuum models. Typical models for the probability density $p(f|u)$ in (3) are Gaussian-, Poisson- or Multiplicative-distributed data f, i.e.

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

$$p(f|u) = \prod_i \frac{(\overline{K}u)_i^{f_i}}{f_i!} e^{-(\overline{K}u)_i} \qquad \text{(Poisson)} \tag{6}$$

$$p(f|u) = \prod_i \frac{n^n}{(\overline{K}u)_i^n\,\Gamma(n)} f_i^{n-1} e^{-n\frac{f_i}{(\overline{K}u)_i}} \ \text{(Multipl.)},$$

where $\bar{K}$ is a semi-discrete operator derived from sampling K.

Most works deal with the case of Gaussian distributed noise so far. However, in real-life there are several applications, in which different types of noise are of a certain interest, such as Positron Emission Tomography (PET), Microscopy, CCD cameras, or radar. For instance, in addition to fluorescence microscopy, Poisson noise appears also in PET in medical imaging. Other non-Gaussian noise models are salt and pepper noise or

the different types of multiplicative noise, for example appearing in Synthetic Aperture Radar (SAR) imaging to reduce speckle noise. In [AA08], Aubert and Aujol assumed η in $f = (Ku)\eta$ to follow a gamma law with mean one and derived the conditional probability $p(f|u)$ above. For such cases different variational models (fidelities related to the log-likelihood of the noise distribution) can be derived in the framework of MAP estimation, which need different analysis than in the case of Gaussian distributed noise.

In the canonical case of additive Gaussian noise (see (6)), the minimization of the negative log likelihood function (3) leads to classical Tikhonov regularization [BLZ08] based on minimizing a functional of the form

$$\min_{u \geq 0} \left\{ \frac{1}{2}\,\|Ku - f\|_{L^2(\Sigma)}^2 \,+\, \alpha\,J(u) \right\} \ . \tag{7}$$

The first term, the so-called data-fidelity term, penalizes the deviation from equality in (1) whereas R is a regularization term as in (4). If we choose $K = Id$ and the total variation (TV) regularization technique $J(u) := |u|_{BV}$, we obtain the well-known ROF-model [ROF92] for image denoising. The additional positivity constraint is necessary in typical applications as the unknown represents a density image.

In nanoscopic imaging measured data are stochastic and pointwise, more precisely, the data are called "photon counts". This property refers to laser scanning techniques in fluorescence microscopy. Consequently, the random variables of measured data are not Gaussian- but Poisson-distributed (see (6)), with expected value given by $(\overline{K}u)_i$. Hence a MAP estimation via the negative log likelihood function (3) leads to the following variational problem [BLZ08]

$$\min_{u \geq 0} \left\{ \int_\Sigma (Ku - f \log Ku)\,d\mu \,+\, \alpha\,J(u) \right\} \ . \tag{8}$$

Up to additive terms independent of u, the data-fidelity term is the so-called Kullback-Leibler functional (also known as cross entropy or I-divergence) between the two probability measures f and Ku. A particular complication of (8) compared to (7) is the strong nonlinearity in the data fidelity term and resulting issues in the computation of minimizers.

The specific choice of the regularization functional R in (8) is important for how a-priori information about the expected solution is incorporated into the reconstruction process. Smooth, in particular quadratic regularizations have attracted most attention in the past, mainly due to the simplicity in analysis and computation. However, such regularization approaches always lead to blurring of the reconstructions, in particular they cannot yield reconstructions with sharp edges.

Recently, singular regularization energies, in particular those of ℓ^1 or L^1-type, have attracted strong attention. In this work, we introduce an approach which uses total variation (TV) as the regularization functional. TV regularization has been derived as a denoising technique in [ROF92] and generalized to various other imaging tasks subsequently. The exact definition of TV [AV94], used in this paper, is

$$J(u) := |u|_{BV} = \sup_{\substack{g \in C_0^\infty(\Omega;\mathbb{R}^d) \\ ||g||_\infty \leq 1}} \int_\Omega u \, \mathrm{div}\, g \;, \qquad (9)$$

which is formally (true if u is sufficiently regular) $|u|_{BV} = \int_\Omega |\nabla u|$. 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 nanoscopic imaging if the goal is to identify object shapes that are separated by sharp edges and shall be analyzed quantitatively.

Unfortunately, images reconstructed by methods using TV regularization suffer from loosing contrast. In this paper, we suggest to extend EM-TV by iterative regularization to Bregman-EM-TV, attaining simultaneous contrast enhancement. More precisely, we apply total variation inverse scale space methods by employing the concept of Bregman distance regularization. The latter has been derived in [OBG⁺05] with a detailed analysis for Gaussian-type problems (7). Furthermore, it has been generalized to time-continuity [BGO⁺06] and L^p-norm data fitting terms [BFO⁺07]. Here, in the case of Poisson-type problems, the method consists in computing a minimizer u^1 of (8) with $J(u) := |u|_{BV}$ first. Updates are determined successively by computing

$$u^{l+1} = \arg\min_{u \in BV(\Omega)} \left\{ \int_\Sigma (Ku - f \log Ku) \, d\mu \right.$$
$$\left. + \; \alpha \left(|u|_{BV} - \langle p^l, u \rangle \right) \right\} \;, \quad (10)$$

with p^l as an element of the subgradient of the total variation semi norm in u^l. Introducing the Bregman distance with respect to $|\cdot|_{BV}$ defined via

$$D^p_{|\cdot|_{BV}}(u,v) = |u|_{BV} - |v|_{BV} - \langle p, u - v \rangle \qquad (11)$$

with $p \in \partial |v|_{BV} \subseteq BV^*(\Omega)$, where $\langle \cdot, \cdot \rangle$ denotes the duality product, allows to characterize u^{l+1} in (10) as

$$u^{l+1} = \arg\min_{u \in BV(\Omega)} \left\{ \int_\Sigma (Ku - f \log Ku) \, d\mu \right.$$
$$\left. + \; \alpha \, D^{p^l}_{|\cdot|_{BV}}(u, u^l) \right\} \;, \quad (12)$$

with an appropriate update formula for u^l. In the case of a Gaussian-noise model with corresponding L^2 data

fidelity, i.e. (7), the iterative Bregman regularization strategy reads as follows,

$$u^{l+1} = \arg\min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \left\| Ku - v^l - f \right\|^2_{L^2(\Sigma)} \right.$$
$$\left. + \alpha \, D^{p^l}_{|\cdot|_{BV}}(u, u^l) \right\} \;, \quad (13)$$

with the update formula

$$v^{l+1} = v^l - (Ku^{l+1} - f) \qquad v^0 = 0 \;,$$

using the setting $p^l = \frac{1}{\alpha} K^* v^l \in \partial J(u^l)$. To derive the dual counterpart of the primal inverse scale space strategy in (12), one can take the dual formulation of (8) and replace the resulting dual regularization term by the corresponding iterative Bregman regularization. This sequence yields a dual inverse scale space strategy. Interestingly, the once more bidual formulation reads as follows

$$u^{l+1} = \arg\min_{u \in \mathcal{BV}(\Omega)} \left\{ \int_\Sigma (Ku + r^l - f \log(Ku + r^l)) \, d\mu \right.$$
$$\left. + \; \alpha \, |u|_{BV} \right\} \;,$$

with the update of the residual function r^l

$$r^{l+1} = r^l + Ku^{l+1} - f \;, \qquad r^0 = 0 \;.$$

In comparison to (13) this dual scale space method also has a very simple structure and a nice interpretation in terms of a dynamically changing background model based on the residual function r^l. In this work we will see, that primal and dual inverse scale space strategies can noticeably improve reconstructions for inverse problems with Poisson statistics like optical nanoscopy.

In the following section we will study primal and dual inverse scale space methods for general, convex variational regularization methods based on Bregman distances. In the dual perspective, we are going to provide general error estimates and convergence rates for the cases of exact and noisy data. In Section 3 we will apply the proposed scale space strategies to Poisson noise modeling and to regularization with TV. Particularly, we will classify the latter in context of EM-TV based reconstruction methods. Subsequently we will illustrate the performance of the proposed techniques by application to synthetic and experimental optical nanoscopy data. In the last section we will conclude and formulate open questions.

2 Inverse Scale Space Methods

In the following we present primal and dual inverse scale space strategies for solving inverse problems such as (2) described in the introduction. These techniques are based on iterative Bregman distance regularization for general, convex functionals and arise in the over-smoothed limit. From a dual view point of the variational model we also derive a dual inverse scale space flow, which coincides with the primal one in the case of the Gaussian noise case. In more general cases of fidelities, the dual flow appears to be easier with respect to analysis and even allows us to derive error estimates. In the course of this work we use the following definitions:

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

denotes a linear and compact operator where $\mathcal{U}(\Omega)$ and $\mathcal{V}(\Sigma)$ are Banach spaces of functions on bounded and compact sets Ω respectively Σ.

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

is a convex data fidelity using the operator K. In order to guarantee that the data fidelity is centered at zero, we use $Ku - f$ as the argument, i.e. $H_f(Ku - f) = 0$ if $Ku = f$. This notation is particularly useful for duality arguments. Moreover

$$J : \mathcal{W}(\Omega) \subset \mathcal{U}(\Omega) \to \mathbb{R} \cup \{\infty\}$$

denotes a convex regularization functional. Furthermore, we call g *exact data* and f *noisy data* with a given noise estimate

$$H_f(g) \leq \delta . \tag{14}$$

A major step for error estimates and multi-scale techniques in the case of regularization with singular energies has been the introduction of (generalized) Bregman distances (cf. [B67,K97]) as an error measure (cf. [BO04]). The Bregman distance for general convex, not necessarily differentiable functionals, is defined as follows.

Definition 1 (Bregman Distance) Let $\mathcal{U}$ be a Banach space and $J : \mathcal{U} \to \mathbb{R} \cup \{\infty\}$ be a convex functional with non-empty subdifferential ∂J. Then, the Bregman distance is defined as

$$D_J^{\partial J(v)}(u, v)$$
$$:= \{J(u) - J(v) - \langle p, u - v \rangle_{\mathcal{U}} \mid p \in \partial J(v)\} .$$

The Bregman distance for a specific subgradient ζ is defined as $D_J^\zeta : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^+$ with

$$D_J^\zeta(u, v) := J(u) - J(v) - \langle \zeta, u - v \rangle_{\mathcal{U}} , \quad \zeta \in \partial J(v) .$$

Since we are dealing with duality throughout this work, we are going to write

$$\langle a, b \rangle_{\mathcal{U}} := \langle a, b \rangle_{\mathcal{U}^* \times \mathcal{U}} = \langle b, a \rangle_{\mathcal{U} \times \mathcal{U}^*} ,$$

for $a \in \mathcal{U}^*$ and $b \in \mathcal{U}$, as the notation for the dual product, for the sake of simplicity. The Bregman distance is no distance in the usual sense; at least $D_J(u, u) = 0$ and $D_J(u, v) \geq 0$ hold; the latter due to convexity of J. If J is strictly convex, we even obtain $D_J(u, v) > 0$ for $u \neq v$. In general, no triangular inequality nor symmetry holds for the Bregman distance. The latter one can be achieved by introducing the so-called symmetric Bregman distance.

Definition 2 (Symmetric Bregman Distance) Let $\mathcal{U}$ be a Banach space and $J : \mathcal{U} \to \mathbb{R} \cup \{\infty\}$ be a convex functional with non-empty subdifferential ∂J. Then, a symmetric Bregman distance is defined as $D_J^{\mathrm{symm}} : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^+$ with

$$D_J^{\mathrm{symm}}(u_1, u_2) := D_J^{p_1}(u_2, u_1) + D_J^{p_2}(u_1, u_2)$$
$$= \langle u_1 - u_2, p_1 - p_2 \rangle_{\mathcal{U}^*} ,$$

with $\quad p_i \in \partial J(u_i) \quad$ for $\quad i \in \{1, 2\}$.

2.1 Primal Inverse Scale Space Methods

Starting from a general, convex variational problem with data fidelity H_f and regularization functional J, we obtain the standard form:

Problem 1 (Variational Problem)

$$\min_{u \in \mathcal{W}(\Omega)} \{H_f(Ku - f) + \alpha J(u)\} . \tag{15}$$

The corresponding iterative Bregman regularization strategy can be written as

Problem 2 (Inverse Scale Space)

$$u^{l+1} = \arg\min_{u \in \mathcal{W}(\Omega)} \left\{ H_f(Ku - f) + \alpha D_J^{p^l}(u, u^l) \right\}$$
$$= \arg\min_{u \in \mathcal{W}(\Omega)} \left\{ H_f(Ku - f) + \alpha (J(u) - \langle u, p^l \rangle) \right\} , \tag{16}$$

with $p^l \in \partial J(u^l)$.

The first-order optimality condition of this Bregman-regularized functional reads as follows,

$$\alpha(p^{l+1} - p^l) = -K^*(\partial H_f(Ku^{l+1} - f)), \text{ with } p^l \in \partial J(u^l),$$

and basically provides an update rule for p^l. In the limit $\frac{1}{\alpha} \downarrow 0$, the latter can be interpreted as a forward Euler discretization of the flow

$$\frac{d}{dt} p(t) = K^*(\partial H_f(Ku(t) - f)), \qquad p(0) = 0 \in \partial J(u(0)) ,$$

which has been termed *nonlinear inverse scale space method* (cf. [BRH07], [BFO$^+$07], [BGO$^+$06]). The terminology inverse scale space method is due to the fact that this approach somehow behaves in an inverse way to the popular scale space methods (cf. [SG01], [W83], [PM90]). In the case of inverse problems with Gaussian noise modeling, i.e. L^2 data fidelity, inverse scale space strategies have been well studied and error estimates could be attained (cf. [BRH07]). Unfortunately, in the general case above, the residuals $Ku(t) - f$ are enclosed by the derivative of the data fidelity. Unlike the special case of a L^2 data fidelity, this aspect leads to mathematical difficulties if you want to establish error estimates respectively convergence rates of the scale space method, since we need to invert ∂H_f enclosed by K^*. A different way to see the issues is to write the inverse scale-space method as

$$\partial H_f^* \left((K^*)^{-1} \frac{d}{dt} p(t) \right) = (Ku(t) - f),$$
$$\text{with } p(0) = 0 \in \partial J(u(0)) \ ,$$

which is a strongly nonlinear equation for the dual variable. We are able to overcome these difficulties in the following section by using an alternative dual scale space strategy.

2.2 Dual Inverse Scale Space Methods

In this subsection we are going to derive a dual inverse scale space method in terms of an iterative Bregman regularization of a dual reconstruction functional. Fortunately, it is possible to derive error estimates and convergence rates of the corresponding dual inverse scale space flow.

In order to derive the dual formulation of the Bregman regularization functional in (16), we use the zero centered data fidelity $H_f(\cdot)$ and introduce the convex conjugates

$$H_f^*(q) = \sup_{v \in \mathcal{V}(\Sigma)} \left(\langle q, v \rangle_{\mathcal{V}(\Sigma)} - H_f(v) \right)$$
$$J^*(p) = \sup_{u \in \mathcal{W}(\Omega)} \left(\langle p, u \rangle_{\mathcal{U}(\Omega)} - J(u) \right) \ .$$

Under appropriate conditions, the Fenchel duality theorem (cf. [ET99]) implies the following primal-dual relation.

$$\inf_{u \in \mathcal{W}(\Omega)} \left\{ H_f(Ku - f) + \alpha \left(J(u) - \langle p^l, u \rangle \right) \right\}$$
$$= \inf_{u,v} \sup_q \left\{ H_f(Ku - f) + \alpha(J(u) - \langle p^l, u \rangle) \right.$$
$$\left. + \langle v - Ku + f, q \rangle \right\}$$
$$= \sup_q \left\{ \inf_v (H_f(v) + \langle v, q \rangle) \right.$$
$$\left. + \alpha \inf_u (J(u) - \langle p^l + \frac{1}{\alpha} K^* q, u \rangle) + \langle f, q \rangle \right\}$$

Using the convex conjugates and $\inf(\cdot) = -\sup(-\cdot)$ we get

$$= \sup_q \left\{ -H_f^*(-q) + \langle f, q \rangle - \alpha J^*(\frac{1}{\alpha} K^* q + p^l) \right\},$$
$$\text{with } \frac{1}{\alpha} K^* q + p^l \in \partial J(u) \subset \mathcal{W}(\Omega)^*$$

Defining $p := \frac{1}{\alpha} K^* q + p^l$, hence $q = \alpha(K^*)^{-1}(p - p^l)$ implies the dual formulation of the (primal) Bregman method above:

Problem 3 (Inverse Scale Space, dual form.)

$$p^{l+1} = \arg\min_{p \in \mathcal{U}(\Omega)^*} \left\{ H_f^*(\alpha(K^*)^{-1}(p^l - p)) \right.$$
$$\left. - \langle f, \alpha(K^*)^{-1}(p - p^l) \rangle + J^*(p) \right\} \ ,$$

with $\alpha(K^*)^{-1}(p^l - p) \in \partial H_f(Ku)$ and $p \in \partial J(u)$.

Now we are going to use the primal-dual relation above to provide a dual iterative Bregman regularization technique. Considering the standard regularized reconstruction model in (15), the described primal-dual relation above, with $p^l = 0$, yields the dual formulation of the variational problem:

Problem 4 (Variational Problem, dual form.)

$$\min_{p \in \mathcal{U}(\Omega)^*} \left\{ \alpha J^*(p) - \langle f, \alpha(K^*)^{-1}p \rangle + H_f^*(-\alpha(K^*)^{-1}p) \right\} .$$

Please consider that the conjugate of J and the duality product act as a fidelity term and the conjugate of H_f as a regularization term in this formulation.

Consequently, the natural dual counterpart of the primal inverse scale space method, using the substitution $q := \alpha(K^*)^{-1}p$ resp. $q^l := \alpha(K^*)^{-1}p^l$, reads as follows

Problem 5 (Dual Inverse Scale Space)

$$p^{l+1} = \operatorname*{arg\,min}_{q \in \mathcal{V}(\Sigma)^*} \left\{ \alpha\, J^*(\frac{1}{\alpha}K^*(q)) - \langle f, -q \rangle \right.$$

$$\left. + D_{H_f^*}^{r^l}(-q, -q^l) \right\}$$

$$= \operatorname*{arg\,min}_{q \in \mathcal{V}(\Sigma)^*} \left\{ \alpha\, J^*(\frac{1}{\alpha}K^*(q)) - \langle f, -q \rangle \right.$$

$$\left. + H_f^*(-q) + \langle r^l, q \rangle \right\}$$

with $r^l \in \partial H_f^*(-q^l)$.

The corresponding dual formulation of this variational problem, i.e. the primal (equal to the bidual) formulation, has a structure we are familiar with. The definition of the convex conjugate and the Fenchel duality theorem under appropriate conditions once more imply the following dual-primal relation:

$$\inf_p \left\{ \alpha J^*(p) - \langle f, -\alpha(K^*)^{-1}p \rangle + H_f^*(-\alpha(K^*)^{-1}p) \right.$$

$$\left. + \langle r^l, \alpha(K^*)^{-1}p \rangle \right\}$$

$$= \inf_{p,q} \sup_v \left\{ H_f^*(-q) + \langle r^l - f, q \rangle + \alpha J^*(p) \right.$$

$$\left. + \langle K^*q - \alpha p, v \rangle \right\}$$

$$= \sup_v \left\{ \inf_q (H_f^*(-q) + \langle r^l - f, q \rangle + \langle Kv, q \rangle) \right.$$

$$\left. + \alpha \inf_p (J^*(p) - \langle p, v \rangle) \right\}$$

$$= \sup_v \left\{ -\sup_q (H_f^*(-q) + \langle Kv + r^l - f, q \rangle) - \alpha J(v) \right\}$$

$$= -\inf_v \left\{ H_f(Kv + r^l - f) + \alpha J(v) \right\}$$

We obtain the simple primal (bidual) iterative regularization technique, equivalent to the dual formulation above:

Problem 6 (Dual Inverse Scale Space, primal form.)

$$u^{l+1} = \operatorname*{arg\,min}_{u \in \mathcal{W}(\Omega)} \left\{ H_f(Ku + r^l - f) + \alpha J(u) \right\}, \qquad (17)$$

with $r^l \in \partial H_f^*(-\alpha(K^*)^{-1}p^l)$.
Since both H_f and J are proper, lower semi-continuous and convex, and since H_f is locally bounded, we have

$$\partial\left(H_f(Ku) + \alpha J(u)\right) = \partial H_f(Ku) + \alpha\, \partial J(u)$$

for all $u \in \mathcal{W}(\Omega)$, due to [ET99]. Hence, the optimality condition of (17) is given via

$$0 \in K^*(\partial H_f(Ku^{l+1} - f + \partial H_f^*(-\alpha(K^*)^{-1}p^l))) + \alpha p^{l+1}$$

$$\Longleftrightarrow (K^*)^{-1}(-\alpha p^{l+1})$$

$$\in \partial H_f(Ku^{l+1} - f + \partial H_f^*(-\alpha(K^*)^{-1}p^l))$$

$$\Longleftrightarrow \partial H_f^*(-\alpha(K^*)^{-1}p^{l+1})$$

$$= \partial H_f^*(-\alpha(K^*)^{-1}p^l) + Ku^{l+1} - f$$

Consequently, the first order optimality condition of this variational problem provides an update of the residual function r^l,

$$r^{l+1} = r^l + Ku^{l+1} - f \qquad (18)$$

for $r^l \in \partial H_f^*(-\alpha(K^*)^{-1}p^l)$ and $r^{l+1} \in \partial H_f^*(-\alpha(K^*)^{-1}p^{l+1})$. This recursion formula yields an interesting decomposition of f involving "noise" at levels l and $l+1$ and signal at level $l+1$.

2.2.1 Well-Definedness of the Iterates

In the following we show that the iterative dual-Bregman procedure is well-defined, i.e. that (18) has a minimizer u^{l+1} and that we may find a suitable subgradient r^{l+1}.

Proposition 1 *Assume H_f to be a strict convex fidelity with operator K having a trivial null space and J to be a convex functional. Let $u^0 := 0$, $p^0 := 0 \in \partial J(u^0)$, $r^0 = 0 \in \partial H_f^*(-\alpha(K^*)^{-1}p^0)$ and $\alpha > 0$. Then, the minimizers u^{l+1} in (17) are well-defined.*

Proof As described above, rewriting the optimality condition of (17) yields the update (18) of the residuals. Since $r^0 = 0$, the $l-th$ residual can be expressed explicitly by

$$r^l = -\sum_{i=1}^l (f - Ku^i) \in \partial H_f^*(-\alpha(K^*)^{-1}p),$$

consequently (18) changes to

$$u^{l+1} = \operatorname*{arg\,min}_{u \in \mathcal{W}(\Omega)} \left\{ H_f(Ku + \sum_{i=1}^l (Ku^i) - (l+1)f) + \alpha J(u) \right\}.$$

Hence, the existence of minimizers can be traced back to existence of minimizers for the original reconstruction problem, just with modified given data, which can be treated as usual. Moreover, as K has only a trivial null space, the strict convexity of H_f and the convexity of J implies the strict convexity of the functional (17), and therefore the minimizers u^{l+1} are unique. □

2.2.2 Dual Inverse Scale Space Flow

To derive a dual nonlinear inverse scale space flow we have to take a look at the update formula (18) due to the optimality condition of (17). In the limit $\alpha \downarrow 0$, this can be interpreted as a forward Euler discretization of the flow

$$\frac{d}{dt}r(t) = Ku(t) - f, \qquad r(0) = 0, \qquad (19)$$

with $r(t) \in \partial H_f^*(-\alpha(K^*)^{-1}p(t))$, which is termed *dual nonlinear inverse scale space method* (in analogy to previous works [BGO+06], [BFO+07]). By defining the integrated residual in that way, we obtain

$$p(t) = \frac{1}{\alpha}K^*(q(t)) . \tag{20}$$

2.2.3 Error Estimates

In order to derive error estimates in the iterative Bregman distance setting we need to introduce the so-called source condition

$$\exists \tilde{p} \in \partial J(\tilde{u}), \exists \tilde{q} \in \mathcal{V}(\Sigma)^* : \quad \tilde{p} = \frac{1}{\alpha}K^*\tilde{q} . \tag{SC}$$

The nowadays standard source condition (SC) will in some sense ensure that a solution $\tilde{u}$ contains features that are enhanced by the regularization term J. Notice the resemblance between the time dependent subgradient $p(t)$ in (20) and $\tilde{p}$ in the source condition. Now we consider the Bregman distance for the convex conjugate of H_f, $D_{H_f^*}^{\partial H_f^*(-q(t))}(-\tilde{q}, -q(t))$, which is finite due to the source condition. Then

$$\frac{d}{dt}\left(D_{H_f^*}^{\partial H_f^*(-q(t))}(-\tilde{q}, -q(t))\right)$$
$$= \frac{d}{dt}\left(H_f^*(-\tilde{q}) - H_f^*(-q(t)) - \partial H_f^*(-q(t))(-\tilde{q} + q(t))\right)$$
$$= \langle r_t, \tilde{q} - q(t)\rangle \stackrel{(19)}{=} \langle f - Ku(t), q(t) - \tilde{q}\rangle$$
$$= \langle f - g, q(t) - \tilde{q}\rangle - \langle Ku(t) - g, q(t) - \tilde{q}\rangle$$
$$= \langle f - g, q(t) - \tilde{q}\rangle - \alpha\langle u(t) - \tilde{u}, p(t) - \tilde{p}\rangle$$
$$= \langle f - g, q(t) - \tilde{q}\rangle - \alpha D_J^{symm}(u(t), \tilde{u})$$
$$\leq \langle f - g, q(t) - \tilde{q}\rangle - \alpha D_J^{p(t)}(\tilde{u}, u(t))$$
$$=: -I(t) . \tag{21}$$

In the following we want to analyse the monotony behavior of $I(t)$. For that purpose, we deduce a relation between q_t and the second derivative of the data fidelity H_f from the dual inverse scale space flow in (19) first, which is

$$\frac{d}{dt}(\partial H_f^*(-q(t))) = \partial_q^2(H_f^*(-q(t)))(-q_t)$$
$$= Ku(t) - f .$$

By using an equivalent definition of the convex conjugate in the differentiable case $(\partial H_f^* = (\partial H_f)^{-1})$ and by using the derivative of inverse functions this yields

$$q_t(t) = \partial_t(\alpha(K^*)^{-1}p(t))$$
$$= [\partial_q^2(H_f^*(-q(t)))]^{-1}(f - Ku(t))$$
$$= [\partial_u[(\partial_u H_f(\partial H_f^*(-q(t)))^{-1})]]^{-1}(f - Ku(t))$$
$$= \partial_u^2 H_f(r(t))(f - Ku(t)), \quad r(t) \in \partial H_f^*(-q(t)) .$$

Consequently, the temporal properties of the estimate $I(t)$ read as follows:

$$\frac{d}{dt}(I(t))$$
$$= -\langle q_t, f - g\rangle + \alpha\frac{d}{dt}\left(D_J^{p(t)}(\tilde{u}, u(t))\right)$$
$$= -\langle q_t, f - g\rangle - \alpha\langle\tilde{u} - u(t), p_t\rangle$$
$$= -\langle q_t, f - g\rangle - \langle g - Ku(t), \partial_t(\alpha(K^*)^{-1}p(t))\rangle$$
$$= -\langle f - g, q_t\rangle$$
$$\qquad - \langle g - Ku(t), H_f''(r(t))(f - Ku(t))\rangle \tag{22}$$
$$= -\langle f - g, H_f''(r(t))(f - Ku(t))\rangle$$
$$\qquad - \langle g - Ku(t), H_f''(r(t))(f - Ku(t))\rangle$$
$$= -\langle f - Ku(t), H_f''(r(t))(f - Ku(t))\rangle$$
$$\leq 0$$
$$\iff \text{ Hessian } H_f'' \text{ is positive semidefinite}$$
$$\iff H_f \text{ convex}$$

with $r(t) \in \partial H_f^*(-q(t))$. Hence, after integrating inequality (22) from 0 to t we get a decrease of I in time and obtain

$$I(t) \leq I(s) \ \forall t \geq s \implies t \cdot I(t) \leq \int_0^t I(s)ds . \tag{23}$$

Now, integrating (21) from 0 to t yields

$$D_{H_f^*}^{r(t)}(-\tilde{q}, -q(t)) - D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0)) \leq -\int_0^t I(s)ds$$
$$\Rightarrow t \cdot I(t) \stackrel{(23)}{\leq}$$
$$\int_0^t I(s)ds \leq D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0)) - D_{H_f^*}^{r(t)}(-\tilde{q}, -q(t)). \tag{24}$$

In the case of noise-free data, i.e. $\delta = 0$, $I(t)$ reduces to the time dependent Bregman distance we want to estimate. Hence, we can conclude

$$t\,\alpha\cdot D_J^{p(t)}(\tilde{u}, u(t)) \leq D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0)) \underbrace{- D_{H_f^*}^{r(t)}(-\tilde{q}, -q(t))}_{\leq 0}$$

and thus:

Theorem 1 (Exact data) *Let $\tilde{u} \in \mathcal{U}(\Omega)$ satisfy $K\tilde{u} = g = f$, and (SC). Moreover, with $q(t) := (K^*)^{-1}p(t)$, let u be a solution of the dual inverse scale space flow*

$$\partial_t r(t) = Ku(t) - g, \qquad r(t) \in \partial H_f^*(-q(t)) .$$

Then the convergence rate $D_J^{p(t)}(\tilde{u}, u(t)) = \mathcal{O}(\frac{1}{t})$ holds, more precisely

$$D_J^{p(t)}(\tilde{u}, u(t)) \leq \frac{D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0))}{\alpha\,t} .$$

In the case of noisy data some further effort is necessary, since the temporal derivative of the Bregman distance with respect to H_f^* in (21) is not only bounded by the negative Bregman distance. Therefore, (24) reads as follows:

$$D_{H_f^*}^{r(t)}(-\tilde{q}, q(t)) - \langle t(f-g), q(t) - \tilde{q} \rangle + t\, \alpha\, D_J^{p(t)}(\tilde{u}, u(t))$$
$$\leq D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0)) \quad (25)$$

To find a lower bound of the first two terms on the left hand side, we provide the following lemma.

Lemma 1 *Let $F : X \to \mathbb{R} \cup \{\infty\}$ be a convex functional, $u, v \in X$, $p, q \in X^*$ and $t \in \mathbb{R}^+$. Then the duality product can be estimated by a sum of Bregman distances,*

$$\langle t(u-v), p-q \rangle_X \leq c\, t^2\, D_F^{\partial F(u)}(v, u) + D_{F^*}^{\partial F^*(q)}(p, q),$$

with a constant,

$$c := \left(\inf_{w \in [v,u]} \|F''(w)\| \right)^{-1} \left(\inf_{\xi \in [q,p]} \|F^{*''}(\xi)\| \right)^{-1},$$

depending on the norm of the second derivative of F and its convex conjugate.

Proof One observes from a Taylor expansion of F^* in p around q with a residual term,

$$F^*(p) = F^*(q) + \langle (F^*)'(q), p-q \rangle + \frac{1}{2}\langle p-q, (F^*)''(\xi)(p-q) \rangle$$

with $\xi \in [q, p]$, that a lower bound for the corresponding Bregman distance is given by

$$D_{F^*}^{(F^*)'(q)}(p, q) = F^*(p) - F^*(q) - \langle (F^*)'(q), p-q \rangle$$
$$= \frac{1}{2}\langle p-q, (F^*)''(\xi)(p-q) \rangle$$
$$\geq \frac{\epsilon}{2}\|p-q\|^2,$$

with $\epsilon := \inf_{\xi \in [q,p]} \|(F^*)''(\xi)\|$.

In analogy a Taylor expansion of F in v around u,

$$F(v) = F(u) + \langle F'(u), v-u \rangle + \frac{1}{2}\langle v-u, F''(w)(v-u) \rangle,$$

with $w \in [u-v]$, yields

$$\frac{1}{\tilde{c}\,\epsilon} D_F^{F'(u)}(v, u) = \frac{1}{\tilde{c}\,\epsilon}(F(v) - F(u) - \langle F'(u), v-u \rangle)$$
$$= \frac{1}{2\,\tilde{c}\,\epsilon}\langle v-u, F''(w)(v-u) \rangle$$
$$\geq \frac{1}{2\,\epsilon}\|u-v\|^2,$$

with $\tilde{c} := \inf_{w \in [u,v]} \|F''(w)\|$.

Using Young's inequality we obtain

$$\langle t(u-v), p-q \rangle \leq \frac{t^2}{2\epsilon}\|u-v\|^2 + \frac{\epsilon}{2}\|p-q\|^2$$
$$\leq c\, t^2\, D_F^{\partial F(u)}(v, u) + D_{F^*}^{\partial F^*(q)}(p, q)$$

With constant $c := \frac{1}{\tilde{c}\epsilon}$ we get the desired estimation. $\square$

Now, applying Lemma 1 to functional H_f yields the estimate

$$\langle t(f-g), q(t) - \tilde{q} \rangle \leq D_{H_f^*}^{r(t)}(-\tilde{q}, -q(t)) + ct^2 D_{H_f}^{\partial H_f(f)}(g, f),$$

such that the Bregman distance regarding to H_f^* is annihilated in (25) and that we can conclude

$$-ct^2 D_{H_f}^{\partial H_f(f)}(g, f) + t\, \alpha\, D_J^{p(t)}(\tilde{u}, u(t)) \leq D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0)).$$

Finally, since $H_f(f) = 0$, $\partial H_f(f) = 0$ and with the upper bound of the noise (14) we have

$$D_{H_f}^{\partial H_f(f)}(g, f) = H_f(g) \leq \delta,$$

provides a general error estimate for the dual inverse scale space method:

Theorem 2 (Noisy data) *Let $\tilde{u} \in \mathcal{U}(\Omega)$ satisfy $K\tilde{u} = g$ and (SC), and let f be noisy data satisfying (14). Moreover, with $q(t) := (K^*)^{-1}p(t)$, let u be a solution of the dual inverse scale space flow*

$$\partial_t\, r(t) = Ku(t) - f, \qquad r(t) \in \partial H_f^*(-q(t)).$$

Then the error estimate

$$D_J^{p(t)}(\tilde{u}, u(t)) \leq \frac{D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0))}{\alpha\, t} + c\, \delta\, t$$

holds. In particular, for the choice $t_*(\delta) := \mathcal{O}(\frac{1}{\delta})$ we obtain the convergence rate $D_J^{p(t_*)}(\tilde{u}, u(t_*)) = \mathcal{O}(\delta)$.

Remark 1 In the case of Poisson noise modeling, i.e. if H_f is the Kullback-Leibler data fidelity, c_{KL} reads as follows:

$$c_{KL}$$
$$= \left(\inf_{w \in [g,f]} \|\partial_w^2 H_f(w)\| \right)^{-1} \left(\inf_{\xi \in [\tilde{q}, q(t)]} \|\partial_\xi^2 H_f^*(\xi)\| \right)^{-1}$$
$$= \frac{\max\{\sup g, \sup f\}^2}{\inf f} \frac{\max\{\sup(1-\tilde{q}), \sup(1-q(t))\}^2}{\inf f},$$

since $\partial_w^2 H_f(w) = \frac{f}{w^2}$, $H_f^*(\xi) = \int_\Sigma -f \log(1-\xi)$ and $\partial_\xi^2 H_f^*(\xi) = \frac{f}{(1-\xi)^2}$.

3 Reconstruction with Poisson Noise: Bregman-EM-TV Methods

In the following we present reconstruction algorithms for inverse problems with measured data drawn from Poisson statistics (e.g. deconvolution problems in fluorescence microscopy). More precisely, we use the Kullback-Leibler functional, i.e.

$$H_f(Ku - f) := \int_{\Sigma} (f \log\left(\frac{f}{Ku}\right) - f + Ku)\, d\mu \quad (26)$$

as the data fidelity and the (exact) total variation (TV), see (9), as regularization $J(u)$. Interestingly, the iteration steps of reconstruction methods proposed in this chapter, including the primal and dual inverse scale space methods of the last section, are all based on the following variational framework:

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0}} \int_{\Sigma} (Ku + b - f \log(Ku + b))\, d\mu$$
$$+ \alpha(|u|_{BV} - \langle p, u \rangle)\,, \quad (27)$$

with

$\alpha := 0, \quad p := 0 \qquad$ for EM

$p := 0, \qquad\qquad$ for EM-TV

$b := 0, \quad p := p^l \qquad$ for Bregman-EM-TV

$b := r^l, \quad p := 0 \qquad$ for Dual-Bregman-EM-TV .

In comparison to (26), we skipped terms independent of u in (27) for simplicity without affecting minimizers. In the following subsection we will explain the EM reconstruction algorithm and will proceed with the presentation of EM combined with TV regularization. Finally we are going to introduce the EM-TV based inverse scale space methods Bregman-EM-TV and Dual-Bregman-EM-TV.

3.1 Reconstruction method: EM algorithm

In literature there are two types of reconstruction methods that are used in general: analytic (direct) and algebraic (iterative) methods. A classical example for a direct method is the Fourier-based filtered backprojection (FBP). Although FBP is well understood and computationally efficient, iterative type methods obtain more and more attention in the applications mentioned above. The major reason is the high noise level (low SNR) and the type of statistics, which cannot be taken into account by direct methods. Hence, we will give a short review on the Expectation-Maximization (EM) algorithm [SV82,DLR77], which is a popular iterative algorithm to maximize the likelihood function $p(u|f)$ in

problems with incomplete data. In the absence of prior knowledge any object u has the same relevance, i.e. the Gibbs a-priori density $p(u)$ in (4) is constant. We can then normalize $p(u)$ such that $J(u) \equiv 0$. Hence (8) reduces to the constrained minimization problem

$$\min_{u \geq 0} \int_{\Sigma} (Ku + b - f \log(Ku + b))\, d\mu\,, \quad (28)$$

which is (27) with $\alpha := 0$ and $p := 0$. With the natural scaling assumption

$$K^* \mathbf{1} = \mathbf{1}\,,$$

a suitable iteration scheme for computing stationary points, which also preserves positivity (assuming K preserves positivity), is the so called EM algorithm (cf. [NW01])

$$u_{k+1} = u_k\, K^*\left(\frac{f}{Ku_k + b}\right)\,, \quad k = 0, 1, \ldots\,. \quad (29)$$

For noise-free data f several convergence proofs of the EM algorithm to the maximum likelihood estimate, i.e. the solution of (28), can be found in literature [NW01, R07, VSK85, I91]. Besides, it is known that the speed of convergence of iteration (29) is slow. A further property of the iteration is a lack of smoothing, whereby the so-called "checkerboard effect" arises.

For noisy data f it is necessary to differentiate between discrete and continuous modeling. In the discrete case, i.e. if K is a matrix and u is a vector the existence of a minimum can be guaranteed since the smallest singular value is bounded by a positive value. Hence, the vectors 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. Again the reason is the ill-posedness of the integral equation (1), which transfers to problem (28). This aspect can be taken as a lack of additional a-priori knowledge about the unknown u resulting from $J(u) = 0$. The EM algorithm converges to a minimizer if it exists. Consequently, in the continuous case it is essential to ensure consistence of the given data to prevent divergence of the EM algorithm. As described in [R07], the EM iterates show the following typical behavior for ill-posed problems. The (metric) distance between the iterates and the solution decreases initially before it increases as the noise is amplified during the iteration process. This issue might be regulated by using appropriate stopping rules to obtain reasonable results. In [R07] it is shown that certain stopping rules indeed allow stable approximations. Ways to improve reconstruction results are TV or Bregman-TV regularization techniques that we will consider in the following subsections.

3.2 Reconstruction method: EM-TV algorithm

The EM or Richardson/Lucy algorithm is currently the standard iterative reconstruction method for deconvolution problems with Poisson noise based on the linear equation (1). However, with the assumption $J(u) = 0$, 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 the multiplicative fixed point iteration (29) delivers unsatisfactory and noisy results even with early termination. For this purpose 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 interesting approach to improve the reconstruction is the EM-TV algorithm. In the classical EM algorithm, the negative log likelihood functional (28) is minimized. We modify the functional by adding a weighted TV term (cf. [ROF92]),

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0}} \left\{ \int_{\Sigma} (Ku + b - f \log(Ku + b)) \, d\mu \; + \; \alpha |u|_{BV} \right\} . \tag{30}$$

This is (27) with $p = 0$. Using $|u|_{BV}$ as a regularization functional ensures, that images with smaller total variation are preferred in the minimization (have higher prior probability). $BV(\Omega)$ is a popular function space in image processing since it can represent discontinuous functions. By minimizing TV the latter are even preferred [EG92, G84]. Hence, expected reconstructions are cartoon-like images. Obviously, such an approach cannot be used for studying very small structures in an object, but it suits well for segmenting different cell structures and analyzing them quantitatively. For the solution of (30), we propose a forward-backward splitting algorithm, which can be realized by alternating classical EM steps with almost standard TV minimization steps as encountered in image denoising. The latter is solved by using duality [C04], obtaining a robust and efficient algorithm. For designing the proposed alternating algorithm, we consider the first order optimality condition of (30). Due to the total variation, this variational problem is not differentiable in the usual sense. But the latter is convex since TV is convex and since we can extend the data fidelity term to a Kullback-Leibler functional, cf. [RA07], without affecting the stationary points. For such problems powerful methods from convex analysis are available, e.g. a generalized derivative called the subdifferential [HL93], denoted by ∂. With the natural scaling assumption $K^* \mathbf{1} = \mathbf{1}$, this generalized notion of gradients and the Karush-Kuhn-Tucker

(KKT) conditions [HL93, ,Theorem 2.1.4] yield the existence of a Lagrange multiplier $\lambda \geq 0$ such that

$$\left\{ \begin{array}{ll} 0 & \in 1 - K^* \left(\dfrac{f}{Ku + b} \right) + \alpha \, \partial |u|_{BV} \; - \; \lambda \\ 0 & = \lambda \, u \end{array} \right\} . \tag{31}$$

By multiplying (31) with u we can eliminate the Lagrange multiplier and derive the following semi-implicit iteration scheme

$$u_{k+1} \; - \; u_k \, K^* \left(\frac{f}{K u_k + b} \right) \; + \; \alpha \, u_k \, p_{k+1} \; = \; 0 \tag{32}$$

with $p_{k+1} \in \partial |u_{k+1}|_{BV}$. Remarkably, the second term within this iteration scheme is the EM step in (29). Consequently, method (32) solving variational problem (30), can be realized as a nested two step iteration,

$$\left\{ \begin{array}{ll} u_{k+\frac{1}{2}} & = u_k \, K^* \left(\dfrac{f}{K u_k + b} \right) & \text{(EM step)} \\ u_{k+1} & = u_{k+\frac{1}{2}} - \alpha \, u_k \, p_{k+1} & \text{(TV step)} \end{array} \right\} . \tag{33}$$

Thus, we alternate an EM step with a TV correction step. The complex second half step from $u_{k+\frac{1}{2}}$ to u_{k+1} can be realized by solving the following variational problem,

$$u_{k+1} = \arg\min_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{(u - u_{k+\frac{1}{2}})^2}{u_k} + \alpha \, |u|_{BV} \right\} . \tag{34}$$

Inspecting the first order optimality condition confirms the equivalence of this minimization with the TV correction step in (33). Problem (34) is just a modified version of the Rudin-Osher-Fatemi (ROF) model, with weight $\frac{1}{u_k}$ in the fidelity term. This analogy creates the opportunity to carry over efficient numerical schemes known for the ROF-model.

For the solution of (34) we use the exact definition of TV (9) with dual variable g and derive an iteration scheme for the quadratic dual problem similar to [C04]. The resulting algorithm reads as follows: We initialize the dual variable g^0 with 0 (or the resulting g from the previous TV correction step) and for any $n \geq 0$ we compute the update

$$g^{n+1} \; = \; \frac{g^n + \tau \, \nabla \alpha \, u_k \, \mathrm{div} g^n - u_{k+\frac{1}{2}})}{1 + \tau \, |\nabla(\alpha \, u_k \, \mathrm{div} g^n - u_{k+\frac{1}{2}})|} \; , \; 0 < \tau < \frac{1}{4 \, \alpha \, u_k} \; ,$$

with the constrained damping parameter τ to ensure stability and convergence of the algorithm.

For a detailed analytical examination of EM-TV we refer to [BSB09].

3.3 Inverse Scale Space Method: Bregman-EM-TV

The presented EM-TV algorithm (33) solves the problem (30) and delivers cartoon-reconstructions with sharp edges due to TV regularization. However, the realization of TV steps via the weighted ROF-model (34) has the drawback that reconstructed images suffer from losing contrast. Thus, we propose to extend (30) and therewith EM-TV by iterative regularization to a simultaneous contrast correction. More precisely, we perform a contrast enhancement by inverse scale space methods and by using the Bregman iteration. These techniques have been derived in [OBG+05], with a detailed analysis for Gaussian-type problems (7), and have been generalized in [BGO+06, BFO+07]. Following these methods, an iterative refinement is realized by a sequence of modified EM-TV problems based on (30).

The inverse scale space methods concerning TV, derived in [OBG+05], follow the concept of iterative regularization by the Bregman distance [B67]. In case of the Poisson-model the method initially starts with a simple EM-TV algorithm, i.e. it consists in computing a minimizer u^1 of (30). Then, updates are determined successively by considering variational problems with a shifted TV, namely (10), where p^l is an element of the subgradient of the total variation in u^l. The Bregman distance concerning TV is defined in (11). The introduction of this definition allows to characterize the sequence of modified variational problems (10) by addition of constant terms as

$$u^{l+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \int_{\Sigma} (Ku - f \log Ku) \, d\mu + \alpha D^{p^l}_{|\cdot|_{BV}}(u, u^l) \,.$$
$$(35)$$

Thus, the first iterate u^1 can also be realized by the variational problem (35), if u^0 is constant and $p^0 := 0 \in \partial |u^0|_{BV}$. We point out, that (35) is (27) with $b := 0$ and $p := p^l$. The Bregman distance $D^p_{|\cdot|_{BV}}$ does not represent a distance in the common (metric) sense, since D is not symmetric in general and the triangle inequality does not hold. Though, compared to (10), the formulation in (35) offers the advantage that $D^p_{|\cdot|_{BV}}$ is a distance measure with

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

Besides, the Bregman distance is convex in the first argument because $|\cdot|_{BV}$ is convex. In general, i.e. for any convex functional J (see e.g. [BFO+07]), the Bregman distance can be interpreted as the difference between $J(\cdot)$ in u and the Taylor linearization of J around $\tilde{u}$ if, in addition, J is continuously differentiable.

Before deriving a two-step iteration corresponding to (33) we will motivate the contrast enhancement by iterative regularization in (35). The TV regularization in (30) prefers functions with only few oscillations. 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 (35) can be interpreted as follows: search for a solution that matches the Poisson distributed data after applying K and simultaneous minimization of the residual of the Taylor approximation of $|\cdot|_{BV}$ around u^l. In the following we will see that this form of regularization does not change the position of gradients with respect to the last computed EM-TV solution u^l but that an increase of intensities is permitted. This leads to a noticeable contrast enhancement.

For the derivation of a two-step iteration we consider the first order optimality condition of the variational problem (10) resp. (35). Due to convexity of the Bregman distance in the first argument we can determine the subdifferential of (35). Analogous to the derivation of the EM-TV iteration the subdifferential of the log likelihood functional can be expressed by the Fréchet derivative in (31). Hence, with the natural scaling assumption $K^*1 = 1$ the optimality condition is given by

$$0 \in 1 - K^* \left(\frac{f}{Ku^{l+1}} \right) + \alpha \, (\, \partial |u^{l+1}|_{BV} - p^l \,), \quad (36)$$

with $p^l \in \partial |u^l|_{BV}$.
For u^0 constant and $p^0 := 0 \in \partial |u^0|_{BV}$ this delivers a well defined update of the iterates p^l,

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

Analogous to EM-TV we can apply the idea of the nested iteration (33) in every refinement step, $l = 1, 2, \cdots$. For the solution of (35) condition (36) yields a strategy consisting of an EM-step $u^{l+1}_{k+\frac{1}{2}}$ followed by solving the adapted weighted ROF-problem

$$u^{l+1}_{k+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ \frac{1}{2} \int_{\Omega} \frac{(u - u^{l+1}_{k+\frac{1}{2}})^2}{u^{l+1}_k} + \alpha \, (\, |u|_{BV} - \langle p^l, u \rangle \,) \right\} \,.$$
$$(37)$$

Following [OBG+05, BGO+06, BFO+07], 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 Bregman distance via a slightly modified EM-TV algorithm. With the

scaling $v^l := \alpha\, p^l$ and (36) we obtain the following update formula

$$v^{l+1} = v^l - \left(1 - K^* \left(\frac{f}{Ku^{l+1}}\right)\right), \qquad v^0 = 0 \ . \quad (38)$$

Using this scaled update we can rewrite the second step (37) to

$$u_{k+1}^{l+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \frac{1}{2} \int_\Omega \frac{(u - u_{k+\frac{1}{2}}^{l+1})^2 - 2uu_k^{l+1}v^l}{u_k^{l+1}} + \alpha\, |u|_{BV} \ .$$

Note that

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

holds, where the last two terms are independent of u. Hence (37) simplifies to

$$u_{k+1}^{l+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \frac{1}{2} \int_\Omega \frac{(u - (u_{k+\frac{1}{2}}^{l+1} + u_k^{l+1}v^l))^2}{u_k^{l+1}} + \alpha\, |u|_{BV} \ , \quad (39)$$

i.e. the second step (37) can be realized by a slight modification of the TV step introduced in (34). Obviously, the efficient numerical implementation of the weighted ROF-problem in Section 3.2 using the exact definition of TV and duality strategies can be applied in complete analogy to (39). The update variable v in (38) is an error function with reference to the optimality condition of the unregularized log-likelihood functional (28). In every refinement step of the Bregman iteration v^{l+1} differs from v^l by the current error in the optimality condition (28). Within the TV-step (39) one observes that an iterative regularization with the Bregman distance leads to contrast enhancement. Instead of fitting to the EM solution $u_{k+\frac{1}{2}}^{l+1}$ in the weighted norm, we use a function in the fidelity term whose intensities are increased by the error function v^l. Resulting from the idea of adaptive regularization v^l is weighted by u_k^{l+1}, too. As usual for iterative methods the described reconstruction method by iterative regularization needs a stopping criterion. The latter should stop at an iteration offering a solution that approximates the true image as good as possible. This is necessary to prevent that too much noise arises by the inverse scale space strategy. In the case of Gaussian noise, the discrepancy principle is a reasonable stopping criterion, i.e. the procedure would stop if the residual $\left\|Ku^l - f\right\|_2$ reaches the variance of the noise. In the case of Poisson noise, however, it makes sense to stop the Bregman iteration if the Kullback-Leibler distances of Ku^l and the given data f reach the

noise level. For synthetic data the noise level is naturally given by the KL distance between $g = K\tilde{u}$ and f, where $\tilde{u}$ denotes the true, noise-free image. For experimental data it is necessary to find a suitable estimate for the noise level from counts.

3.4 Dual Inverse Scale Space Method: Dual-Bregman-EM-TV

In section 2.2 we presented a dual inverse scale space method in terms of an iterative Bregman regularization technique for general, convex data fidelities and regularization terms. This strategy based on a dual representation of the initial variational problem (15). A bidual formulation of the dual inverse scale space strategy offers a simple interpretation in terms of a familiar (primal) problem (17). In the special case of Poisson noise modeling and TV regularization, this reads as follows

$$u^{l+1} = \operatorname*{arg\,min}_{\substack{u \in \mathcal{BV}(\Omega) \\ u \geq 0}} \left\{ \int_\Sigma (Ku + r^l - f\log(Ku + r^l))\, d\mu \right.$$
$$\left. + \alpha|u|_{BV} \right\} \ , \quad (40)$$

with the update of the residual function r^l (see (18))

$$r^{l+1} = r^l + Ku^{l+1} - f \ , \qquad \text{with } r^0 = 0 \ .$$

The variational problem above can simply be interpreted as (27), if $b := r^l$ and $p := 0$. Comparing the proposed iterative regularization technique with the EM-TV problem in (30), reveals the noise function r^l as a dynamically updated background model instead of a time-constant background b. Shifting the argument of the data fidelity with r^l in that appropriate way, leads to the expected contrast enhancing behavior in each time step $l \to l+1$.

Although the minimization problem (40) for a specific l can intuitively be implemented in analogy to the splitting strategy of EM-TV,

$$\begin{cases} u_{k+\frac{1}{2}} & = u_k\, K^* \left(\dfrac{f}{Ku_k + r^l}\right) & \text{(EM step)} \\[2ex] u_{k+1} & = u_{k+\frac{1}{2}} - \alpha\, u_k\, p_{k+1} & \text{(TV step)} \end{cases} ,$$

we need to be aware of division-by-zero problems in the EM step. For the dual inverse scale method, we can overcome this problem by a partially explicit approximation. For this sake we rewrite the optimality

condition in the following way, using $K^*1 = 1$,

$$1 - K^* \left(\frac{f}{Ku + r} \right) + \alpha p = 0$$

$$\implies 1 - \frac{f}{Ku + r} + \alpha (K^*)^{-1} p = 0$$

$$\implies Ku + r - f + \alpha (K^*)^{-1} p(Ku + r) = 0$$

$$\implies K^* \left(1 - \frac{f - r - \alpha r (K^*)^{-1} p}{Ku} \right) + \alpha p = 0$$

$$\implies 1 - K^* \left(\frac{f - r(1 + \alpha (K^*)^{-1} p)}{Ku} \right) + \alpha p = 0 \, .$$

Now we use an approximation of the first term including the subgradient p from the last Bregman step and obtain

$$u_{k+1} = u_k K^* \left(\frac{f - r^l(1 + q^l)}{Ku_k} \right) - \alpha \, u_k \, p_{k+1} \, ,$$

with $q^l = \alpha (K^*)^{-1} p^l$. Note that q^l does not need to be computed by inverting K^*, but can be obtained from the update formula

$$q^{l+1} = \frac{f - r^l(1 + q^l)}{Ku^{l+1}} - 1.$$

4 Application to Optical Nanoscopy: Results

In recent years revolutionary imaging techniques have been developed in light microscopy with enormous importance for biological and material sciences or medicine. For a couple of decades the technology of light microscopy has been considered to be exhausted, as the resolution is basically limited by Abbe's law for diffraction of light. By developing stimulated emission depletion (STED)- and 4Pi-microscopy now resolutions are achieved that are way beyond these diffraction barrier [K00, HS06]. To get an impression of nanoscopic images blurred by different convolution kernels (PSFs), we refer to Figure 1. In this section we present the performance of the proposed techniques by reconstructing synthetic and experimental data.

Figure 2 illustrates our techniques at a simple synthetic object by applying a 4Pi convolution and adding Poisson noise. With EM-TV (see 2(c) and 2(g)) we get rid of noise and oscillations, but we are not able to separate the objects sufficiently. Using Bregman-EM-TV a considerable improvement resulting from contrast enhancement can be achieved. This aspect is underlined by the values of the KL-distance for the different reconstructions. In Figure 4 we compare the primal and dual inverse scale space strategy using the same synthetic object, but in this case with a Gaussian convolution kernel. As expected, both inverse scale space strategies
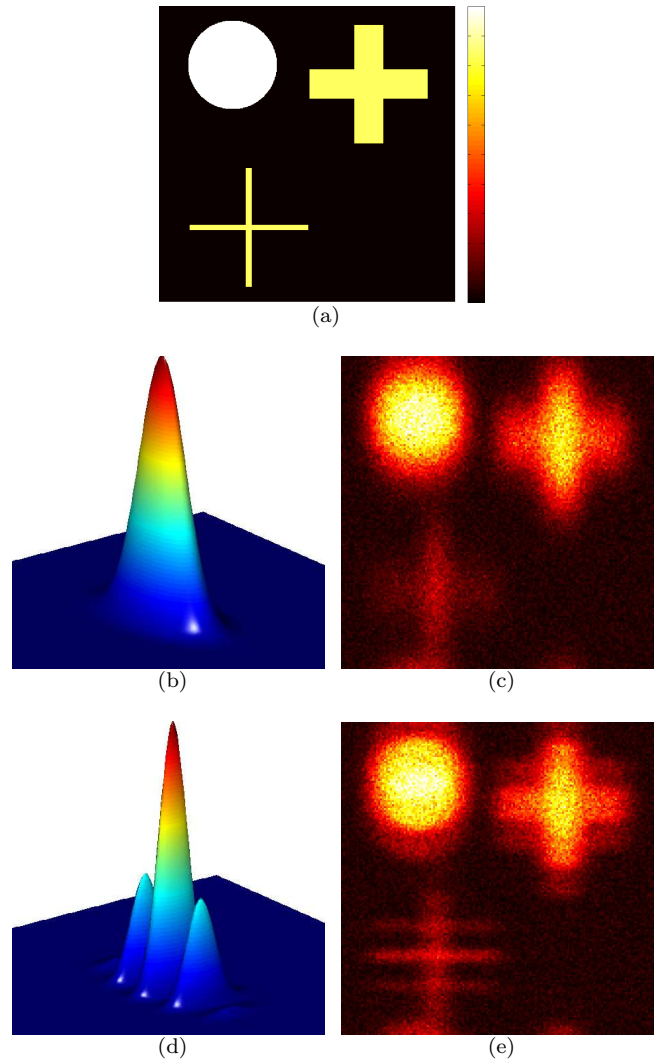
Fig. 1 Synthetic data concerning different PSFs: (a) true image; (b) Gaussian PSF; (c) is convolved with Gaussian PSF and Poisson noise; (d) PSF appearing in 4Pi microscopy; and (e) is convolved with 4Pi PSF and Poisson noise.

compute very similar iterates and we can observe a decrease of the Kullback-Leibler distance between u and $\tilde{u}$ until the noise level is reached. Taking a closer look at the distance measurements reveals a slightly better decrease in the case of Dual-Bregman-EM-TV.

Figure 3, (a)-(c) demonstrate the protein Bruchpilot [K06] and its EM-TV and Bregman-EM-TV reconstruction. Particularly, the latter delivers well separated object segments and a high contrast level. In Figure 3, (d)-(f) we illustrate our techniques by reconstructing Syntaxin [WHM+07], a membrane integrated protein participating in exocytosis. Here, the contrast enhancing property of Bregman-EM-TV is observable as well, compared to EM-TV. It is possible to preserve fine structures in the image.
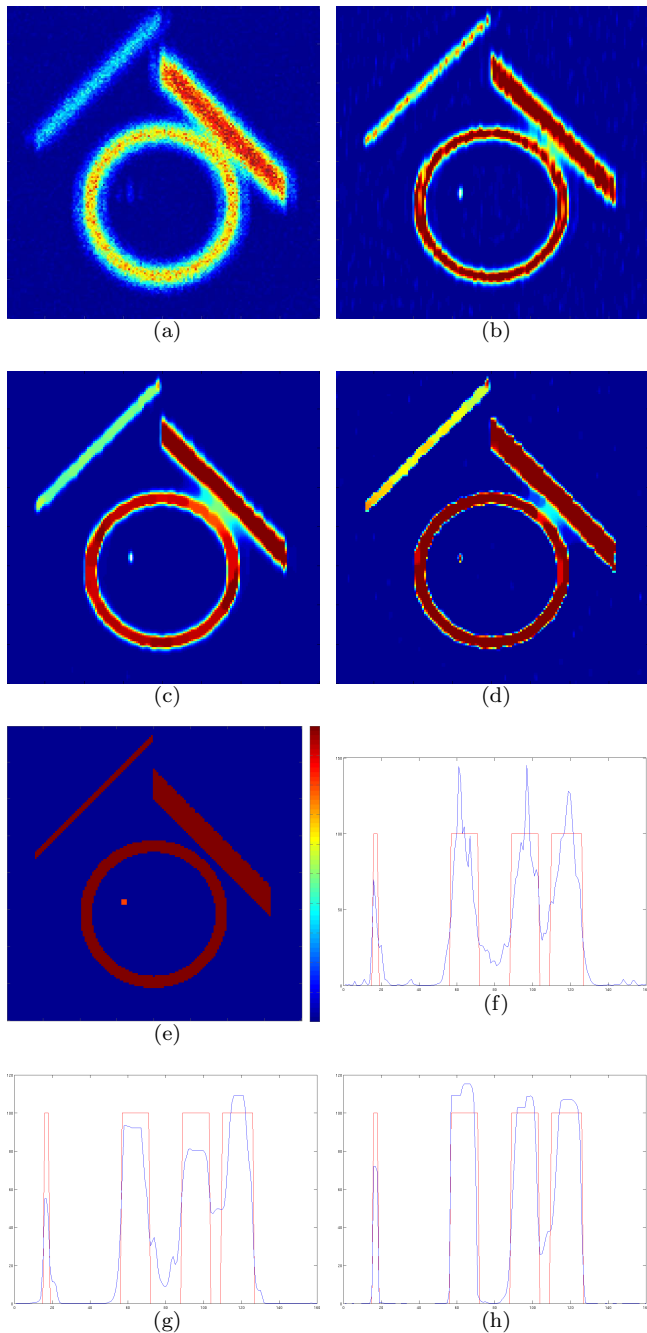
Fig. 2 Synthetic data: (a) raw data using 4Pi PSF; (b) EM reconstruction, 20 its, KL-distance: 3.20; (c) EM-TV, $\alpha = 0.04$, KL-distance: 2.43; (d) Bregman-EM-TV, $\alpha = 0.1$, after 4 updates, KL-distance: 1.43; (e) true image; (f)-(h) horizontal slices EM, EM-TV and Bregman-EM-TV compared to true image slice.

5 Conclusions & Open Questions

We have derived reconstruction methods for inverse problems with Poisson noise. Particularly, we concentrated on deblurring problems in nanoscopic imaging, although the proposed methods can easily be adapted
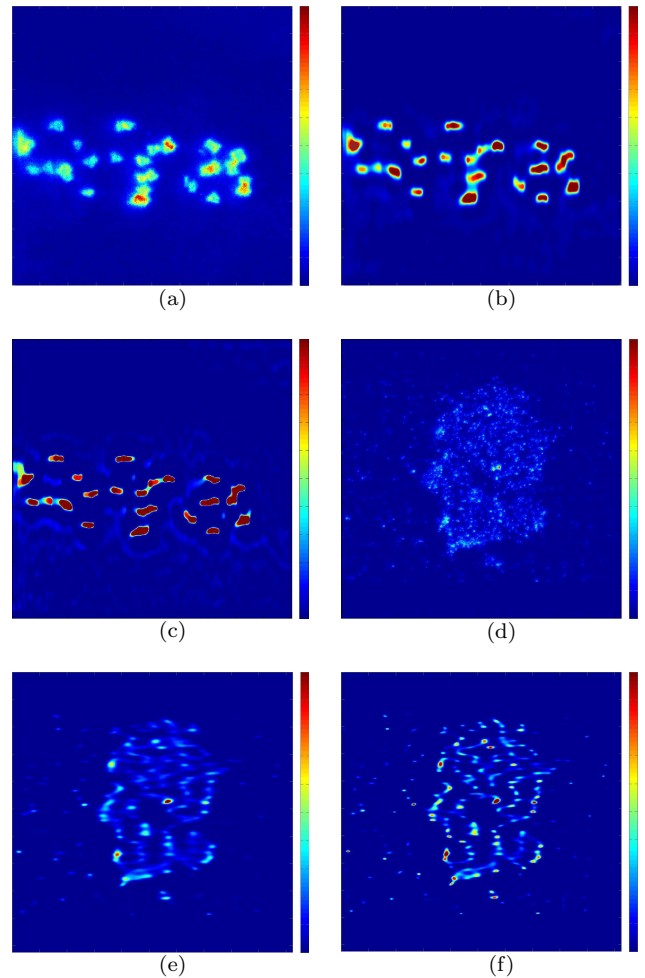
Fig. 3 Experimental data: (a) Protein Bruchpilot in active zones of neuromuscular synapses in larval Drosophila; (b) EM-TV; (c) Bregman-EM-TV; (d) Protein Syntaxin in cell membrane, fixed mamalian (PC12) cell; (e) EM-TV; and (f) Bregman-EM-TV.

to other imaging tasks, i.e. medical imaging (PET, [SBW+08]). Motivated by a statistical modeling we developed a robust EM-TV algorithm that incorporates a-priori knowledge into the reconstruction process. By combining EM with simultaneous TV regularization we can reconstruct cartoon-images with sharp edges, which yield a reasonable basis for quantitative investigations. To overcome the problem of contrast reduction, we extended the reconstruction to Bregman iterations and inverse scale space methods. We applied the proposed methods to optical nanoscopy and pointed out their improvements in comparison to standard reconstruction techniques.

An open issue remains the error estimation for the primal inverse scale space method, which - if applicable at all - will require novel theoretical approaches. On the other hand the derivation of appropriate stopping rules for the dual inverse scale space methods is an important
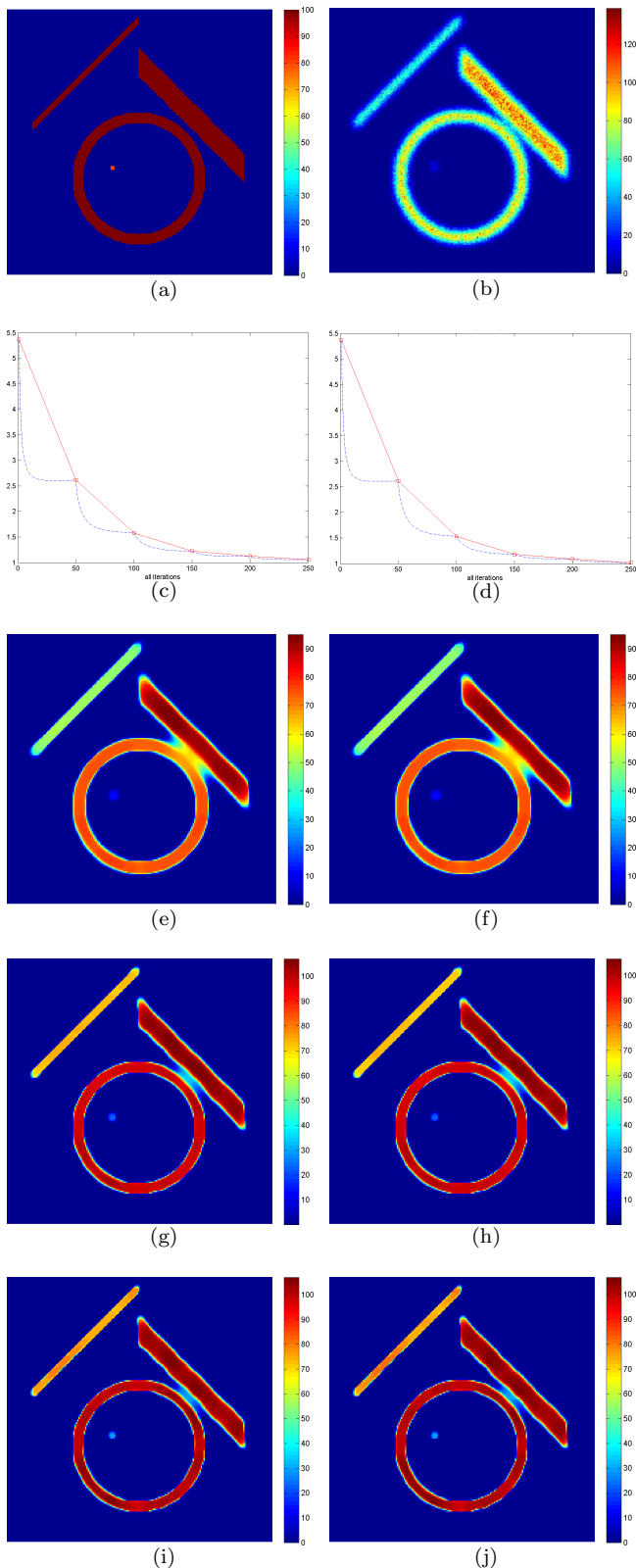
Fig. 4 Synthetic data: Comparison of primal and dual inverse scale space methods; (a): true image; (b): raw data f using Gaussian PSF; (c)-(d): KL-distance between u and $\tilde{u}$ for Bregman-EM-TV resp. Dual-Bregman-EM-TV, blue line: distance at all 250 iterations, red marker: distance at every Bregman step (intervals of 50 interior iterations)(e),(g),(i): iterates u^1,u^3 and u^5 of Bregman-EM-TV; (f),(h),(j): iterates u^1,u^3 and u^5 of Dual-Bregman-EM-TV.

open problem. For practical purposes visual inspection may however be sufficient in most cases.

References

[AA08]   G.Aubert, J.F.Aujol, *A Variational approach to remove multiplicative noise*, SIAM J. Appl. Math. 68 (2008), 925-946.

[AV94]   Acar, R., Vogel, C.R.: Analysis of bounded variation penalty methods for ill-posed problems. Inverse Problems **10**, 1217–1229 (1994)

[BL09]   J.Bardsley, A.Luttman, *Total variation-penalized Poisson likelihood estimation for ill-posed problems* , Adv. Comp. Math. (2009), to appear.

[BB09]   Benning, M., Burger, M.: Error estimates for variational models with non-gaussian noise. Tech. Rep. 09-40, UCLA CAM (2009)

[BLZ08]   Bertero, M., Lantéri, H., Zanni, L.: Iterative image reconstruction: a point of view. In: Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT), *CRM series*, vol. 8 (2008)

[BHM$^+$07]   N.Bissantz, T.Hohage, A.Munk, F.Ruymgaart, *Convergence rates of general regularization methods for statistical inverse problems and applications*, SIAM J. Numer. Anal. **45** (2007), 2610-2636.

[B67]   Bregman, L.M.: The relaxation method for 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)

[BKM$^+$08]   T.Bonesky, K.S.Kazimierski, P.Maass, F.Schöpfer, T.Schuster, *Minimization of Tikhonov functionals in Banach spaces*, Abstr. Appl. Anal. 19 (2008), 192679

[BSB09]   C.Brune, A.Sawatzky, M.Burger, *Bregman-EM-TV Methods with Application to Optical Nanoscopy*, Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision, vol. 5567 (2009), 235-246.

[BSB09]   Brune, C., Sawatzky, A., Wübbeling, F., Kösters, T., Burger, M.: EM-TV methods for inverse problems with poisson noise, in preparation (2009)

[BO04]   Burger, M., Osher, S.: Convergence rates of convex variational regularization. Inverse Problems **20**, 1411–1421 (2004)

[BGO$^+$06]   Burger, M., Gilboa, G., Osher, S., Xu, J.: Nonlinear inverse scale space methods. Commun. Math. Sci. **4**(1), 179–212 (2006)

[BFO$^+$07]   Burger, M., Frick, K., Osher, S., Scherzer, O.: Inverse total variation flow. SIAM Multiscale Modelling and Simulation **6**(2), 366–395 (2007)

[BRH07]   Burger, M., Resmerita, E., He, L.: Error estimation for bregman iterations and inverse scale space methods. Computing **81**, 109–135 (2007)

[BSH08]   M.Burger, C.Schönlieb, L.He, *Cahn-Hilliard inpainting and a generalization to gray-value images*, Preprint (2008), and submitted.

[CDL⁺98] A.Chambolle, R.DeVore, N.Y.Lee, B.Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Proc., 7 (1998), 319–335.

[C04] Chambolle, A.: An algorithm for total variation minimization and applications. J. of Mathematical Imaging and Vision **20**, 89–97 (2004)

[CE05] T.F.Chan, S.Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math. 65 (2005), 1817-1837.

[CS05] T.Chan, J.Shen, *Image Processing and Analysis* (SIAM, Philadelphia, 2005).

[C91] I.Csiszar, *Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems*, Ann. Stat.19 (1991), 2032-2066.

[DLR77] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. J. of the Royal Statistical Society, B **39**, 1–38 (1977)

[DBZ⁺06] N.Dey, L.Blanc-Feraud, C.Zimmer, Z.Kam, P.Roux, J.C.Olivo-Marin, J. Zerubia, *Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution*, Microscopy Research Technique 69 (2006), 260-266.

[EHN96] H.W.Engl, M.Hanke, A.Neubauer, Regularization of Inverse Problems, Kluwer Academic Publishers, Dordrecht (1996) (Paperback edition 2000).

[EHK05] H.W.Engl, A.Hofinger, S.Kindermann, *Convergence rates in the Prokhorov metric for assessing uncertainty in ill-posed problems*, Inverse Problems **21** (2005), 399-412.

[EKN89] H.W.Engl, K.Kunisch, A.Neubauer, *Convergence rates for Tikhonov regularisation of non-linear ill-posed problems*, Inverse Problems 5 (1989) 523-540.

[ET99] Ekeland, I., Temam, R.: Convex analysis and variational problems, Corrected Reprint Edition. SIAM, Philadelphia (1999)

[EG92] Evans, L.C., Gariepy, R.F.: Measure theory and fine properties of functions. Studies in Advanced Mathematics, CRC Press: Boca Raton, FL (1992)

[GG84] S.Geman, D.Geman, *Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1984), 721741.

[GM85] S.Geman, D.E.McClure, *Bayesian image analysis: an application to single photon emission tomography*, Proc. Statistical Computation Section (American Statistical Association, Washington, 1985) 1218.

[G84] Giusti, E.: Minimal surfaces and functions of bounded variation. Birkhäuser (1984)

[HS06] Hell, S., Schönle, A.: Nanoscale resolution in far-field fluorescence microscopy. In: P.W. Hawkes, J.C.H. Spence (eds.) Science of Microscopy. Springer (2006)

[HL93] Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I, Volume 305 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag (1993)

[HP06] A.Hofinger, H.K.Pikkarainen, *Convergence rates for the Bayesian approach to linear inverse problems*, Inverse Problems **23** (2006), 2469-2484.

[HP09] A.Hofinger, H.K.Pikkarainen, *Convergence rates for linear inverse problems in the presence of an additive normal noise*, Stoch. Anal. Appl. **27** (2009), 240-257.

[HKP⁺07] B.Hofmann, B.Kaltenbacher, C.Pöschl, O.Scherzer, *A Convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems 23 (2007), 987–1010.

[H09] T.Hohage, *Variational regularization of inverse problems with Poisson data*, Preprint (2009).

[HNW09] Y.M.Huang, M.K.Ng, Y.W.Wen, *A new total variation method for multiplicative noise Removal*, SIAM J. Imaging Sciences 2 (2009), 20-40.

[I91] Iusem, A.N.: Convergence analysis for a multiplicatively relaxed EM algorithm. Mathematical Methods in the Applied Sciences **14**, 573–593 (1991)

[JHC98] E.Jonsson, S.C.Huang, T.Chan, *Total variation regularization in positron emission tomography*, CAM Report 98-48 (UCLA, 1998).

[K06] Kittel, J., et al.: Bruchpilot promotes active zone assembly, Ca^{2+} channel clustering, and vesicle release. Science **312**, 1051–1054 (2006)

[K97] Kiwiel, K.: Proximal minimization methods with generalized bregman functions. SIAM J. Cont. Optim. **35**, 1142 – 1168 (1997)

[K00] Klar, T.A., et al.: Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. PNAS **97**, 8206–8210 (2000)

[LCA07] Le, T., Chartrand, R., Asaki, T.J.: A variational approach to reconstructing images corrupted by Poisson noise. J. Math. Imaging Vision **27**, 257–263 (2007)

[L08] D.A.Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, J. Inverse Ill-Posed Problems 16(2008), 463-478.

[LT08] D.A.Lorenz, D.Trede, *Optimal Convergence rates for Tikhonov regularization in Besov Scales*, Inverse Problems 24 (2008), 055010.

[L74] Lucy, L.B.: An iterative technique for the rectification of observed distributions. The Astronomical Journal **79**, 745–754 (1974)

[NW01] Natterer, F., Wübbeling, F.: Mathematical methods in image reconstruction. SIAM Monographs on Mathematical Modeling and Computation (2001)

[N97] A.Neubauer, *On converse and saturation results for Tikhonov regularization of linear ill-posed problems*, SIAM J. Numer. Anal. 34 (1997), 517527.

[OBG⁺05] Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation based image restoration. Multiscale Modelling and Simulation **4**, 460–489 (2005)

[PZG99] V.Y.Panin, G.L.Zeng, G.T.Gullberg, *Total variation regulated EM Algorithm*, IEEE Trans. Nucl. Sci. NS-46 (1999) 2202-2010.

[PM90] Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell **12**, 629–639 (1990)

[RS] S.Remmele, M.Seeland, J.Hesser, *Fluorescence microscopy deconvolution based on Bregman iteration and Richardson-Lucy algorithm with TV regularization*, Preprint (University Heidelberg, 2008).

[R05] E.Resmerita,*Regularization of ill-posed problems in Banach spaces: convergence rates*, Inverse Problems 21 (2005), 13031314.

[RS06] E.Resmerita, O.Scherzer, *Error estimates for nonquadratic regularization and the relation to enhancing*, Inverse Problems 22 (2006) 801-814.

[RA07] Resmerita, E., Anderssen, S.: Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems. Math. Meth. Appl. Sci **30**, 1527–1544 (2007)

[R07] Resmerita, E., et al.: The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis. Inverse Problems **23**, 2575–2588 (2007)

[RI72] Richardson, W.H.: Bayesian-based iterative method of image restoration. J. Opt. Soc. Am. **62**, 55–59 (1972)

[RLO03]   L.Rudin, P.L.Lions, S.Osher, *Multiplicative denoising and debluring: Theory and algorithms*, in: S.Osher, N.Paragios, eds., *Geometric Level Sets in Imaging, Vision, and Graphics* (Springer, New York, 2003), 103119.

[ROF92]   Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)

[SBW$^+$08]   Sawatzky, A., Brune, C., Wübbeling, F., Kösters, T., Schäfers, K., M., B.: Accurate EM-TV algorithm in PET with low SNR. In: IEEE Nucl. Sci. Symp. (2008)

[SG01]   Scherzer, O., Groetsch, C.: Inverse scale space theory for inverse problems, in: (Kerckhove, M., ed.) Scale-space and morphology in computer vision. Proc. 3rd Int. Conf. Scale-space, pp. 317-325. Springer, Berlin (2001)

[SV82]   Shepp, L.A., Vardi, Y.: Maximum likelihood reconstruction for emission tomography. IEEE Transactions on Medical Imaging **1**(2), 113–122 (1982)

[SO08]   J.Shi, S.Osher, *A nonlinear inverse scale space method for a convex multiplicative noise model*, SIAM Journal on Imaging Sciences (2008) 1(3):294321

[SHW93]   D.L.Snyder, A.M.Hammoud, R.L.White, *Image recovery from data acquired with a charge- coupled-device camera*, J. Opt. Soc. Amer. A 10 (1993), 1014-1023.

[SHL$^+$95]   D.L.Snyder,   C.W.Helstrom,   A.D.Lanterman, M.Faisal,   R.L.White,   *Compensation for readout noise in CCD images*, J. Opt. Soc. Amer. A 12 (1995), 272-283.

[VSK85]   Vardi, Y., Shepp, L.A., Kaufman, L.: A statistical model for positron emission tomography. J. of the American Statistical Association **80**(389), 8–20 (1985)

[V02]   C.Vogel, *Computational Methods for Inverse Problems* (SIAM, Philadelphia, 2002).

[WA04]   M.N.Wernick, J.N.Aarsvold, eds., *Emission Tomography: The Fundamentals of PET and SPECT*, (Academic Press, San Diego, 2004)

[WHM$^+$07]   Willig, K.I., Harke, B., Medda, R., Hell, S.W.: STED microscopy with continuous wave beams. Nature Meth. **4**(11), 915–918 (2007)

[W83]   Witkin, A.P.: Scale-space filtering. In: Proc. Int. Joint Conf. on Artificial Intelligence, Karlsruhe, pp. 1019–1023 (1983)