A Guide to the TV Zoo

Martin Burger and Stanley Osher

Abstract Total variation methods and similar approaches based on regularizations with ℓ^1-type norms (and seminorms) have become a very popular tool in image processing and inverse problems due to peculiar features that cannot be realized with smooth regularizations. In particular total variation techniques had particular success due to their ability to realize cartoon-type reconstructions with sharp edges. Due to an explosion of new developments in this field within the last decade it is a difficult task to keep an overview of the major results in analysis, the computational schemes, and the application fields. With these lectures we attempt to provide such an overview, of course biased by our major lines of research.

We are focusing on the basic analysis of total variation methods and the extension of the original ROF-denoising model due various application fields. Furthermore we provide a brief discussion of state-of-the art computational methods and give an outlook to applications in different disciplines.

1 Introduction

Reconstructing and processing images with appropriate edges is of central importance in modern imaging. The development of mathematical techniques that preserve or even favour sharp edges has become a necessity and created various interesting approaches. The two most succesful frameworks are two variational approaches: total variation models on the one hand and models with explicit edges

M. Burger (✉)
Institute for Computational and Applied Mathematics, University of Münster, Münster, Germany
e-mail: martin.burger@www.de

S. Osher
Department of Mathematics, UCLA, Los Angeles, USA

M. Burger et al. (eds.), *Level Set and PDE Based Reconstruction Methods in Imaging*, Lecture Notes in Mathematics 2090, DOI 10.1007/978-3-319-01712-9_1,

in the framework of Mumford and Shah [140] on the other. The latter usually lead to various difficulties in the analysis (cf. [138]) and numerical realization due to the explicit treatment of edges and arising nonconvexity, consequently they have found limited impact in practical applications beyond image segmentation. In these lectures notes we will discuss various developments for total variation methods, which can be formulated as convex variational methods. A whole zoo of approaches to the modelling, analysis, numerical solution, and applications has developed in the last two decades, through which we shall try to develop a guide.

The starting point of total variation (TV) methods has been the introduction of a variational denoising model by Rudin et al. [163], consisting in minimizing total variation among all functions within a variance bound

$$TV(u) \to \min_u \quad \text{subject to} \int_\Omega (u-f)^2 dx \leq \sigma^2. \tag{1}$$

Introducing a Lagrange multiplier λ it can be shown that this approach is equivalent to the unconstrained problem of minimizing

$$E_{ROF}(u) := \frac{\lambda}{2} \int_\Omega (u-f)^2 dx + TV(u), \tag{2}$$

in the following often referred to as the ROF model. In subsequent years this model was generalized for many imaging tasks and inverse problems and found applications in different areas.

We will provide a more detailed motivation for total variation regularization in a rather general setup in Sect. 2. In Sects. 3–5 we provide an overview of various aspects in the analysis of total variation regularization. In Sect. 6 we discuss the concepts of Bregman iterations and inverse scale space methods, which allow to compensate systematic errors of variational methods, e.g. contrast loss in the case of TV regularization, and gave another boost to research in this area in recent years. In Sects. 7 and 8 we discuss some variants of the models, with changes concerning the fidelity term in Sect. 7 and the regularization term in Sect. 8. In Sect. 9 we discuss some approaches for the numerical solution of the variational problems. Section 10 is devoted to geometric aspects of total variation minimization and the relaxation of segmentation problems into convex models for functions of bounded variation. We then proceed to applications, which we mainly incorporate as further links to literature in Sect. 11. Finally we present some open questions in the modelling and analysis of TV methods in Sect. 12.

2 The Motivation for TV and Related Methods

In the following we provide basic motivations for the general setup of TV methods with respect to forward operators, data fidelity terms, and regularization.

2.1 MAP and Penalized ML Estimation

In order to obtain a general variational model including total variation and similar penalties we resort to the Bayesian approach for computing solutions to an operator equation

$$Au = fv \tag{3}$$

in the presence of stochastic effects such as noise (cf. e.g. [110, 114, 129]). In a classical log-likelihood estimation technique one computes a solution by minimizing the negative log-likelihood of observing f under u, i.e.,

$$u_{LL} = \arg\min_u(-\log p(f|u)), \tag{4}$$

where $p(f|u)$ denotes an appropriate probability density for observing f given u. This can usually be identified with the probability density of the noise, e.g. in the frequently investigated case of additive noise

$$f = Au + \eta, \tag{5}$$

where η is a stochastic perturbation, i.e. noise, we find

$$p(f|u) = p_\eta(f - Au).$$

The computationally efficient part of the Bayesian approach is to compute the MAP (maximum a-posteriori probability) estimate, by

$$u_{MAP} = \arg\min_u(-\log p(u|f)). \tag{6}$$

The posterior probability density is obtained from Bayes formula

$$p(u|f) = \frac{p(f|u)\,p_0(u)}{\tilde{p}_0(f)}, \tag{7}$$

where p_0 denotes the prior probability for u and $\tilde{p}_0$ is the prior probability for f. Since the latter will only contribute a constant term in the minimization of the negative logarithm of the prior probability, $\tilde{p}_0(f)$ is not important for the MAP estimate. We can rewrite the MAP estimation in a similar form to log-likelihood estimation as

$$u_{MAP} = \arg\min_u(-\log p(f|u) - \log p_0(u)). \tag{8}$$

The negative logarithm of the prior probability density hence acts as a penalty or regularization functional, which creates a detailed link to total variation methods. Since the second term penalizes nonsmooth functions in addition to the logarithmic likelihood, this approach is also called penalized maximum likelihood (ML) method.

Since we assume that images of low total variation have a higher prior probability than those with high total variation, it seems obvious that p_0 should be a monotonously decreasing function of $TV(u)$. For related methods we of course need to replace TV by the appropriate prior, we consequently shall write the general functional J instead of TV. The natural choice in probability is a Gibbs form (cf. [94])

$$p_0(u) \sim e^{-\beta J(u)} \tag{9}$$

with some constant $\beta > 0$. We can then rewrite the MAP estimation as

$$u_{MAP} = \arg\min_u(-\log p(f|u) + \beta J(u)). \tag{10}$$

Rescaling to

$$\lambda H(u, f) := -\frac{1}{\beta} p(f|u) \tag{11}$$

where λ is a parameter depending on β and possibly the noise (see below), we see that MAP estimation yields a minimization problem of the form

$$\lambda H(u, f) + J(u) \to \min_u. \tag{12}$$

For specific imaging tasks and specific noise models the form of $H(u, f)$ can be written down, usually by straight-forward calculations. We shall now and in the following consider a linear image formation model $u \mapsto Au$ and a Gaussian additive noise η. We assume that A is a bounded linear operator to some Banach space. This means that for the "exact" image $\hat{u}$ the data are generated from

$$f = A\hat{u} + \eta, \tag{13}$$

where η follows a normal distribution with variance σ^2 and zero mean. This means that (formally)

$$p(f|u) = p_{Gauss}(f - Au) \sim \exp\left(-\frac{\|Au - f\|^2}{2\sigma^2}\right).$$

Hence we can define $\lambda := \frac{1}{\beta\sigma^2}$ and obtain the MAP estimation from the variational problem

$$\frac{\lambda}{2}\|Au - f\|^2 + J(u) \to \min_u. \tag{14}$$

We shall investigate non-Gaussian noise models leading to nonquadratic fidelity terms in the variational problem in Sect. 7.

2.2 Imaging Tasks

Above we have simply modelled image formation by a linear operator A mapping between some Banach (or ideally Hilbert) spaces. In our case the preimage space will consist of functions of bounded variation on a domain $\Omega \subset \mathbb{R}^d$ (usually $d = 2, 3$). In the following we give an overview how various imaging tasks can be formulated as such (cf. e.g. [65]):

- *Denoising:* In the case of denoising (with Gaussian noise) the given (noisy) image is measured as an element of the Hilbert space $L^2(\Omega)$. The clean image is usually modelled as an element of a smaller function space X, e.g. a Sobolev space $W^{1,p}(\Omega)$, a Besov space, or in our context usually $BV(\Omega) \cap L^2(\Omega)$. The operator A is an embedding operator into $L^2(\Omega)$ and hence in particular bounded.
- *Deblurring:* In the case of deblurring, A is modelled as an integral operator of the form

$$(Au)(x) = \int_\Omega k(x, y)u(y)\, dy \tag{15}$$

  in most cases with a convolution kernel $k(x, y) = \tilde{k}(x - y)$. For typical kernels A is a bounded operator from $L^1(\Omega)$ or $L^2(\Omega)$ to $L^2(\Omega)$.
- *Decompositon:* In an image decomposition model one aims to separate certain components of the image, e.g. smooth ones from texture. A typical model for decomposition into two parts is an operator

$$A : X_1 \times X_2 \to Y, \quad A(u_1, u_2) = u_1 + u_2. \tag{16}$$

  A possible choice of the output space $Y \supset X_i$ is $L^2(\Omega)$, but if one seeks decompositions also into oscillatory components such as texture a larger space such as $H^{-1}(\Omega)$ is needed.
- *Inpainting:* The aim of inpainting is to restore respectively extend an image given in $\Omega \setminus D$ into the inpainting domain D. Hence, the operator A is an embedding operator from a function space on $\Omega \setminus D$ into a function space on Ω, which is again linear and bounded.
- *Medical image reconstruction:* In medical imaging, the image formation model depends on the specific modality (cf. e.g. [141]). Examples are the Radon/X-ray transform (in computerized tomography and PET) or samples of the Fourier transform (in MRI), in both cases linear bounded operators. Certain modalities (e.g. electron and optical tomography) need to be modelled by nonlinear operators, which are still compact however.
- *Image segmentation:* In edge-based image segmentation one seeks a decomposition of the image into the smooth parts of the image in some function space and in particular an edge set $\Gamma \subset \Omega$ of finite $(n - 1)$-dimensional Hausdorff-measure. If the edge set is the boundary of some subset of Ω such as in object-based segmentation, it is often replaced by a function of bounded variation taking only the values zero (outside this subset) and one (inside).

2.3 Regularization Functionals

In the following we consider the variational model

$$E(u) = \frac{\lambda}{2}\|Au - f\|^2 + J(u) \tag{17}$$

and discuss some implications of different choices of J.

A simple choice of a regularization functional is a quadratic energy of the form

$$J(u) = \frac{1}{2}\int_\Omega |Du|^2\, dx, \tag{18}$$

where $D : X \to L^2(\Omega)^m$ is a linear operator, e.g. a gradient in $X = H^1(\Omega)$, $m = d$. We assume that A is a bounded linear operator on X. Then the optimality condition for (17) with regularization (18) becomes

$$0 = E'(u) = \lambda A^*(Au - f) + D^*Du. \tag{19}$$

Hence we see that for D^*D being invertible

$$u = \lambda(D^*D)^{-1}A^*(f - Au).$$

Consequently $u \in \mathcal{R}((D^*D)^{-1}A^*)$, the range of the smoothing operator $(D^*D)^{-1}A^*$, and hence the reconstructed image will be blurred. Consider e.g. the denoising case where $D = \nabla$ and $A : H^1(\Omega) \to L^2(\Omega)$ is the embedding operator, then

$$-\Delta u = \lambda(f - u) \in L^2(\Omega).$$

By elliptic regularity we thus expect $u \in H^2(\Omega)$, i.e. the image is oversmoothed and there are no edges in particular. On the other hand consider $X = L^2(\Omega)$, D being the identity, and A is an integral operator as in deblurring or image reconstruction. Then u is in the range of A^*, which is again an integral operator (with kernel $k(y, x)$). This again implies smoothness of u and the nonexistence of edges.

An obvious next step would be to replace the square in the regularization (18) by a smooth strictly convex function R, i.e. to choose

$$J(u) = \frac{1}{2}\int_\Omega R(Du)\, dx, \tag{20}$$

in (17). The corresponding optimality condition becomes

$$0 = E'(u) = \lambda A^*(Au - f) + D^*(R'(Du)). \tag{21}$$

The linearization of this equation is given by

$$\lambda A^* A v + D^*(R''(Du)Dv) = g. \tag{22}$$

If R is smooth and strictly convex, then the "diffusion" coefficient $R''(Du)$ is strictly positive and the linearized operator has analogous smoothing properties as (19). Consequently we need to expect a similar smoothing behaviour as in the quadratic case.

As a consequence one might argue that only a regularization functional based on R being not strictly convex can indeed prevent oversmoothing and should be able to maintain edges. As we shall see this is indeed the case for total variation. In order to gain a formal understanding it is instructive to consider $R(u) = \frac{1}{p}|u|^p$ for p approaching 1. The corresponding optimality condition becomes

$$0 = E'(u) = \lambda A^*(Au - f) + D^*(|Du|^{p-2}Du). \tag{23}$$

For $p > 1$ a large "gradient" Du will have an impact on the differential operator, since the Euclidean norm of $|Du|^{p-2}Du$ is $|Du|^{p-1}$. In the limit $p \to 1$ this behaviour changes and the involved vector field $|Du|^{-1}Du$ has Euclidean norm one no matter how large Du becomes. Hence, large (infinite) "gradients" Du will not have particular impact on $D^*(|Du|^{p-2}Du)$.

Visually the improvement when choosing $p = 1$ compared to larger values (e.g. $p = 2$) can be observed in Figs. 1 and 2. The noisy versions of the images are used as input f for two variational denoising methods, a linear scheme (on the lower left)

$$\frac{\lambda}{2}\int_\Omega (u - f)^2\, dx + \int_\Omega |\nabla u|^2\, dx \to \min_u \tag{24}$$

equivalent to

$$\lambda(u - f) - \Delta u = 0, \tag{25}$$

and the ROF model (on the lower right)

$$\frac{\lambda}{2}\int_\Omega (u - f)^2\, dx + \int_\Omega |\nabla u|\, dx \to \min_u . \tag{26}$$

In both cases, a visually optimal value of λ is chosen. The linear scheme leads to significant blurring and non-sharp images, while the ROF model succeeds in computing reconstructions with sharp edges, which seem much more appealing to the human eye (cf. [100] for a further interpretation of this effect). Comparing the ROF reconstructions in Figs. 1 and 2 one observes that the quality of the ROF solution is better in the case of the blocky image, which ideally corresponds to the properties of total variation as we shall also see below. In the case of a natural image the denoising is of lower visual quality (still clearly outperforming the

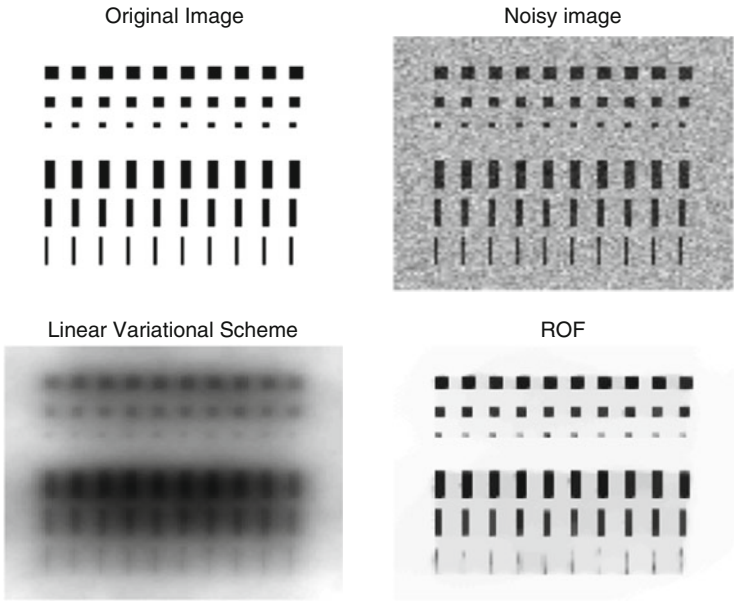

Fig. 1 A blocky image and its noisy version (*top*), reconstruction with linear scheme (quadratic regularization) and with the ROF model

linear scheme) due to various oscillatory patterns (texture) appearing in the image (compare e.g. the flower bush in the background). In such cases the ROF model still manages to reconstruct a cartoon of the image, i.e. its major structure without small-scale textures. Further techniques can subsequently be applied to find textures in the residual.

A similar reasoning holds for schemes enforcing sparsity. Suppose $(\psi_j)_{j\in\mathbb{N}}$ is an orthonormal basis of a separable Hilbert space X and

$$B : \ell^2(\mathbb{N}) \to X,\ \mathbf{c} = (c_j)_{j\in\mathbb{N}} \mapsto \sum c_j \psi_j \tag{27}$$

is the map from coefficients to elements in X. Then one can rephrase linear methods via the variational problem (Fig. 3)

$$\frac{\lambda}{2}\|AB\mathbf{c} - f\|^2 + \frac{1}{2}\sum_{j=1}^{\infty} \omega_j |c_j|^2, \tag{28}$$

with (ω_j) being a suitable sequence of weights (e.g. $\omega_j = 1$ corresponds to the original squared norm in X). The optimality condition for (28) becomes

$$\omega_j c_j = -(B^* A^* (AB\mathbf{c} - f)) \cdot e_j, \tag{29}$$

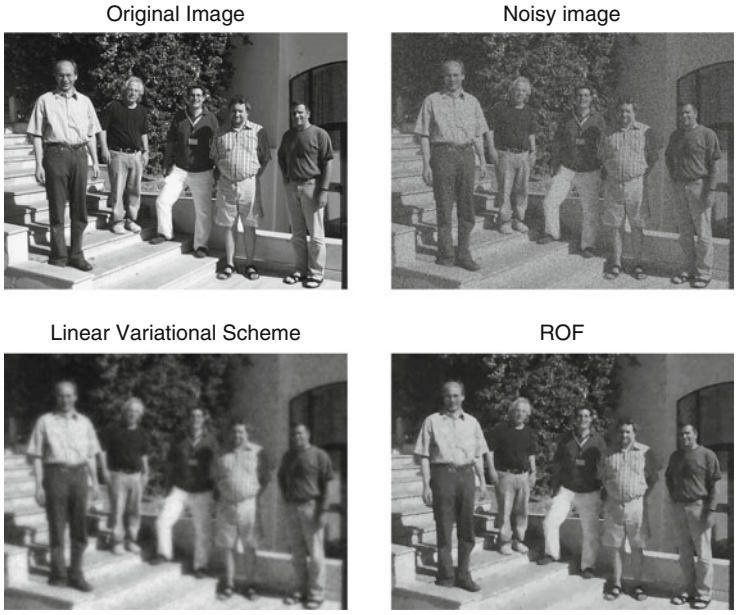

Fig. 2 A natural image (lecturers of the summer school) and its noisy version (*top*), reconstruction with linear scheme (quadratic regularization) and with the ROF model

where e_j is the j-th unit vector. Hence the decay of the coefficients c_j is mainly determined by the properties of the adjoint operator B^*A^* and the weights ω_j. There is no particular reason for achieving a sparse (in particular finite) weight sequence, one will rather encounter a slowly decaying one. If one generalizes the regularization to

$$\frac{\lambda}{2}\|AB\mathbf{c} - f\|^2 + \frac{1}{p}\sum_{j=1}^{\infty}\omega_j|c_j|^p, \tag{30}$$

the optimality condition becomes

$$\omega_j c_j |c_j|^{p-2} = -(B^*A^*(AB\mathbf{c} - f)) \cdot e_j. \tag{31}$$

For $p > 1$ there still exists a unique simple monotone relation between c_j and the right-hand side, so the decay behaviour is still dominated by ω_j and B^*A^*. Again the limit $p \to 1$ gives an interesting change, since for $c_j \to 0$ the limit of the left-hand side can be quite arbitrary. Thus the case $p = 1$ (or $p < 1$) will be particularly interesting also in such situations.

Fig. 3 From *left* to *right*: clean image, denoising with the isotropic ROF model with three different scale parameters λ (decreasing from *left* to *right*)

2.4 Total Variation and Variants

The total variation of a function is formally defined as

$$TV(u) = \int_\Omega |\nabla u| \, dx, \tag{32}$$

a definition which makes sense if $u \in W^{1,1}(\Omega)$. Since neither the structure of $W^{1,1}(\Omega)$ is very convenient (it is not the dual of a Banach space) nor it contains piecewise constant functions, it is reasonable to consider a slightly larger space which has the desired properties, namely the space of functions of bounded variation. An exact definition of the total variation is (cf. [95])

$$TV(u) := \sup_{\substack{g \in C_0^\infty(\Omega;\mathbb{R}^d) \\ ||g||_\infty \le 1}} \int_\Omega u \, \nabla \cdot g \, dx. \tag{33}$$

Here we use the vectorial norm

$$||g||_\infty := \operatorname*{ess\,sup}_{x \in \Omega} \sqrt{g_1(x)^2 + \ldots + g_d(x)^2}, \tag{34}$$

variants in the choice of the vector norm will be discussed below. The space of functions of bounded variation is then defined as

$$BV(\Omega) = \{ u \in L^1(\Omega) \mid TV(u) < \infty \}. \tag{35}$$

We shall later verify that BV has indeed the desired properties, see also [6, 86, 95] for detailed discussions of functions of bounded variation.

The total variation as defined so far was isotropic, in the sense that it is invariant with respect to rotations in the argument x. This is often an advantage, since the result becomes independent of the rotation of the input image. Variants of total variation are basically obtained by changing the norm of the vector g in (33) from the standard Euclidean norm to different vector norms

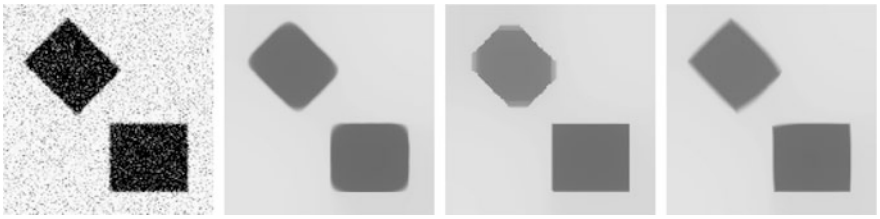

Fig. 4 From *left* to *right*: noisy image, denoising with the isotropic ROF model (Euclidean vector norm), denoising with the anisotropic ROF Model (supremum vector norm), and a ROF model with adapted anisotropy. From [30]

$$TV(u) := \sup_{\substack{g \in C_0^\infty(\Omega;\mathbb{R}^d) \\ \gamma(g) \leq 1 \text{ a.e.}}} \int_\Omega u \, \nabla \cdot g \, dx, \tag{36}$$

where $\gamma : \mathbb{R}^d \to \mathbb{R}$ is a nonnegative one-homogeneous functional satisfying a triangle-inequality. Of course the arising total variation norms are equivalent and give the same space $BV(\Omega)$ in the above definition. However, in the reconstruction the anisotropy of vector norms different from the Euclidean one can have a significant impact (cf. [30, 83]). While the isotropic total variation (based on the Euclidean norm) favours rounded edges in reconstructions, anisotropic definitions can favour different structures, e.g. corners or rectangular structures. This behaviour is illustrated in Fig. 4.

3 Existence, Uniqueness, and Optimality

In the following we investigate the well-posedness of the general model

$$\frac{\lambda}{2}\|Au - f\|^2 + TV(u) \to \min_{u \in BV(\Omega)}. \tag{37}$$

This problem has been analyzed first by Acar and Vogel [2], using strong L^p topologies that are suitable due to compact embedding results for $BV(\Omega)$. We shall here present a different analysis based on the weak* topology of $BV(\Omega)$. In order to avoid technicalities in the analysis we first eliminate the mean value of u, which later allows to define an equivalent norm on those functions with mean zero. We introduce the corresponding subspace

$$BV_0(\Omega) = \{ u \in BV(\Omega) \mid \int_\Omega u \, dx = 0 \}. \tag{38}$$

In the remainder of the paper we shall assume that $A\mathbf{1} \neq 0$, where $\mathbf{1}$ is the constant function taking the value 1 everywhere. We further assume

$$\langle Av, A\mathbf{1} \rangle = 0 \qquad \forall\, v \in BV_0, \tag{39}$$

which is true for most of the operators we consider (e.g. denoising, inpainting, deblurring, decomposition). The main results still hold true if (39) is not satisfied but with additional technical effort related to the mean value of u.

Lemma 3.1. *Let $\lambda > 0$. Then the minimizer of (37) is of the form*

$$u = v + \frac{\langle f, A\mathbf{1} \rangle}{\|A\mathbf{1}\|^2}\mathbf{1} \tag{40}$$

with $v \in BV_0(\Omega)$.

Proof. Each $u \in BV(\Omega)$ can be written as $u = v + c\mathbf{1}$ with mean value $c \in \mathbb{R}$ and $v \in BV_0(\Omega)$. Since $TV(u) = TV(v)$, we can rewrite the functional to be minimized using (39) as

$$\frac{\lambda}{2}\|Au - f\|^2 + TV(u) = \frac{\lambda}{2}\|cA\mathbf{1} - f\|^2 + \frac{\lambda}{2}\|Av\|^2 - \lambda\langle Av, f\rangle + TV(v).$$

Hence in order to compute the optimal c we can simply minimize $\|cA\mathbf{1} - f\|^2$ with respect to $c \in \mathbb{R}$, which yields the desired result. □

As a result of Lemma 3.1, we can now reduce the minimization over $BV_0(\Omega)$, shifting f to $f - \frac{\langle f, A\mathbf{1}\rangle}{\|A\mathbf{1}\|^2}A\mathbf{1}$. With abuse of notation we also call the shifted data f and consider the minimization

$$E(u) = \frac{\lambda}{2}\|Au - f\|^2 + TV(u) \to \min_{u \in BV_0(\Omega)}. \tag{41}$$

3.1 Coercivity

At the first glance we observe that the total variation is bounded if the functional in (41) is bounded. Thus we seek coercivity in the norm of $BV(\Omega)$. A first useful property of BV-spaces is embedding into Lebesgue spaces (cf. [86, 95]):

Lemma 3.2. *Let $\frac{q}{q-1} \geq d$, then*

$$BV_0(\Omega) \hookrightarrow L^q(\Omega).$$

The embedding is compact if $\frac{q}{q-1} > d$.

From Lemma 3.2 we see that in any dimension there exists $q > 1$ with $BV_0(\Omega) \hookrightarrow L^q(\Omega)$, even with compact embedding. If the operator A is defined on such an L^q-space, the analysis could be carried out in the strong topology of $L^q(\Omega)$, an approach used e.g. in [2]. We take a different approach using a different topology in $BV_0(\Omega)$ directly. The main result we shall use to verify coercivity of E is the following classical theorem (cf. e.g. [122]):

Theorem 3.3 (Banach-Alaoglu). *Let X be the dual of some Banach space Z. Then each bounded set in X is precompact in the weak-* topology.*

Hence we need to define a weak-* topology on $BV_0(\Omega)$, respectively find a space whose dual $BV_0(\Omega)$ is. For this sake we define a normed space

$$Z_0 = \{ \nabla \cdot g \mid g \in C_0^\infty(\Omega; \mathbb{R}^d) \}, \tag{42}$$

with norm (the norm properties can be checked in a straight-forward way)

$$\|p\|_Z = \inf_{g \in C_0^\infty(\Omega;\mathbb{R}^d), \nabla \cdot g = p} \|g\|_{L^\infty}. \tag{43}$$

Its completion in this norm is denoted by

$$Z := \overline{Z_0}. \tag{44}$$

We now find $BV_0(\Omega)$ as the dual of Z:

Proposition 3.4. *$BV_0(\Omega)$ can be identified with the dual space of Z defined by* (44)

Proof. First of all we observe that for each $u \in BV_0(\Omega)$ we can construct a linear functional in Z_0 given by

$$\ell_u : p \mapsto \int_\Omega up \, dx.$$

For $p = \nabla \cdot g_p$ we have

$$|\ell_u(p)| = |\int_\Omega u \nabla \cdot g_p \, dx| \leq \|g_p\|_{L^\infty} \sup_{g \in C_0^\infty(\Omega;\mathbb{R}^d), \|g\|_{L^\infty} \leq 1} \int_\Omega u \nabla \cdot g \, dx = \|g_p\|_{L^\infty} \, TV(u).$$

Taking the infimum over all such g_p we conclude

$$|\ell_u(p)| \leq \|p\|_Z \, \|u\|_{BV}.$$

Thus, ℓ_u is bounded on Z_0 and can therefore be extended in a unique way to a bounded linear functional on Z. Moreover, for $u_1 \neq u_2$ it is easy to see that $\ell_{u_1} \neq \ell_{u_2}$, otherwise

$$TV(u_1 - u_2) = \sup_{g \in C_0^\infty(\Omega;\mathbb{R}^d), \|g\|_{L^\infty} \leq 1} (\ell_{u_1}(g) - \ell_{u_2}(g)) = 0.$$

This implies that we can identify $BV_0(\Omega)$ with a subspace of Z^*.

On the other hand we see that for q_* sufficiently large we have $L_0^{q*}(\Omega) \hookrightarrow Z$, where

$$L_0^{q*}(\Omega) = \{ v \in L^{q*}(\Omega) \mid \int_\Omega v \, dx = 0 \}.$$

This can be shown e.g. by regularity results for the homogeneous Neumann problem for the Poisson equation (note that functions in $L_0^{q*}(\Omega)$ satisfy the solvability criterion). Thus we obtain the opposite inclusion for the dual spaces, $Z^* \subset L_0^q(\Omega)$ with $q = \frac{q_*}{q_*-1}$. Hence, each functional $\ell \in Z^*$ can be identified with a functional of the form

$$\ell(p) = \int_\Omega v \, p \, dx$$

for some $v \in L_0^q(\Omega)$. The boundedness of the linear functional ℓ further implies $TV(v) < \infty$, hence $v \in BV_0(\Omega)$. Thus, we can also identify Z^* with a subspace of $BV_0(\Omega)$. □

In order to complete our analysis of coercivity, we need to replace the BV-norm by the seminorm, i.e., the total variation, since only the latter is bounded on sub-level sets of E. Hence we prove a Poincare-Wirtinger inequality on $BV_0(\Omega)$:

Lemma 3.5. *The total variation is an equivalent norm on $BV_0(\Omega)$.*

Proof. It is obvious that

$$\|u\|_{BV} = \|u\|_{L^1} + TV(u) \geq TV(u)$$

and it remains to show that there exists $c > 0$ with $\|u\|_{BV} \leq c TV(u)$ for all $u \in BV_0(\Omega)$. We prove this assertion by contradiction and hence assume that for each $n \in N$ there exists $u_n \in BV_0(\Omega)$ with $\|u_n\|_{BV} > n \, TV(u_n)$. Now let $v_n := \frac{u_n}{\|u_n\|_{BV}}$, then $TV(v_n) < \frac{1}{n}$, i.e. $TV(v_n) \to 0$. Since $\|v_n\|_{BV} = 1$, we conclude $\|v_n\|_{L^1} \to 1$. Since v_n is uniformly bounded in $BV_0(\Omega)$, there exists a weak-* convergent subsequence v_{n_k} with some limit v. Due to the compact embedding of $BV_0(\Omega)$ into $L^1(\Omega)$, this sequence converges strongly in $L^1(\Omega)$ and thus

$$\|v\|_{L^1} = \lim_k \|v_{n_k}\|_{L^1} = 1.$$

By the lower semicontinuity of the total variation in the weak-* topology (see below) we further conclude

$$TV(v) \leq \liminf_k TV(v_n) = 0,$$

hence $TV(v) = 0$, which implies that v is constant and since $v \in BV_0(\Omega)$ even that $v \equiv 0$. The latter contradicts $\|v\|_{L^1} = 1$. □

We are finally able to state the main coercivity result as a direct consequence of the properties shown above:

Lemma 3.6. *The sub-level set*

$$\mathcal{M}_C = \{ u \in BV_0(\Omega) \mid E(u) \leq C \} \tag{45}$$

is precompact in the weak- topology of* $BV_0(\Omega)$.

We mention that the weak-* topology in BV always implies strong convergence in $L^1(\Omega)$, hence we can work with sequences in all arguments in the following.

3.2 Lower Semicontinuity

Besides coercivity, a major ingredient is the lower semicontinuity of the functional E in (41). Since a sum of two functionals is lower semicontinuous if both of them are, we verify this property separately, starting with the total variation:

Proposition 3.7. *The total variation is weak-* lower semicontinuous on* $BV_0(\Omega)$.

Proof. Let $u_n \rightharpoonup^* u$ and let $\psi_k \in C_0^\infty(\Omega; \mathbb{R}^d)$ with $\|\psi_k\|_{L^\infty} \leq 1$ such that

$$\int_\Omega u \nabla \cdot \psi_k \, dx \to TV(u).$$

Then we have due to the weak-* convergence

$$\begin{aligned} \int_\Omega u \nabla \cdot \psi_k \, dx &= \lim_n \int_\Omega u_n \nabla \cdot \psi_k \, dx \\ &= \liminf_n \int_\Omega u_n \nabla \cdot \psi_k \, dx \\ &\leq \liminf_n TV(u_n). \end{aligned}$$

Taking now the limit $\psi_k \to \infty$ we find

$$TV(u) = \lim_k \int_\Omega u \nabla \cdot \psi_k \, dx \leq \liminf_n TV(u_n),$$

hence TV is weak-* lower semicontinuous. □

For the fitting functional we need to make a basic assumption on the operator A in order to obtain weak-* lower semicontinuity, namely a range property of the adjoint operator

$$\mathcal{R}(A^*) \subset Z. \tag{46}$$

Note that for A being a bounded linear operator on $BV_0(\Omega)$, $\mathcal{R}(A^*)$ is always a subset of the dual space (cf. [122]) of $BV_0(\Omega)$. The additional assumption that $\mathcal{R}(A^*)$ lies in the smaller predual space is a regularity assumption on the operator A.

Lemma 3.8. *For $\lambda > 0$ the functional $u \mapsto \lambda \|Au - f\|^2$ is weak-* lower semicontinuous on $BV_0(\Omega)$.*

Proof. Since a positive multiple and a square preserve lower semicontinuity of a positive functional it suffices to show that $u \mapsto \|Au - f\|$ is weak-* lower semicontinuous. By the dual characterization of a Hilbert space norm (cf. [122]) we have

$$\|Au - f\| = \sup_{\varphi, \|\varphi\|=1} \langle Au - f, \varphi \rangle.$$

Now we can again choose a sequence φ_k with norm equal one such that $\langle Au - f, \varphi_k \rangle \to \|Au - f\|$. Since $A^*\varphi_k \in Z$ we can employ weak-* convergence of u_n to see

$$\begin{aligned}
\langle Au - f, \varphi_k \rangle &= \langle u, A^*\varphi_k \rangle - \langle f, \varphi_k \rangle \\
&= \lim_n \langle u_n, A^*\varphi_k \rangle - \langle f, \varphi_k \rangle \\
&= \liminf_n \langle u_n, A^*\varphi_k \rangle - \langle f, \varphi_k \rangle \\
&= \liminf_n \langle Au_n - f, \varphi_k \rangle \\
&\leq \liminf_n \|Au_n - f\|.
\end{aligned}$$

Again with the limit $k \to \infty$ we obtain weak-* lower semicontinuity. □

3.3 Existence

With the above prerequisites we can now show the existence of a minimizer of E by a standard variational technique:

Theorem 3.9. *There exists a minimizer of E in $BV_0(\Omega)$, i.e. a solution of (41).*

Proof. The proof technique is a standard combination of coercivity and lower semicontinuity, which we give here for completeness. Since

$$+\infty > \frac{\lambda}{2}\|f\|^2 = E(0) \geq \inf_u E(u) \geq 0 > -\infty,$$

the infimum of E is finite and hence there exists a minimizing sequence u_k with $E(u_k) \to \inf_u E(u)$. For k sufficiently large, this minimizing sequence lies in the sub-level set $E(u) \leq C$ with $C = E(0) + 1$, which is precompact in the weak-* topology. Hence u_k has a weak-* convergent subsequence, again denoted by u_k without loss of generality. If $\hat{u}$ is the weak-* limit of u_k, then the lower semicontinuity implies

$$E(\hat{u}) \leq \liminf_k E(u_k) = \inf_u E(u),$$

thus $\hat{u}$ is a solution of (41). □

3.4 Uniqueness

The uniqueness of (41) is related to the convexity of the functional. Since both the quadratic fitting term and the total variation are convex, (41) is a convex variational problem, hence the following general result applies (cf. [81]).

Lemma 3.10. *The set of minimizers of* (41) *is convex.*

As a consequence we see that we can at least uniquely pick a certain solution by a different criterion, e.g. by minimizing some squared norm of the set of minimizers. An obvious candidate is the solution of minimal L^2-norm among all minimizers of (41). Since in this case a strictly convex functional is minimized over a convex set, the solution is unique.

A general uniqueness result can only be formulated if A is injective:

Theorem 3.11. *Let A have trivial nullspace and $\lambda > 0$. Then* (41) *has a unique minimizer.*

Proof. If A has trivial nullspace, then the second variation of $F(u) = \frac{\lambda}{2}\|Au - f\|^2$ is given by

$$F''(u)(v, v) = \lambda\|Av\|^2,$$

which is positive for $v \neq 0$. Hence the functional is strictly convex, which implies that the minimizer is unique. □

3.5 Optimality Conditions

In order to obtain optimality conditions for the variational problem (41) we need a general notion of derivative of the functional E since the total variation is not differentiable in a classical sense. However, since the total variation is convex, we can adopt the multivalued notion of a *subdifferential* (cf. [81]). The subdifferential of a convex functional $J : X \to \mathbb{R} \cup \{+\infty\}$ at $u \in X$ (here with $X = BV_0(\Omega)$ is given by

$$\partial J(u) = \{\ p \in X^* \mid \langle p, v - u\rangle \leq J(v) - J(u), \forall\ v \in X\ \}. \tag{47}$$

A subgradient $p \in \partial J(u)$ can be identified with the slope of a plane (of codimension one) in $X \times \mathbb{R}$ through $(u, J(u))$ that lies under the graph of J.

From the definition of the subdifferential it is straight-forward to see that $\hat{u}$ is a minimizer of J if and only if $0 \in \partial J(\hat{u})$. Due to convexity the first-order optimality condition is not only necessary, but also sufficient. It thus remains to compute the subdifferential of E to characterize minimizers of (41). Since the quadratic fitting term is Frechet-differentiable, we can decompose the subdifferential of E into this derivative and the subdifferential of the total variation (cf. [81]). Thus, the optimality condition for a minimizer u of E becomes

$$\lambda A^*(Au - f) + p = 0, \quad p \in \partial TV(u). \tag{48}$$

It thus remains to characterize the subdifferential of the total variation, which is a rather difficult task. We start with a general property of one-homogeneous functional (i.e. $J(tx) = |t|J(x)$ for all $t \in \mathbb{R}$)

Lemma 3.12. *Let $J : X \to \mathbb{R} \cup \{+\infty\}$ be a convex one-homogeneous functional. Then*

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

Proof. Using $v = 0$ and $v = 2u$ in the definition of a subgradient we find

$$-\langle p, u\rangle \leq -J(u)$$

and

$$\langle p, u\rangle \leq J(2u) - J(u) = 2J(u) - J(u) = J(u).$$

Thus, $\langle p, u\rangle = J(u)$ and the assertion follows. □

In the case of the total variation we see that for each subgradient the dual norm is bounded by

$$\|p\| = \sup_{v \in BV_0(\Omega), TV(v)=1} \langle p, v \rangle \leq \sup_{v \in BV_0(\Omega), TV(v)=1} TV(v) = 1.$$

Hence, we see that

$$\partial TV(u) = \{\, p \in BV_0(\Omega)^* \mid \|p\| \leq 1, \langle p, u \rangle = TV(u) \,\}. \tag{49}$$

From the structure of the dual space of $BV_0(\Omega)$ we see that for each $p \in BV_0(\Omega)^*$ with $\|p\| \leq 1$ there exists $g \in L^\infty(\Omega; \mathbb{R}^d)$ with $p = \nabla \cdot g$ (the opposite is not true in general). Hence, the optimality condition can be stated equivalently as

$$\lambda A^*(Au - f) + \nabla \cdot g = 0, \quad \|g\|_{L^\infty} \leq 1, \quad \langle \nabla \cdot g, u \rangle = TV(u). \tag{50}$$

For detailed investigations of subgradients of the total variation and properties of the dual space and its norm (often called G-norm) we refer to [15, 134, 143, 146, 179].

4 Examples

In the following we present some examples that can be computed exactly in order to provide further insight into the behaviour of total variation regularization in general and the ROF-functional (2). In this way one obtains further insight into structural properties of solutions, but also into remaining deficiencies of total variation regularization such as staircasing and loss of contrast. Since solutions cannot be found in closed form in general, the only way to obtain exact solutions for some functions f is to guess the form of the solution u and verify that $\lambda(f - u)$ is a subgradient of the total variation at u. Of course some general structural results, in particular for piecewise constant data (cf. [58, 162]) and numerical experiments can be a good guideline to find such solutions. The solutions that can be found in the easiest way are related to eigenvalues, more precisely to the nonlinear eigenvalue inclusion problem $\mu u \in \partial TV(u)$, which has been discussed in [26, 27] and effectively used in [134] for a celebrated example of an exact solution of the ROF functional, namely for Ω being the whole $\mathbb{R}^d$ and

$$f = \chi_{B_R(0)} = \begin{cases} 1 & \text{if } |x| < R \\ 0 & \text{else.} \end{cases} \tag{51}$$

The solution of this problem is given by

$$u = \left(1 - \frac{2}{\lambda R}\right)^+ \chi_{B_R(0)}, \tag{52}$$

i.e. u is proportional to f and thus also the subgradient p, hence the solution is based on an eigenfunction of the total variation. For further exact solutions we refer to [27, 178].

Note that for each function $f \in BV(\Omega)$ satisfying $\mu f \in \partial TV(f)$ for some $\mu \geq 0$ we can explicitly construct the unique minimizer of the ROF functional, which is of the form $u = cf$ for some $c \geq 0$. With this form the optimality condition gives for $\mu < \lambda$

$$cf - f + \frac{\mu}{\lambda} f = 0,$$

hence $c = 1 - \frac{\mu}{\lambda}$ yields a solution. If $\mu \geq \lambda$ then we can choose $c = 0$ to obtain the solution, which follows from a general results shown below: For λ sufficiently small (with explicit bound depending on f) $u \equiv 0$ is always the unique minimizer.

We shall in the following work on a bounded domain, for simplicity considering $d = 1$ and $\Omega = [0, 1]$, which allows some elementary computations, since the dual norm of p is simply $\|P\|_\infty$ with P being the primitive of p with $P(0) = 0$. We do not only provide the explicit examples but also check the computations verifying them, since this might be useful for research on other ℓ^1-type regularizations in the future. Multidimensional examples can be constructed in a rotationally symmetric setup by completely analogous methods.

4.1 Indicator Functions

We start with an example related to indicator functions of subintervals $[a, b]$ related to Meyer's example (cf. [134]) with input data from (51). Since we are considering a bounded image domain and not the whole space we need to subtract a normalizing constant to achieve $\int_\Omega f \, dx = 0$, i.e.

$$f = \alpha \chi_{[a,b]} - \alpha(b - a). \tag{53}$$

We are looking for a subgradient $p = \mu f$, which is characterized by $\|P\|_\infty \leq 1$ and

$$TV(f) = 2\alpha(1 + b - a) = \langle p, f \rangle = \mu \int_0^1 f(x)^2 \, dx = \mu\alpha^2(1 - b + a)(b - a)(1 + b - a), \tag{54}$$

which can be solved for μ to obtain

$$\mu = \frac{2}{\alpha(b - a)(1 - b + a)}. \tag{55}$$

The primitive is a piecewise linear function given by

$$P(x) = \begin{cases} -\mu\alpha(b-a)x & \text{if } x \le a, \\ -\mu\alpha(b-a)x + \mu\alpha(x-a) & \text{if } a < x < b, \\ -\mu\alpha(b-a)x + \mu\alpha(b-a) & \text{if } x \ge b. \end{cases}$$

The absolute value of P will attain its maximum at $x = a$ or $x = b$, hence $\|P\|_\infty \le 1$ is equivalent to the two conditions

$$\frac{2a}{1-b+a} \le 1 \qquad \text{and} \qquad \frac{2-2b}{1-b+a} \le 1, \tag{56}$$

which is obviously satisfied for $a = 1 - b$, i.e. the interval $[a, b]$ is placed symmetrically. Hence, any function of this type is a solution of the nonlinear eigenvalue problem. We also see the scale aspect in the eigenvalue μ, the small eigenvalues (large scales) are related to a high product $\alpha(b-a)$, which somehow characterizes the size of the step in f (the additional factor $1 - b + a$ can be interpreted as the effect of the finite domain).

4.2 Staircasing

It has been observed frequently in numerical experiments that total variation methods inherit staircasing phenomena, i.e. they often tend to produce stair-like step functions instead of smoothly increasing ones for 1D signals, respectively blocky structures for 2D images. In order to obtain a better understanding why staircasing is favoured, we consider two explicit examples in the following. Unfortunately it is not possible to construct an analytical example of noise that exhibits staircasing due to Lipschitz regularity results for minimizers of the ROF functional (cf. [59]), but at least we can give some hints in the right direction. Note also that is was believed for a long while that real staircasing in the sense of piecewise constant regions separated by discontinuities is a typical structure in ROF-denoising, however the recent results in [59] rather indicate that the piecewise constant regions are often rather separated by steep continuous parts instead.

We start with a stair-like structure

$$f = \alpha \sum_{k=1}^{N-1} \chi_{(k/N,1]} - \alpha \frac{N-1}{2}. \tag{57}$$

for N being an even number. The total variation of f equals $\alpha(N-1)$ and thus the a subgradient $p = \mu f$ needs to satisfy

$$\alpha(N-1) = \mu \int f^2\,dx = \frac{2\mu}{N} \sum_{k=0}^{N/2-1} \left(\frac{\alpha(N-1}{2} - k\alpha \right)^2 = \frac{\mu\alpha^2}{12}(2N^2-2N-1),$$

which yields

$$\mu = \frac{12(N-1)}{\alpha(2N^2-2N-1)}. \tag{58}$$

The primitive is given by

$$P(x) = \mu\alpha \sum_{k=1}^{N-1} \left(x - \frac{k}{N} \right)^+ - \mu\alpha \frac{N-1}{2} x,$$

and $|P|$ attains its maximum at $x = \frac{1}{2}$ as

$$|P(\frac{1}{2})| = \mu\alpha \sum_{k=1}^{N/2} \left(\frac{1}{2} - \frac{k}{N} \right) - \mu\alpha \frac{N-1}{4} = \mu\alpha \frac{N}{8} = \frac{3N(N-1)}{2(2N^2-2N-1)}.$$

Thus $\|P\|_\infty \leq 1$ and hence μf is a subgradient, i.e. a solution of the eigenvalue problem.

The second example is a linearly growing function

$$f(x) = \alpha(x - \frac{1}{2}), \tag{59}$$

which due to staircasing we do not expect to solve the eigenvalue problem. The candidate for a minimizer of the ROF functional is of the form

$$u(x) = \max\{-\beta, \min\{\beta, f(x)\}\}. \tag{60}$$

We thus have to verify that $p = \lambda(f-u)$ is a subgradient in $\partial TV(u)$. First of all we compute β from the relation

$$2\beta = TV(u) = \int_0^1 pf\,dx = \lambda\alpha\beta \left(\frac{1}{2} - \frac{\beta}{\alpha} \right)^2,$$

which yields

$$\beta = \alpha \left(\frac{1}{2} - \sqrt{\frac{2}{\lambda\alpha}} \right)^+.$$

The primitive $|P|$ attains its maximum at $x = \frac{1}{2} - \frac{\beta}{\alpha}$ with $|P(x)| = 1$. Thus $p = \lambda(f - u) \in \partial TV(u)$ and hence u is a minimizer of the ROF model, and hence f is partly replaced by piecewise constant structures.

5 Asymptotics and Stability Estimates

In the following we discuss some asymptotic properties in terms of the Lagrange parameter λ as well as stability estimates with respect to λ and the data f. Such asymptotics have first been investigated by Acar and Vogel [2], error estimates are due to [47, 72].

5.1 Asymptotics in λ

We shall now discuss the asymptotic behaviour of minimizers of (41) with respect to λ. For simplicity we assume that A has trivial nullspace, hence the minimizer is unique, and we shall denote the functional E and the minimizer for specific λ as E^λ and u^λ, respectively. The simplest asymptotic for the variational problem (41) is the one for $\lambda \to 0$, for which of course we expect the minimization of the total variation as the asymptotic problem:

Theorem 5.1. *For $\lambda \downarrow 0$ the minimizers of (41) satisfy $u^\lambda \to 0$ in the strong topology of $BV_0(\Omega)$.*

Proof. By the minimization property of u^λ we have

$$\frac{\lambda}{2}\|Au^\lambda - f\|^2 + TV(u^\lambda) = E^\lambda(u^\lambda) \leq E^\lambda(0) = \frac{\lambda}{2}\|f\|^2.$$

Hence $TV(u^\lambda)$ converges to zero (even with order λ). Since TV is an equivalent norm on $BV_0(\Omega)$ the convergence is also in the norm topology. □

The above convergence result and its proof can be applied for various regularization functionals, not only for total variation. A specific feature of total variation (and other ℓ^1-type functionals) is that the convergence to zero arises already for finite λ. The key is to investigate the norm in the dual space of BV, which can be characterized as

$$\|p\|_* = \inf_{g \in L^\infty(\Omega;\mathbb{R}^d), \nabla \cdot g = p} \|g\|_\infty. \tag{61}$$

Meyer [134] showed that

Theorem 5.2. *Let $\lambda\|A^* f\|_* \leq 1$. Then $u \equiv 0$ is a minimizer of (41).*

Proof. Since

$$\partial J(0) = \{\ p \in BV^* \mid \|p\|_* \leq 1\ \}$$

we obtain

$$p = 0 - \lambda A^* f \in \partial J(0),$$

hence 0 satisfies the optimality condition and is a minimizer. □

More involved and also more interesting is the asymptotic for $\lambda \to \infty$. Since the fitting term dominates in this asymptotic one expects convergence towards a solution of $Au = f$, which can however only be true if $f \in \mathcal{R}(A)$. In this case convergence can be verified in the weak-* topology as we shall see below. Before we provide a general monotonicity property:

Proposition 5.3. *The map $\lambda \mapsto \|Au^\lambda - f\|$ is nonincreasing and the map $\lambda \mapsto TV(u^\lambda)$ is nondecreasing. If $A^* f \neq 0$, then the decrease respectively increase are strict for λ sufficiently large.*

Proof. Let $\mu > \lambda$, then from the definition of u^λ and u^μ as minimizers of the regularized functional we find

$$\frac{\lambda}{2}\|Au^\lambda - f\|^2 + TV(u^\lambda) \leq \frac{\lambda}{2}\|Au^\mu - f\|^2 + TV(u^\mu)$$

and

$$\frac{\mu}{2}\|Au^\mu - f\|^2 + TV(u^\mu) \leq \frac{\mu}{2}\|Au^\lambda - f\|^2 + TV(u^\lambda).$$

By adding these inequalities and simple rearrangement we find $\|Au^\mu - f\| \leq \|Au^\lambda - f\|$. Dividing the first inequality by λ and the second by μ and subsequent analogous comparison yields

$$TV(u^\lambda) \leq TV(u^\mu).$$

We finally show the strict decrease (increase of TV) by contradiction. Assuming $\|Au^\mu - f\| = \|Au^\lambda - f\|$, we conclude from the first two inequalities $TV(u^\lambda) \leq TV(u^\mu)$. Hence u^λ is also a minimizer for the functional with parameter μ and from the optimality conditions we see that there exist $p, q \in \partial TV(u^\lambda)$ such that

$$0 = \lambda A^*(Au^\lambda - f) + p = \mu A^*(Au^\lambda - f) + q,$$

which implies

$$p = \frac{\lambda}{\mu} q.$$

Since both p and q are subgradients of the total variation, we have

$$TV(u^\lambda) = \langle q, u^\lambda \rangle = \frac{\lambda}{\mu} \langle q, u^\lambda \rangle,$$

which can only hold for $u^\lambda = 0$ since $\frac{\lambda}{\mu} < 1$. $u^\lambda = 0$ can only be a solution if

$$A^* f = \frac{1}{\lambda} p.$$

Since $\|p\| \leq 1$ for a subgradient, this can only be true if $\|A^* f\| \leq \frac{1}{\lambda}$, which contradicts $A^* f \neq 0$ if λ is sufficiently large. □

For exact data, i.e. f in the range of A, we can show the following weak convergence result:

Theorem 5.4. *Let $f \in \mathcal{R}(A)$, then every subsequence of (u^λ) has a weak-* convergent subsequence. Every weak-* accumulation point is a solution of $Au = f$ with minimal total variation.*

Proof. Let $\hat{u} \in BV_0(\Omega)$ be an arbitrary solution of $Au = f$. Then by the definition of u^λ we have

$$\frac{\lambda}{2} \|Au^\lambda - f\|^2 + TV(u^\lambda) \leq TV(\hat{u}).$$

Hence $\|Au^\lambda - f\|$ is of order $\lambda^{-1/2}$ and converges to zero, while $TV(u^\lambda) \leq TV(\hat{u})$ is uniformly bounded. From the latter we infer the existence of a weak-* convergent subsequence. For such a weak-* convergent sequence, which we denote by $u_k := u^{\lambda_k}$ with limit $\overline{u}$ we find from the lower semicontinuity of the fitting term

$$\|A\overline{u} - f\| \leq \liminf_k \|Au_k - f\| \leq \liminf_k \sqrt{\frac{2\, TV(\hat{u})}{\lambda}} = 0$$

and from the lower semicontinuity of the total variation

$$TV(\overline{u}) \leq \liminf_k TV(u_k) \leq TV(\hat{u}).$$

Thus, $\overline{u}$ is a solution of $Au = f$ with minimal total variation. □

In the more relevant case of noisy data, i.e. f being a perturbation of $g \in \mathcal{R}(A)$, the convergence has to be interpreted relative to the noise, more precisely the estimated variance of the noise,

$$\sigma^2 := \|f - g\|^2. \tag{62}$$

As usual for ill-posed problems (cf. [82]) convergence is achieved only if λ does not converge to infinity too fast relative to the decrease of the noise level.

Theorem 5.5. *Let $g \in \mathcal{R}(A)$ and f^σ be noisy data with variance σ^2, i.e. satisfying (62). Moreover, let δ_σ be an estimate of the noise variance such that there exists $c > 0$ with*

$$\delta_\sigma \geq c\sigma \qquad \forall\, \sigma > 0. \tag{63}$$

Let the Lagrange parameter be chosen such that $\lambda = \lambda(\delta_\sigma) \to \infty$ as $\delta_\sigma \to 0$ such that $\lambda(\delta_\sigma)\delta_\sigma^2 \to 0$. Then every subsequence of (u^λ) has a weak- convergent subsequence and every weak-* accumulation point is a solution of $Au = g$ with minimal total variation.*

Proof. In the following we always write λ meaning $\lambda(\delta_\sigma)$ for brevity. Let $\hat{u}$ be any solution of $Au = g$. From the definition of the minimizer we have

$$\frac{\lambda}{2}\|Au^\lambda - f^\sigma\|^2 + TV(u^\lambda) \leq \frac{\lambda}{2}\|f^\sigma - g\|^2 + TV(\hat{u}).$$

Using the noise variance and the properties of the estimator we find

$$\frac{\lambda}{2}\|Au^\lambda - f^\sigma\|^2 + TV(u^\lambda) \leq \frac{\lambda}{2}\sigma^2 + TV(\hat{u}) \leq \frac{\lambda}{2c}\delta_\sigma^2 + TV(\hat{u}).$$

Since the right-hand side converges to $TV(\hat{u})$, the total variation of u^λ is uniformly bounded and $\|Au^\lambda - f^\sigma\|$ converges to zero, from which the convergence properties can be infered as in the proof of Theorem 5.4. □

5.2 *Stability Estimates*

Error estimation for variational models like the ROF-model was an open problem in imaging and inverse problems for a rather long period, which is due to the difficulty to find an appropriate error measure. Intuitively it becomes obvious quite soon that the total variation norm (or seminorm) cannot be an appropriate measure, since it penalizes small visual differences too strongly.

Moreover, the above asymptotic analysis gives only weak-* convergence, but no strong convergence in $BV_0(\Omega)$, hence also error estimation seems out of reach. In [47] a generalized Bregman distance was introduced as an error measure, which allowed to derive suitable estimates consistent with the norm estimates in the case of penalization with a squared norm. The estimation in Bregman distances has been widely accepted since then and extended to various other situations (cf. [27, 28, 88, 98, 99, 107, 127, 160, 161]). We shall here mainly recall the results obtained in [47], and start by defining Bregman distances.

Definition 5.6. Let $J : X \to \mathbb{R} \cup \{+\infty\}$ be a convex functional. For each $p \in \partial J(u)$,

$$D_J^p(v,u) = J(v) - J(u) - \langle p, v-u\rangle \tag{64}$$

is a generalized Bregman distance between u and v. Moreover, for $q \in \partial J(v)$

$$D_J^{q,p}(v,u) = \langle p-q, u-v\rangle \tag{65}$$

is called a symmetric Bregman distance between u and v.

The Bregman distance is not a distance in the classical sense, since it is not symmetric in its original form, and it is not strict, i.e. $D_j^p(v,u) = 0$ is possible if $v \neq u$. For the total variation in particular contrast changes are not detectable in the Bregman distance, i.e. for a smooth monotone map F the functions u and $F(u)$ have zero distance (for any choice of subgradients). On the other hand differences in edges can be measured well in Bregman distances for the total variation (cf. [47, 50]), which is of particular interest for total variation techniques.

The stability of the variational problem in terms of the data f is given by the following result:

Theorem 5.7. *Let u_i, $i = 1, 2$ be the minimizers of (41) with data f_i. Then there exists $p_i \in \partial TV(u_i)$ such that the estimate*

$$\lambda \|Au_1 - Au_2\|^2 + 2D_{TV}^{p_1,p_2}(u_1,u_2) \leq \lambda \|f_1 - f_2\|^2 \tag{66}$$

holds.

Proof. Since the u_i are minimizers, they satisfy the optimality condition

$$\lambda A^*(Au_i - f_i) + p_i = 0, \qquad p_i \in \partial TV(u_i).$$

Subtracting them and taking the duality product with $u_1 - u_2$ we find

$$\lambda\langle A^*(Au_1 - Au_2), u_1 - u_2\rangle + \langle p_1 - p_2, u_1 - u_2\rangle = \lambda\langle A^*(f_1 - f_2), u_1 - u_2\rangle.$$

Using the definition of the adjoint we can rewrite the first term on the left-hand side as $\lambda\|Au_1 - Au_2\|^2$ and we estimate the right-hand side as

$$\begin{aligned}\langle A^*(f_1 - f_2), u_1 - u_2\rangle &= \langle f_1 - f_2, A(u_1 - u_2)\rangle \leq \|f_1 - f_2\|\|Au_1 - Au_2\| \\ &\leq \frac{1}{2}\|f_1 - f_2\|^2 + \frac{1}{2}\|Au_1 - Au_2\|^2.\end{aligned}$$

Inserting these relations and multiplying by two we obtain (66). □

Besides stability at fixed λ, quantitative estimates concerning the asymptotics in λ are of interest. Due to the ill-posedness of the operator equation with A additional conditions on the solution are needed. In particular the existence of a Lagrange parameter for the constrained problem

$$TV(u) \to \min_{u \in BV_0(\Omega), Au = g}, \tag{67}$$

characterizing the solution of minimal total variation, is a standard regularity assumption. This can be formulated as the source condition (cf. [47, 72])

$$A\hat{u} = g, \qquad \hat{p} = A^* w \in \partial TV(\hat{u}). \tag{68}$$

Theorem 5.8. *Let u^λ be the minimizer of E and let $\hat{u}$, $\hat{p}$ satisfy (68). Then there exists $p \in \partial TV(u^\lambda)$ such that the estimate*

$$\frac{\lambda}{2}\|Au^\lambda - A\hat{u}\|^2 + D_{TV}^{p,\hat{p}}(u^\lambda, \hat{u}) \le \frac{1}{\lambda}\|w\|^2 + \lambda\|g - f\|^2. \tag{69}$$

holds.

Proof. We start from the optimality condition for u^λ and subtract $(\hat{p} + \lambda A^* g)$ on both sides to have

$$\lambda A^*(Au^\lambda - A\hat{u}) + p - \hat{p} = -\hat{p} + \lambda A^*(f - g).$$

The duality product with $u^\lambda - \hat{u}$ yields

$$\lambda\|Au^\lambda - A\hat{u}\|^2 + D_{TV}^{p,\hat{p}}(u^\lambda, u) = -\langle \hat{p}, u^\lambda - \hat{u}\rangle + \lambda\langle f - g, Au^\lambda - A\hat{u}\rangle.$$

For the first term on the right-hand side we insert (68) and apply Young's inequality to obtain

$$-\langle \hat{p}, u^\lambda - \hat{u}\rangle = -\langle w, Au^\lambda - A\hat{u}\rangle \le \frac{\lambda}{4}\|Au^\lambda - A\hat{u}\|^2 + \frac{1}{\lambda}\|w\|^2,$$

and for the second term we directly apply Young's inequality, which finally yields the desired estimate. □

The error estimate (69) consists of two parts, a decreasing term of order $\frac{1}{\lambda}$ with a constant depending on the source condition (thus on the smoothness of $\hat{u}$) and a second term simply depending on the noise variance. Balancing the terms gives an indication how to choose the optimal λ at given noise, which is of course not directly possible in practice since $\|w\|$ is not known.

6 Bregman Iterations and Scale Spaces

Although the ROF-model (2) had great success in applied imaging, there is still one deficiency remaining, which is loss of contrast in the reconstruction compared to the original image. This aspect was highlighted by Meyer [134], who showed that an application of the ROF-model (2) to the characteristic function of a ball (f from (51)) results in a shrunk version as the minimizer, with the shrinkage proportional to $\frac{1}{\lambda}$ (cf. Sect. 4). A fundamental reason for the loss of contrast is a systematic error of variational regularization methods, since in the case of exact data $A\hat{u} = f$ we obviously have

$$J(u) \leq J(\hat{u}) - \frac{\lambda}{2}\|Au - f\|^2 < J(\hat{u})$$

for (17), since typically $Au \neq f$. For J being the total variation, this means that the total variation of the reconstruction is smaller than the total variation of the exact solution, which results in the above mentioned contrast loss.

An approach to overcome this issue has been presented in [148], we here start with a derivation in the case of the ROF-model (2). Assume that u_1 is the minimizer of (2) for fixed λ chosen rather to oversmooth (hence there should be no significant noise part in u_1). Then the main deficiency will be that the residual

$$v_1 = f - u_1 \tag{70}$$

still contains too much signal—actually enough to restore quite detailed information about the original image as demonstrated in [117]. Since this part is obviously shrunk too strongly by the ROF-model one might try and add the "noise" back to the image, so that in another run of the minimization this compensates the loss of contrast. This means a new image u_2 is computed from minimizing

$$J_2(u) := \frac{\lambda}{2}\int_\Omega (u - f - v_1)^2 dx + TV(u). \tag{71}$$

This minimization can be interpreted in an alternative way if we take a look at the optimality condition for (2), which together with the definition of v_1 implies

$$p_1 = \lambda v_1 = \lambda(f - u_1) \in \partial TV(u_1). \tag{72}$$

Hence, we can rewrite J_2 as

$$J_2(u) = \frac{\lambda}{2}\int_\Omega (u - f)^2 dx + TV(u) - \langle p_1, u\rangle + \langle p_1, f + v_1\rangle. \tag{73}$$

Ignoring the constant term $\langle p_1, f + v_1 \rangle$ we observe that minimizing J_2 is equivalent to minimize the fitting functional penalized by the Bregman distance to the last iterate u_1 (with subgradient p_1).

As a consequence of the above reasoning we can generalize the one-step contrast correction introduced via (73) to an iterative scheme. We start with u_0 and $p_0 \in \partial TV(u_0)$ and then subsequently compute a sequence u_k via

$$u_{k+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left[\frac{\lambda}{2} \int_\Omega (u - f)^2 dx + D_{TV}^{p_k}(u, u_k) \right], \quad p_k \in \partial TV(u_k). \tag{74}$$

This iterative refinement scheme introduced and analyzed in [148] turns out to be equivalent to the so-called *Bregman iteration* introduced in [37] for quite general, but continuously differentiable functionals. The more general form of the Bregman iteration for computing a minimizer of

$$J(u) \to \min_u \quad \text{subject to } u \in \operatorname*{arg\,min}_v H(v, f) \tag{75}$$

is given by

$$u_{k+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left[H(u, f) + D_J^{p_k}(u, u_k) \right], \quad p_k \in \partial J(u_k). \tag{76}$$

The problem with convex fitting term $H(., f)$ can be generalized to convex and possibly nondifferentiable J, but in this case some regularity of H is still needed to guarantee well-definedness of the iteration and suitable properties of subgradients.

6.1 Interpretations of the Bregman Iteration

In the following we further discuss different interpretations of the Bregman iteration in the case of a quadratic fitting functional (Gaussian noise) and general convex regularization, i.e., we consider

$$J(u) \to \min_u \quad \text{subject to } u \in \operatorname*{arg\,min}_v \frac{\lambda}{2} \|Av - f\|^2, \tag{77}$$

where A is a linear operator. The corresponding Bregman iteration is defined by

$$u_{k+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left[\frac{\lambda}{2} \|Au - f\|^2 + D_J^{p_k}(u, u_k) \right], \quad p_k \in \partial J(u_k). \tag{78}$$

We observe that the optimality condition for the minimizer in (78) yields

$$\lambda A^*(Au_{k+1} - f) + p_{k+1} = p_k, \tag{79}$$

which can also be seen as an update rule for the dual variable. Such a dual interpretation of the Bregman iteration will be derived in the following.

Primal-Dual Iteration

If the data f lie in the domain of A, then we can write (77) equivalently as

$$J(u) \to \min_u \quad \text{subject to } Au = f. \tag{80}$$

The associated Lagrangian for this problem is

$$L_0(u, w) = J(u) + \langle w, Au - f \rangle \tag{81}$$

which can be augmented without changing minimizers to

$$L(u, w) = J(u) + \frac{\lambda}{2}\|Au - f\|^2 + \langle w, Au - f \rangle \tag{82}$$

and the optimality conditions for a saddle point (if it exists, which is a regularity property not automatic for A being compact) are given by

$$\begin{aligned} 0 &\in \partial J(u) + \lambda A^*(Au - f) + A^* w = \partial_u L(u, w), \\ 0 &= Au - f = \partial_w L(u, w). \end{aligned}$$

A primal dual scheme iterates on both optimality conditions, e.g. by using a quadratic penalty on w. This yields

$$\begin{aligned} 0 &\in \partial J(u_{k+1}) + \lambda A^*(Au_{k+1} - f) + A^* w_k, \\ 0 &= Au_{k+1} - f = \tau(w_{k+1} - w_k). \end{aligned}$$

From the first relation we see that w_{k+1} is related to the dual variable in the Bregman iteration by $p_{k+1} = -A^* w_{k+1}$. With $\lambda = \frac{1}{\tau}$ and applying A^* to the second relation, we see that this primal-dual scheme is exactly equivalent to the Bregman iteration, cf. (79). The iteration in the variable w_k can be interpreted as a generalization of the original motivation (70), it describes the update of the residuals. It also delivers an alternative way to realize the Bregman iteration as in (71), the minimizer can be computed from minimizing

$$u_{k+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left[\frac{\lambda}{2}\|Au - f + \frac{1}{\lambda} w_k\|^2 + J(u) \right], \tag{83}$$

such that again the problem can be realized as the original variational model just changing the input data.

The primal-dual interpration also induces an alternative to the Bregman iteration, by considering an Uzawa-type iteration with an explicit treatment of the constraint

$$0 \in \partial J(u_{k+1}) + \lambda A^*(Au_{k+1} - f) + A^* w_k,$$
$$0 = Au_k - f = \tau(w_{k+1} - w_k).$$

This scheme is equivalent to the iteration

$$u_{k+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left[\lambda \langle Au, Au_k - f \rangle + D_J^{p_k}(u, u_k) \right], \tag{84}$$

which linearizes the fitting term and was hence called *linearized Bregman iteration* in [55–57] in applications to compressed sensing (J being an ℓ^1-norm). The scheme was also called *Landweber-type method* in an application to total variation methods (even allowing A nonlinear) in [20, 21]—due to analogies with classical iteration schemes for inverse problems.

Updated Bayesian Prior

In the Bayesian setting we computed the MAP estimate by maximizing the posterior probability density

$$p(u|f) \sim p(f|u)p_0(u). \tag{85}$$

The prior density is usually centered at zero, i.e., $u \equiv 0$ maximizes p_0. Given the knowledge from the solution of (17) one could however update the prior probability distribution. Instead of having a probability centered at zero, it makes more sense to use a probability at least locally centered at u_1, where u_1 is the solution of (17). A simple way to do so is to shift the log likelihood by an affinely linear term, namely the first-order Taylor approximation around u_1, in order to obtain a new prior log likelihood and probability density, respectively. The shifted prior probability

$$p_1(u) \sim \exp\left(\log p_0(u) - \log p_0(u_1) + \frac{p_0'(u_1)(u - u_1)}{p_0(u_1)} \right) \tag{86}$$

now has a maximum at u_1 instead of zero. This process of shifting the prior probability is repeated during the Bregman iteration.

In the case of total variation we even find that the prior probability p_1 has a maximum at each multiple of u_1, since $\partial TV(u_1) = \partial TV(cu_1)$ for any $c \in \mathbb{R}$. This means that the contrast is left competely open by the new prior probability, a reason why the Bregman iteration can improve upon contrast losses.

Dual Interpretation

If A^*A is invertible, the Bregman iteration can be interpreted as a dual ascent scheme, preconditioned with $(A^*A)^{-1}$. Defining the standard convex conjugate (dual functional)

$$J^*(p) = \sup_u [\langle p, u\rangle - J(u)], \tag{87}$$

we find $u_{k+1} \in \partial J^*(p_{k+1})$ under appropriate conditions, since $p_{k+1} \in \partial J(u_{k+1})$. Hence, (79) can be rewritten as

$$(A^*A)^{-1}(p_{k+1} - p_k) \in \lambda A^* f - \lambda \partial J^*(p_{k+1}) \tag{88}$$

Hence, the Bregman iteration in the dual variable computes

$$p_{k+1} = \arg\max_p \left(-\frac{1}{2\lambda} \|(A^*A)^{-1/2}(p_{k+1} - p_k)\| + \langle p, A^* f\rangle - J^*(p) \right). \tag{89}$$

This is just a proximal point algorithm (with special norm) for the dual problem

$$\langle p, A^* f\rangle - J^*(p) \to \max_p . \tag{90}$$

A similar relation holds for the linearized Bregman iteration, in this case we find

$$(A^*A)^{-1}(p_{k+1} - p_k) \in \lambda A^* f - \partial \lambda J^*(p_k), \tag{91}$$

i.e., we recover a dual (sub)gradient ascent scheme.

Geometric Interpretation for Total Variation

In the special case of J denoting the total variation we can add a further interpretation based on a geometric viewpoint. The formal version of the Bregman distance is

$$D_J^p(v, u) = \int_\Omega (|\nabla u| - \frac{\nabla v}{|\nabla v|} \nabla u)\, dx = \int_\Omega (\frac{\nabla u}{|\nabla u|} - \frac{\nabla v}{|\nabla v|}) \cdot \nabla u\, dx. \tag{92}$$

The difference $\frac{\nabla v}{|\nabla v|} - \frac{\nabla u}{|\nabla u|}$ can be interpreted as the difference in normals to level sets or edge sets. Hence the Bregman distance measures the alignment of level and edge sets, and in the iteration (76) there is no further penalty for those edge sets aligned with the ones of the previous iterates. In particular there is no penalty if the edge sets are aligned, but the height of the jump is different. This allows the Bregman iteration to correct contrast losses inherent in (2). The idea of matching normal fields has also

been used previously, e.g. in pan-sharpening of color and spectral images (cf. [22]), the connection to the Bregman distance has been established recently (cf. [136]).

6.2 Convergence Analysis

In the following we provide a convergence analysis of the Bregman iteration (78). For simplicity we assume $u_0 = 0$ and $p_0 = 0 \in \partial J(u_0)$, an extension to arbitrary initial values satisfying the consistency condition $p_0 \in \partial J(u_0)$ is quite obvious. Our basic assumption is again the existence of an exact solution $\hat{u}$ satisfying $A\hat{u} = g$ and f is a noisy version of the exact data. We do not further discuss the well-definedness of the iterates u_k, since due to (83) this issue is equivalent to existence and uniqueness of the original variational problem for arbitrary data, which has been discussed above. Besides monotone decrease of the residual, the fundamental property of the Bregman iteration is that the Bregman distance to the exact solution is decreasing up to some point, with a dissipation governed by the least-squares functional.

Lemma 6.1. *Let u_k be a sequence generated by the Bregman iteration (78). Then the inequalities*

$$\|Au^{k+1} - f\| \leq \|Au^k - f\| \tag{93}$$

and

$$D_J^{p_{k+1}}(\hat{u}, u_{k+1}) + \frac{\lambda}{2}\|Au_{k+1} - f\|^2 + D_J^{p_k}(u_{k+1}, u_k) \leq D_J^{p_k}(\hat{u}, u_k) + \frac{\lambda}{2}\|g - f\|^2 \tag{94}$$

hold for all $k \geq 0$.

Proof. Due to the positivity of the Bregman distance and since $D_J^{p_k}(u_k, u_k) = 0$, we directly obtain (93) from comparing the functional values in (78) for the minimizer u_{k+1} and for u_k. We further have

$$\begin{aligned} D_J^{p_{k+1}}(\hat{u}, u_{k+1}) - D_J^{p_k}(\hat{u}, u_k) + D_J^{p_k}(u_{k+1}, u_k) \\ &= \langle p_{k+1} - p_k, u_{k+1} - \hat{u} \rangle \\ &= -\lambda \langle Au_{k+1} - f, Au_{k+1} - g \rangle \\ &= -\lambda \|Au_{k+1} - f\|^2 + \lambda \langle Au_{k+1} - f, g - f \rangle \\ &\leq -\frac{\lambda}{2}\|Au_{k+1} - f\|^2 + \frac{\lambda}{2}\|g - f\|^2, \end{aligned}$$

where we have inserted the optimality condition (12) for $p_{k+1} - p_k$. □

Of particular importance for the convergence of the scheme is the boundedness of the total variation, which will again yield weak-* convergence of subsequences. In the case of noisy data the uniform bound can only be guaranteed for a finite number of iterations, which can be used determine the stopping index:

Lemma 6.2. *Let u_k with subgradients $p_k \in \partial J(u_k)$ be a sequence generated by (78), then*

$$J(u_m) \leq 5J(\hat{u}) + 2\lambda m \|f - g\|^2 \tag{95}$$

for all $m \in \mathbb{N}$.

Proof. Summing (94) and (79) we obtain

$$D_J^{p_m}(\hat{u}, u_m) + \frac{\lambda}{2} \sum_{k=1}^{m} \|Au_k - f\|^2 \leq D_J^{p_0}(\hat{u}, u_0) + \frac{\lambda m}{2} \|g - f\|^2$$

respectively

$$p_m = -\lambda \sum_{k=1}^{m} A^*(Au_k - f).$$

From the first inequality we obtain using nonnegativity of $D_J^{p_m}(\hat{u}, u_m)$ that

$$\|Au_m - f\|^2 \leq \frac{1}{m} \sum_{k=1}^{m} \|Au_k - f\|^2 \leq \frac{2}{\lambda m} J(\hat{u}) + \|g - f\|^2$$

and from the identity for p_m we estimate

$$\begin{aligned} J(u_m) - J(\hat{u}) \leq \langle p_m, u_m - \hat{u} \rangle &= -\lambda \sum_{k=1}^{m} \langle Au_k - f, Au_m - g \rangle \\ &\leq \lambda \sum_{k=1}^{m} \|Au_k - f\| \left(\|Au_m - f\| + \|f - g\| \right) \\ &\leq \lambda \sum_{k=1}^{m} \|Au_k - f\|^2 + \frac{\lambda m}{2} \left(\|Au_m - f\|^2 + \|f - g\|^2 \right) \\ &\leq 4J(\hat{u}) + 2\lambda m \|f - g\|^2, \end{aligned}$$

which yields the assertion. □

Hence in the case of exact data ($f = g$) we have $J(u_m) \leq 5J(\hat{u})$ and thus uniform boundedness of the iterates. Thus we have the following result by standard reasoning as above:

Theorem 6.3 (Convergence for Exact Data). *Let u_k with subgradients $p_k \in \partial J(u_k)$ be a sequence generated by (78) and let $f = g$. Then there exists a weak-* convergent subsequence and every weak-* accumulation point solves $Au = g$. If A has trivial nullspace, then the sequence u_k converges to the unique solution $\hat{u}$ of $A\hat{u} = g$ in the weak-* topology.*

In the case of noisy data, boundedness is obtained if $m\|f - g\|^2 \leq C$ for some constant C. This provides a criterion for the choice of a stopping index k_* in dependence of the noise level $\|f - g\|$:

Theorem 6.4 (Semi-Convergence for Noisy Data). *Let u_k^f with subgradients $p_k^f \in \partial J(u_k^f)$ be a sequence generated by (78) for specific noisy data f and let $k_* = k_*(f)$ be a stopping index satisfying*

$$k_*(f)\|f - g\|^2 \leq C \tag{96}$$

for some constant $C > 0$. Then, if f_n is a sequence with $\|f_n - g\| \to 0$, there exists a weak- convergent subsequence of $\{u_{k_*(f_n)}^{f_n}\}$ and every weak-* accumulation point solves $Au = g$. If A has trivial nullspace, then the sequence $u_{k_*(f_n)}^{f_n}$ converges to the unique solution $\hat{u}$ of $A\hat{u} = g$ in the weak-* topology.*

The convergence analysis can be refined further: In [148] an a-posterior stopping via the discrepancy principle, i.e.,

$$k_*(f) = \min\{\, k \mid \|Au_k^f - f\| \leq \tau\|g - f\| \,\} \tag{97}$$

for $\tau > 1$ fixed, has been shown to yield convergence. In [50] error estimates for the Bregman iteration have been derived under the source conditions discussed above.

6.3 Inverse Scale Space Methods

In general it seems favourable to choose a rather small λ in the Bregman iteration, i.e. each iteration step is actually oversmoothing. In this way one slowly iterates to a point where a stopping criterion is satisfied and should obtain a reasonable reconstruction at this point of stopping. If λ is too large one might end up with a final reconstruction of bad quality just since the last iteration step went too far. The most extreme case for large λ could be that the Bregman iteration stops after only one step, i.e. at the solution of (17) for large λ—hence it cannot lead to an improvement. This reasoning indicates to study the limit $\lambda \to 0$ with a simultaneous increase of

iteration steps, i.e. $t = k\lambda$ being constant. We can observe the result of this limit by rewriting (79) as

$$\frac{p(t+\lambda) - p(t)}{\lambda} = A^*(f - Au(t+\lambda)), \tag{98}$$

where we define $u(k\lambda) := u_k$ and $p(k\lambda) := p_k$. We see that the Bregman iteration is just the implicit Euler discretization of a flow, the so-called *inverse scale space flow*

$$\partial_t p = A^*(f - Au), \quad p \in \partial J(u). \tag{99}$$

This flow has been derived in [49] and called *inverse scale space flow*, since it coincides in the case of quadratic regularization with the inverse scale space methods in [169]. The wording inverse is due to the opposite behaviour to classical scale space methods or diffusion filters (cf. [194]). While scale space methods start with the noisy image and smooth increasingly with proceeding time (coarsening of scales), inverse scale space methods start with the coarsest scale (e.g. the mean value of the image intensity) and roughen the image with increasing time (refinement of scales). Inverse scale space methods are computationally more involved than classical diffusion filtering techniques, but there are several advantages: First of all, inverse scale space methods can be applied to very general imaging tasks such as deblurring or other inverse problems, while scale space techniques are restricted mainly to denoising-type problems. Secondly, appropriate stopping rules can be derived for inverse scale spaces such as the discrepancy principle

$$t_*(f) = \inf\{\, t \mid \|Au(t) - f\| \le \tau \|g - f\| \,\}. \tag{100}$$

Finally, the quality of reconstructions obtained with inverse scale space methods seems to be better compared to scale space methods in particular in the case of total variation and ℓ^1-regularization, both visually and with respect to some error measures (cf. [49, 53]).

Here we shall neither provide details on the analysis of the well-posedness of the flow (cf. [51] for the challenging case of total variation) nor on the convergence analysis (cf. [49]) or error estimates (cf. [50]), which are analogous to the case of Bregman iterations.

It is interesting to consider a relaxation of the inverse scale space method introduced in [49], with an auxiliary variable w:

$$\epsilon \partial_t u = \gamma A^*(f - Au) + w - p \tag{101}$$

$$\partial_t w = A^*(f - Au) \tag{102}$$

$$p \in \partial J(u) \tag{103}$$

with small parameters ϵ and γ. In the limit $\epsilon \to 0$ and $\gamma \to 0$ one expects that $w - p$ tends to zero, and hence one recovers the inverse scale space flow. A detailed analysis of this scheme has been given in [126]. An advantage for the numerical realization is the fact that the relaxed inverse scale space method consists of two evolution equations, which can be integrated more efficiently than the original inverse scale space flow.

The application of Bregman iterations and inverse scale space methods to wavelet denoising has been investigated in [200]. It is well-known that the variational problem analogous to the ROF functional (cf. [62])

$$\frac{\lambda}{2}\int_\Omega (\sum c_j \psi_j - f)^2 + \sum |c_j| \to \min_{(c_j)}, \tag{104}$$

with $\{\psi_j\}$ denoting the wavelet basis, results into soft-thresholding (also called soft shrinkage, cf. [79]). The inverse scale space method instead yields a hard-thresholding formula (with thresholding parameter related to $\frac{1}{t}$) and Bregman iterations yields firm thresholding (cf. [93]), an intermediate between hard and soft thresholding.

6.4 Linearized Bregman Methods

The realization of the Bregman iteration still needs the solution of variational problems of similar structure as the total variation regularized least-squares problem. Since one expects small steps anyway, it seems reasonable to approximate the fidelity. The simplest approximation is a linearization of the first term, which yields the linearized Bregman iteration

$$u_{k+1} = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left[H(u_k, f) + \partial_u H(u_k, f)(u - u_k) + D_J^{p_k}(u, u_k) \right] \tag{105}$$

with $p_k \in \partial J(u_k)$. In the case of a quadratic fidelity, the linearization yields the optimality condition

$$\lambda A^*(Au_k - f) + p_{k+1} = p_k. \tag{106}$$

One observes that the operator A and its adjoint only need to be evaluated in order to carry out the linearized Bregman iterations, which makes the scheme particularly attractive for applications where A is nonlocal, e.g. a convolution operator or even a more complicated image formation model.

A possible disadvantage, which indeed is observed in all numerical tests, is that (106) is not solvable if J is not strictly convex, since in this case the functional to be minimized in (105) is not bounded below. Therefore a multiple of a squared norm is

usually added to the regularization in such cases in order to make the scheme work, e.g. in the case of total variation

$$J(u) = \frac{\kappa}{2}\|u\|_{L^2}^2 + |u|_{TV}. \tag{107}$$

Ideally one tries to choose κ small in order to approximate pure total variation regularization, which can indeed be achieved. Thus a similar behaviour of reconstructions as in the original Bregman iteration is obtained, with a significant improvement of computational efficiency if A can be evaluated easily. A key observation for the realization of efficient computational methods is that the subproblem in each step of such a linearized Bregman iteration can be rewritten as the minimizer of

$$J^k(u) = \frac{\kappa}{2}\|u - f^k\|_{L^2}^2 + |u|_{TV}, \tag{108}$$

with f^k precomputed from the previous iteration. Thus one just solves a TV-denoising problem via the ROF-model in each step. Recently the linearized Bregman method and variations were applied with some success to compressive sampling problems (cf. [55, 56]), with a regularization term of the form

$$J(u) = \frac{\kappa}{2}\|u\|_{L^2}^2 + |\langle u, \varphi_k\rangle|_{\ell^1}, \tag{109}$$

where (φ_k) is an orthonormal basis of some Hilbert space. In that case the subproblem in each step can be computed explicitely by thresholding of coefficients. The convergence analysis of linearized Bregman iterations can be found in [20, 21] in the total variation case, in [56] in the case of ℓ^1 regularization, and in [170] for regularization with powers of strictly convex norms in Banach spaces.

6.5 Total Variation Flow

Total variation flow (or TV flow) is the scale space version of total variation in the denoising case (cf. [38, 40, 41, 194]). It is obtained if the ROF-model (2) is considered as an implicit time discretization of a flow with time step $\tau = \frac{1}{\lambda}$ and iterated to compute $u(t + \tau)$ as a solution of

$$\frac{\lambda}{2}\int_\Omega (u - u(t))^2\, dx + TV(u). \tag{110}$$

From the optimality condition one can derive an evolution law in the form of the differential inclusion

$$\partial_t u(t) = -p(t), \qquad p(t) \in \partial TV(u(t)) \tag{111}$$

with initial value

$$u(0) = f. \tag{112}$$

An analysis of total variation flow can be given in two ways: In [7–9, 25] the evolution is interpreted as a degenerate parabolic problem and appropriate entropy solution techniques are constructed. The approach in [87] rather stays with the variational interpretation of (111) as a gradient flow for total variation and defines weak solutions via a variational inequality. The existence of the flow as well as numerical schemes are then analyzed using a smooth approximation of total variation. Uniqueness and large time-behaviour (cf. [87]), as well as stability estimates (cf. [50]), can be infered directly from the gradient flow interpretation. Note that for two solutions u_1 and u_2 of (111) with subgradients p_1 respectively p_2 we obtain using the duality product with $u_1 - u_2$

$$\frac{1}{2}\frac{d}{dt}\int_\Omega (u_1 - u_2)^2\, dx = \langle \partial_t u_1 - \partial_t u_2, u_1 - u_2\rangle = -\langle p_1 - p_2, u_1 - u_2\rangle.$$

The right-hand side equals a negative Bregman distance and is hence not positive, which directly yields that the flow is non-expansive in the L^2-norm. Further inspections of the behaviour of the dual variable p can then be used to derive error estimates also in the Bregman distance (cf. [50]).

Visually solutions obtained with TV flow behave similarly than those obtained from minimizing the ROF-model, in particular they suffer from the same systematic errors. A surprising result in [39] shows that they are actually the same in the case $d = 1$. In particular for small noise levels (for which one uses only a small final time in the flow), the solutions of (111) can be integrated very efficiently by explicit methods and hence outperform the variational and inverse scale space model with respect to computational speed. However, due to the systematic errors and the difficulties to choose a stopping time in a robust automatic way, some care is needed. For problems different from denoising the total variation flow and similar scale space methods cannot be used, since the construction would rely on the inversion of the operator A, which is neither stable nor computationally efficient in most typical imaging tasks.

6.6 Other Multiscale Approaches

A multiscale approach with similar appearance as Bregman iterations was introduced by Tadmor et al. [181], who constructed hierarchical decompositions based on the ROF-functional. The starting point is the usual decomposition

$$f = u_0 + v_0, \quad u_0 = \operatorname*{arg\,min}_{u\in BV(\Omega)} \left[\frac{\lambda_0}{2}\|u - f\|_{L^2}^2 + TV(u)\right], \tag{113}$$

typically with λ_0 small in order to obtain really a coarse scale in u_0. The residual v_0 respectively a further sequence of residuals v_j can now be further decomposed into

$$v_{j-1} = u_j + v_j, \quad u_j = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left(\frac{\lambda_j}{2} \|u - v_{j-1}\|_{L^2}^2 + TV(u) \right). \tag{114}$$

In order to guarantee a decrease of scales λ_j needs to be decreased, usually via the dyadic sequence

$$\lambda_j = 2\lambda_{j-1}. \tag{115}$$

From this scheme one obtains a decomposition in the form

$$f = \sum_{j=0}^{\infty} u_j, \tag{116}$$

where increasing index marks decreasing scales.

Variants of this approach have been investigated in [182], where rather general norms have been investigated, and in [185], where the fidelity term has been changed to a squared H^{-1}-norm. A time-continuous version analogous to inverse scale space flows has been investigated by Athavale and Tadmor [11].

7 Nonquadratic Fidelity

So far we have laid our attention on quadratic fidelity terms, i.e. Gaussian noise models in the Bayesian framework, since they allow an introduction to the analysis without too many technical issues. In practice one sometimes encounters non-quadratic fidelities instead, due to two reasons: non-Gaussianity of the noise model or / and nonlinearity of the image formation model. We will briefly give an overview on recent developments for those two cases, which are both still subject of very active research.

7.1 Non-Gaussian Noise

In several applications we find noise models different from the simple additive Gaussian assumptions. Such cases can still be treated in the MAP framework, since only $p(f|u)$ is to be changed. In the following we discuss some interesting cases and provide useful references.

Laplace Noise

Motivated by noise models with additive Laplacian distributed errors, various authors have considered models of the form

$$E(u) = \lambda \int_{\Sigma} |f - Au| \, dy + TV(u), \tag{117}$$

with $A : BV(\Omega) \to L^1(\Sigma)$ (cf. [70, 101]).

The existence of minimizers is analogous to the quadratic case considered above, but uniqueness remains unclear since the fidelity term is not strictly convex even for A having trivial nullspace. An interesting property of L^1 fidelities is the possibility of exact reconstruction. If the data f satisfy a source condition, then the reconstruction is exact for finite λ, i.e. for λ sufficiently large a minimizer of (117) satisfies $Au = f$ (cf. [28]). Bregman iterations have been generalized to the case of L^1-fidelity in [104], it turned out that well-posedness and convergence can only be achieved for λ changing during the iteration.

For data being indicator functions, the L^1-fidelity model has also interesting geometrical properties, first of all there is an indicator function solution again (cf. [70]) and the structure of the resulting edge set can be analyzed (cf. [188, 189]).

Poisson Statistics

Poisson-distributed data are the natural case for many image formation models, where one registers counts of photons (e.g. CCD-cameras, cf. [173, 174]) or positrons (e.g. PET, cf. [183, 197]). In the case of high count rates the Poisson distribution can be well approximated by Gaussians, but for lower count rates it becomes more appropriate to use variational models asymptotically derived directly from the Poisson statistics. In the case of denoising this yields variational problems of the form

$$E(u) = \lambda \int_{\Omega} \left(f \log \frac{f}{u} - f + u \right) dx + TV(u) \tag{118}$$

to be minimized subject to nonnegativity. The extension to image reconstruction problems is given by

$$E(u) = \lambda \int_{\Sigma} \left(f \log \frac{f}{Au} - f + Au \right) dy + TV(u) \tag{119}$$

with Σ denoting the measurement domain.

Such models have been proposed and tested already in the late nineties (cf. [113, 150]), motivated in particular by the bad noise statistics in PET. Due to the algorithms used at that time, the application for large regularization parameters

and effective implementation was not able at this time, which prevented significant practical impact despite the enormous potential in such applications. With improvements in the computational approaches such as splitting methods these nonlinear models recently again received strong attention and now seem to find their practice in microscopy and PET (cf. [23, 42, 44, 77, 123, 165]).

An interesting aspect of nonlinear fidelities motivated by the Poisson case (but actually more general) has been investigated in [44], namely the appropriate construction of Bregman iterations. It turns out that Bregman iterations can be constructed in a primal as well as in a dual fashion, which is equivalent in the case of quadratic fidelities, but produces different methods in the case of nonquadratic ones.

Multiplicative Noise

Another interesting class of noise models are multiplicative ones, i.e. $f = (A\hat{u})\,\eta$, which is found in several applications.

A Gaussian statistic for the noise has been assumed in [164], which leads to the nonconvex variational problem

$$\frac{\lambda}{2}\int_\Omega (\frac{f}{u}-1)^2\,dx + TV(u) \to \min_u, \quad \int_\Omega \left(\frac{f}{u}-1\right)\,dx = 0. \tag{120}$$

Aubert and Aujol [13] modeled the noise as Gamma distributed instead, leading to

$$\lambda\int_\Omega (\log\frac{Au}{f} + \frac{f}{Au})\,dx + TV(u) \to \min_u, \tag{121}$$

which is closer to the Poisson noise case and at least convex in a reasonable subset of BV functions. In the case of denoising a substitution to a new variable $v = \ln u$ is possible, which leads to a convex variational model if the total variation of v is used for regularization instead of the total variation of e^v (cf. [108]).

Other Fidelities

Several other fidelity terms have been investigated recently, with different motivations. Bonesky et al. [34] have investigated powers of Banach space norm, to which also inverse scale space methods and their analysis were generalized in [51].

Motivated by the examples mentioned above and the modeling of textures, Meyer [134] proposed to use the dual norm of BV as a fidelity, i.e. the variational problem

$$\lambda\|u-f\|_* + TV(u) \to \min_{u\in BV(\Omega)}, \tag{122}$$

for which also interesting theoretical results have been obtained. Due to difficulties in the numerical solution, several authors have considered other dual Sobolev norms instead (cf. [147, 185–187]).

Other data fidelities (and also a constraint to probability densities) have been considered in applications to density estimation (cf. [135, 144]). In [54] the Wasserstein metric has been used as a fidelity, and the arising regularized optimal transport has been analyzed.

7.2 Nonlinear Image Formation Models

Another motivation for considering nonquadratic fidelity terms are nonlinear image formation models as appearing e.g. in ultrasound imaging, optical tomography, or other inverse problems in partial differential equations. The main theoretical change in this case is to replace the linear forward operator A by a nonlinear one, which we denote by F in the following. The natural extension of the variational model is the minimization

$$\frac{\lambda}{2}\|F(u) - f\|^2 + J(u) \to \min_u . \tag{123}$$

Such an approach has been proposed and tested e.g. by Luce and Perez [131] in an application to parameter identification. Under standard conditions on the operator F, the existence analysis can be carried out as in the linear case (cf. [2]). A major theoretical difference to the previously discussed cases is that the fidelity can now become non-convex, hence there is no uniqueness in general and there might be undesired local minima.

Error estimates in the nonlinear case can be derived similar to the linear case [107, 161], however there are differences related to the conditions needed on the nonlinearity of the operator. Depending on the degree of linearization used, several variants of Bregman iterations can be constructed for nonlinear models (cf. [20, 21, 170]).

8 Related Regularization Techniques

In the following we discuss some extensions and variants of regularization with total variation, which can be used for several purposes in image processing and analysis.

8.1 Higher-Order Total Variation

In several applications it is interesting to obtain piecewise linear instead of piecewise constant structures. This can be achieved by the higher-order total variation

$$TV_2(u) := \sup_{\substack{g \in C_0^\infty(\Omega;\mathbb{R}^{d\times d}) \\ \|g\|_\infty \le 1}} \int_\Omega u \, (\nabla\cdot)^2 g \, dx, \tag{124}$$

or, alternatively by the variation defined via the Laplacian

$$TV_\Delta(u) := \sup_{\substack{g \in C_0^\infty(\Omega) \\ \|g\|_\infty \le 1}} \int_\Omega u \, \Delta g \, dx. \tag{125}$$

The second order total variation has been investigated and applied in [68, 105, 171, 176], duality properties have been investigated in [175]. The extension to higher orders is obvious, but so far hardly investigated, in particular in applications, probably also due to the difficulty of treating higher-order functionals.

The higher-order total variation has received some attention recently in combination with the total variation in inf-convolution functionals as proposed in [61],

$$ICTV(u) = \inf_{w \in BV_2(\Omega)} (TV(u-w) + TV_2(w)). \tag{126}$$

The idea of regularization with the inf-convolution is an optimal decomposition of the solution into a component in $BV(\Omega)$ and another one in

$$BV_2(\Omega) = \{ u \in BV(\Omega) \mid TV_2(u) < \infty \}.$$

In this way staircasing can be avoided (at least to some extent) since in smoothly varying regions the higher order total variation dominates. Clearly the staircasing then appears in the derivative or in other words the variational approach favours a combination of piecewise constant and piecewise linear structures. This can further be improved by considering inf-convolutions with even higher order total variations, of course at the expense of higher computational effort. A slight modification of the inf-convolution approach that further improves the quality of results has been proposed recently (cf. [29, 36, 171, 172]), a detailed analysis of the schemes in terms of underlying function spaces and detailed structure of solutions (e.g. eigenvalue problems as in the TV case above) remains an interesting open problem.

8.2 Total Variation of Vector Fields

So far we have only discussed methods for scalar-valued functions of bounded variation. In several cases one would however prefer to work with vector fields, e.g.

for color or spectral images (cf. [33, 69]) or flow fields (cf. [195]), or even tensors such as in DT-MRI techniques (cf. [73, 177]). The extension of total variation to vector valued functions is rather obvious, namely by taking the total variation of each component. However, there are some subtleties, in particular in the way the components are combined. Formally we may think of

$$TV(\mathbf{u}) = \int_\Omega \|(|\nabla u_1|, |\nabla u_2|, \dots, |\nabla u_m|)\|_p \, dx \tag{127}$$

for each vector valued function $\mathbf{u} = (u_1, \dots, u_m) : \Omega \to \mathbb{R}^m$.

Different choices of p have similar effects as in sparsity regularization. For $p = 1$ the norm favours the vector $(|\nabla u_1|, \dots, |\nabla u_m|)$ to be locally sparse, which means in particular that edge sets for the different components u_i will usually not coincide. This can be reasonable in some applications if the u_i represent some kind of complementary variables, but it is e.g. not desirable for color images, where most edges are expected to be at the same places for the different color components. In such cases $p \geq 2$ is more appropriate. For tensor images an analogous discussion applies to the choice of the suitable matrix norm.

8.3 Smoothed Total variation

In order to avoid difficulties with the non-differentiability of the total variation, i.e. of the Euclidean norm at zero, smoothed versions are frequently used (cf. [78, 190, 191]).

The most frequently used approximation of the total variation is of the form

$$TV^\epsilon(u) = \int_\Omega \sqrt{|\nabla u|^2 + \epsilon^2} \, dx. \tag{128}$$

The Huber norm is given by

$$H_\epsilon(u) = \int_\Omega h_\epsilon(|\nabla u|), \tag{129}$$

where h_ϵ is a locally quadratic approximation of the identity, i.e.

$$h_\epsilon(t) = \begin{cases} t & t > \epsilon \\ \frac{t^2}{\epsilon} & t \leq \epsilon \end{cases} \tag{130}$$

The Huber norm can also be obtained as a Moreau-Yosida regularization (cf. [125] of the total variation

$$h_\epsilon(\nabla u) = \min_s \left(|s| + \frac{1}{2\epsilon} |s - \nabla u|^2 \right). \tag{131}$$

A disadvantage of a smoothed approximation is the structure of solutions. Due to the differentiability at zero, the sharp edges can become smeared out. For very small ϵ that is used to obtain an appropriate approximation of the total variation, the parameter dependence can lead to significant slow down in computational methods for differentiable problems.

8.4 Nonlocal Regularization

Nonlocal regularization techniques have emerged recently from a proposed linear filter by Buades et al. [45] of the form

$$u(x) = \frac{1}{C_f(x)} \int_\Omega w_f(x,y) f(y)\, dy \tag{132}$$

with weight function

$$w_f(x,y) = \exp\left(-\frac{1}{h^2} \int_{\mathbb{R}^d} G_a(t) |f(x+t) - f(y+t)|^2\, dt\right) \tag{133}$$

with Gaussian G_a. This nonlocal filter generalizes neighbourhood filters in the pixel domain (like Yaroslavsky and SUSAN filters). It allows to take advantage of patches appearing in similar form throughout the image domain, which in particular applies to textures. The significant visual improvement in this respect led to an enormous increase of research in this direction.

The relation to total variation techniques was established first by Kindermann et al. [116] with the observation that the nonlocal filter can be realized equivalently by solving a quadratic variational problem of the form

$$\frac{\lambda}{2} \int_\Omega (u-f)^2\, dx + \frac{1}{2} \int_\Omega \int_\Omega w_f(x,y)(u(x)-u(y))^2\, dx\, dy. \tag{134}$$

Modifications of this quadratic problem towards a nonlocal version of the total variation (respectively the ROF functional) are possible, but not completely unique. In particular one can define a nonlocal total variation as

$$TV_{NL1}(u) = \int_\Omega \int_\Omega \sqrt{w_f(x,y)} |u(x)-u(y)|\, dx\, dy \tag{135}$$

and

$$TV_{NL2}(u) = \int_\Omega \sqrt{\int_\Omega w_f(x,y)(u(x)-u(y))^2\, dx}\, dy. \tag{136}$$

This variational problem can be analyzed in appropriate function spaces and generalized for other imaging and inversion tasks. An additional difficulty for problems where the forward operator A is deviating strongly from the identity is that the weights cannot be computed directly from the data. Two approaches have been proposed, a computation from a rough solution by linear inversion (cf. [128]) or an updating of the weights from the current solution in an iterative process (cf. [154, 204]).

As mentioned above, nonlocal techniques can lead to high quality texture reconstructions in images with repeated structures. A major drawback of nonlocal methods is the computational effort, since effectively the dimension of the problem is doubled (i.e. the number of degrees of freedoms in numerical computations is squared). If no spatial restriction is introduced already the construction of the weights can lead to overwhelming effort for large size 2D and for 3D images.

9 Numerical Methods

Over the last decades, a variety of numerical methods for TV-minimization has been proposed, we only provide a short overview here focusing on state-of-the art approaches.

The basic setup we consider in this section is the minimization of functionals $J : X \to \mathbb{R} \cup \{+\infty\}$ of the form

$$J(u) = \frac{\lambda}{2}\|Au - f\|^2 + \sup_{p \in K} \langle u, Dp \rangle, \tag{137}$$

where $A : X \to Y$ and $D : Z \to X$ are linear operators, usually A being bounded and compact, while D might be unbounded or with high norm. We assume that X and Z are Banach spaces with a convex bounded subset $K \subset Z$, and that Y is a Hilbert space. This covers the continuous (with D being the divergence) as well as the discretized setting.

The regularization functional

$$R(u) = \sup_{p \in K} \langle u, Dp \rangle \tag{138}$$

is the convex conjugate of the characteristic function of K and hence convex and positively one-homogeneous. We consider the typical case of K representing a point-wise constraint, which is the case in all common applications we have in mind. If X is a space of (generalized) functions on Ω (either a continuous or discrete set equipped with a positive measure), then the constraint set K is of the form

$$K = \{\, p \in Z \mid p(x) \in M \text{ for almost every } x \in \Omega \,\} \tag{139}$$

with $M \subset \mathbb{R}^m$ being a bounded convex set. With the convex conjugate of the charactistic function of M, which we denote by $F : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$,

$$F(s) = \sup_{q \in M} q \cdot s \tag{140}$$

the regularization functional can (at least formally) be written as

$$R(u) = \int F(D^* u)\, d\mu, \tag{141}$$

with μ an appropriate measure (see below for standard examples) and D^* being the adjoint of D in the L^2-scalar product with this measure.

The driving example is of course total variation regularization. In this case Z is the space of continuous vector fields vanishing at the boundary of a domain $\Omega \subset \mathbb{R}^d$, $D = \nabla\cdot$ is the divergence operator, X an L^p-space on Ω, and K is the unit ball

$$K = \{\, p \in Z \mid \|p\|_\infty \leq 1 \,\}. \tag{142}$$

Note that K can indeed be defined by the pointwise constraint $|p(x)| \leq 1$ for all $x \in \Omega$ and hence (141) holds for weakly differentiable functions u, where $D^* = -\nabla$ and $F(p) = |p|$, i.e., the total variation as defined in the first part.

A second important case are so-called *sparse* or *compressive* models, which are based on penalizing the coefficients in some orthonormal basis (or more generally in a frame) of a separable Hilbert space by a weighted ℓ^1-norm. By a suitable redefinition of the operator A as the effective operator acting on the coefficients, such algorithms can be rewritten in the above form with $X = \ell^2(\mathbb{N})$, $Z = \ell^\infty(\mathbb{N})$, D is the multiplication operator

$$(u_n)_{n\in\mathbb{N}} \mapsto (w_n u_n)_{n\in\mathbb{N}},$$

where (w_n) is a suitable sequence of positive weights, often identical to one, and

$$K = \{\, p \in Z \mid |p_n| \leq 1, \forall\, n \in \mathbb{N} \,\}. \tag{143}$$

The resulting regularization functional is then the weighted ℓ^1-norm

$$R(u) = \sum_{n\in\mathbb{N}} w_n |u_n|. \tag{144}$$

This is a special case of (141) with μ being the sum of weighted discrete point measures.

9.1 Optimality Conditions

In the following we provide a unifying formulation of optimality conditions for (137), which gives a common viewpoint on many existing methods. For this sake we introduce a gradient variable v and the corresponding constraint

$$D^*u + v = 0. \tag{145}$$

Under the main assumption that the regularization functional has the representation (141), the unconstrained variational problem (137) can then be rephrased as the constrained problem

$$\frac{\lambda}{2}\|Au - f\|^2 + \int F(v)\, d\mu \to \min \quad \text{subject to} \quad D^*u + v = 0. \tag{146}$$

The solution of the constrained problem is a saddle-point of the Lagrangian (minimum with respect to u and v, maximum with respect to w) given by

$$L(u, v, w) = \frac{\lambda}{2}\|Au - f\|^2 + \int F(v)\, d\mu + \int (D^*u + v)w\, d\mu. \tag{147}$$

The Lagrangian is differentiable with respect to u and w, hence for an optimal solution the variations with respect to this variable need to vanish. Moreover, the Lagrangian is convex with respect to v, hence zero needs to be an element of the subgradient. Note that the minimization problem with respect to v for fixed u and w is equivalent to the minimization of

$$\int (F(v) + vw)\, d\mu, \tag{148}$$

which is achieved by the pointwise minimization of $F(v) + vw$ for $v(x)$ μ-almost everywhere.

As a consequence of the above considerations we obtain the following optimality system for the constrained problem

$$\lambda A^*Au + Dw = \lambda A^*f \tag{149}$$

$$\partial F(v) + w \ni 0 \tag{150}$$

$$D^*u + v = 0 \tag{151}$$

As we shall see below, most of the methods indeed work without the variable v, which is the special case of an iteration method that always keeps (151) exactly satisfied. The above formulation has several advantages however. First of all it is quite general and allows to interpret many schemes as iterations on these optimality conditions, in particular also the recently proposed split Bregman method. Secondly,

the difficulties due to the gradient operator D and the nondifferentiability of F become separated. This allows to immediately generalize the schemes below to quite arbitrary operators D, and allows to investigate (150) as a pointwise relation. Several approximations of this relation will indeed distinguish different schemes, they can be more easily understood when considered in the pointwise sense.

9.2 Primal Approaches

We start with primal iteration approaches, corresponding to iterations on (150) or smoothed versions thereof, usually with additional damping in (149).

Most of these approaches are based on simplifying the subgradient relation as an approximation by smoothing, i.e., F is approximated by a family of parameter-dependent differentiable functionals F_δ such that $F_\delta \to F$ in a suitable sense (e.g. uniformly) as $\delta \to 0$. Several special approximations have been proposed depending on the specific choice of F. In the case of F being the Euclidean norm in $\mathbb{R}^m$, the C^∞-approximation

$$F_\delta(s) = \sqrt{|s|^2 + \delta^2} \tag{152}$$

received most attention (cf. [78, 163, 190–192]) due to its simplicity and high regularity. A quite general way to obtain first-order regular (differentiable with Lipschitz-continuous derivative) approximations is the *Moreau-Yosida regularization* (cf. [81]) defined by

$$F_\delta(s) = \inf_\sigma [F(\sigma) + \frac{1}{2\delta}|s - \sigma|^2]. \tag{153}$$

Even if it cannot be computed in an explicit form it is useful for computation, since one can just minimize the whole functional with respect to the new variable as well. In the case of F being the Euclidean norm in $\mathbb{R}^m$, the Moreau-Yosida regularization can be explicitely calculated as the so-called *Huber norm* (cf. [10, 109])

$$F_\delta(s) = \begin{cases} \frac{1}{2\delta}|s|^2 & \text{if } |s| \le \epsilon \\ |s| - \epsilon & \text{if } |s| > \epsilon. \end{cases} \tag{154}$$

A potential disadvantage of such approaches is the slow growth of the regularized functional F_δ at zero. Hence, sparsity is not promoted, since replacing small values by zero has only little gain in the regularization term (which is then completely balanced by losses in fidelity terms).

The first scheme proposed in [163] was a gradient descent method, which iterates (150) explicitely as

$$w^{k+1} = F_\delta'(v^k) \tag{155}$$

and then inserts w^{k+1} in (149). In order to obtain convergence, damping (with rather small damping parameter τ depending on the image size) needs to be used to compute

$$u^{k+1} = u^k - \tau(\lambda A^*(Au^k - f) + Dw^{k+1}). \tag{156}$$

Finally, (151) is enforced exactly, i.e. one computes

$$v^{k+1} = D^* u^{k+1}. \tag{157}$$

The main disadvantages of the gradient descent approach are the dependence on δ and in particular on the grid size (respectively number of pixels). In total variation, the role of τ corresponds to a time step in the explicit time discretization of a parabolic partial differential equation and thus needs to be very small (of the order of the grid size squared) to maintain stability and convergence to the minimizer.

A popular approach to reduce the severe restriction on τ is to use some kind of semi-implicit time stepping, which still yields linear problems in each time step. In this way (156) and (157) are solved as a linear system coupled with

$$L(v^k)v^{k+1} + w^{k+1} = L(v^k)v^k - F_\delta(v^k), \tag{158}$$

where $L(v^k)$ is a (discretized) differential operator such that $L(v)v$ approximates $F'_\delta(v)$. A quite popular method of this kind is the lagged diffusivity approximation (cf. [191, 192]) using (152) and

$$L(v)\varphi = \frac{\varphi}{\sqrt{|v|^2 + \delta^2}}, \tag{159}$$

which yields $L(v)v = F'_\delta(v)$. In the denoising case $A = I$ also the Newton method on the smoothed optimality system can be put in this form if $L = F''_\delta$ and $\tau = \frac{1}{\lambda}$. Besides the dependence of δ, a major disadvantage is that the matrix for the linear system to be solved has to be changed in each iteration step (an approximation L independent of v^k usually does not provide reasonable results) and may be badly conditioned for small δ.

9.3 Dual Approaches

Instead of using (150), the dual approach rather iterates the equivalent dual subgradient relation

$$v + \partial F^*(w) \ni 0. \tag{160}$$

Note that the convex conjugate $F^*(w)$ is again the characteristic function of M. Thus, (160) can be interpreted as the optimality condition for minimizing the linear functional $\langle v, w\rangle$ with respect to $w \in M$.

As in the primal case one can approximate the dual subgradient relation, respectively the inclusion $q \in M$, via smoothing. Again one can construct the Moreau-Yosida regularization

$$F_\delta^*(q) = \inf_r [F^*(r) + \frac{1}{2\delta}|r-q|^2] = \inf_{r\in M} \frac{1}{2\delta}|r-q|^2.$$

It is straight-forward to show that the minimum is attained for $r = P_M(q)$, where P_M is the projection onto M. Hence, we have

$$F_\delta^*(q) = \frac{1}{2\delta}|q - P_M(q)|^2 = \frac{1}{2\delta}\mathrm{dist}(q, M)^2.$$

This smoothing is a penalty method approximating F^* from below, i.e. we have $F^*(q) \geq F_\delta^*(q)$ for all q. If F is the Euclidean norm, i.e. M the Euclidean unit ball in $\mathbb{R}^m$, we have $P_M(q) = \frac{q}{|q|}$ if $q \notin M$. Hence

$$2\delta F_\delta^*(q) = |q - \frac{q}{|q|}|^2 = |q|^2 - 2|q| + 1 = (|q|-1)^2 \qquad \text{for } q \notin M.$$

A compact form for F_δ^* is the quadratic penalty

$$F_\delta^*(q) = \frac{1}{2\delta}\max\{|q|-1, 0\}^2. \tag{161}$$

A standard alternative to penalty schemes are barrier approaches, which amounts to approximating F^* from above, i.e. by a family of functionals F_δ^* differentiable on M such that $F_\delta^*(q) \geq F^*(q)$ for all q at least close to the boundary of M. The derivation of specific barrier functionals is usually based on a particular perturbation related to the complementarity condition for the constraint $q \in M$. Since the complementarity condition needs to be written for the specific inequalities defining M, the barrier functional has to be derived separately in each case. If F is the Euclidean norm respectively M is the unit ball in $\mathbb{R}^m$, then the resulting standard approximation in interior point methods is based on the logarithmic barrier

$$F_\delta^*(q) = -\delta\log(1-|q|). \tag{162}$$

In [139] Newton-type methods for barrier as well as penalty methods are constructed and also parallelization techniques are discussed. It is shown that in the case of penalty methods some slope is introduced in the usually flat parts of the reconstruction but the edges are sharp, while barrier methods keep the flat regions with introducing some smoothing of the edges. However, if the penalty parameter is sufficiently small compared to the grid size, this effect disappears.

In dual iteration approaches, (149) and (151) are satisfied exactly, i.e.

$$\lambda A^* A u^{k+1} + D w^{k+1} = \lambda A^* f \tag{163}$$

$$D^* u^{k+1} + v^{k+1} = 0, \tag{164}$$

while iterative approximation is carried out on (160) in order to determine w^{k+1}. Note that A^*A needs to be regular in order to solve (163) and (164) for u^{k+1} and v^{k+1}, which is not an issue for denoising, but prevents the use of dual methods for ill-posed problems such as deblurring or inpainting.

An obvious approach is to use a fixed point iteration on (160), i.e.,

$$w^{k+1} + \tau \partial F^*(w^{k+1}) \ni w^k - \tau v^k. \tag{165}$$

It is easy to see that this is the optimality condition for minimizing

$$J_k(w) = \frac{1}{2}\|w - w^k + \tau v^k\|^2 + \tau F(w),$$

and since F it is the characteristic function of M it is further equivalent to minimize the least-squares term subject to $w \in M$, which yields

$$w^{k+1} = P_M(w^k - \tau v^k),$$

where P_M is the projection onto M. Since v^k is the gradient of the dual objective with respect to w, this iteration is just the dual projected gradient method proposed in [205]. An older and more popular approach in the same spirit is due to Chambolle [60] in the case of M being the unit ball (in the Euclidean norm, but extensions to other norms are obvious). Chambolle did not use exact projection, but

$$w^{k+1} = \frac{w^k - \tau v^k}{1 + \tau |v^k|},$$

which is still consistent. The behaviour of the iteration is very similar to the projected gradient method. Again a major disadvantage is that τ needs to be chosen very small in order to obtain a stable iteration (note that the dual gradient is a differential operator of the form $v \mapsto \nabla \cdot (a \nabla v)$).

9.4 Augmented Lagrangian Methods

Recently the most popular methods are of Augmented Lagrangian type, although the seminal paper formulated them differently motivated from the Bregman iteration above, see [96] for this Split Bregman method in detail. The major idea in our setup

of (149)–(151) is to use an augmented Lagrangian approach, i.e. add multiples of the (151) to (149) and (150), which clearly yields an equivalent system that however can be solved uniquely for u and v. Then an iterative update of w in (151) is performed.

The augmented optimality system, with parameter $\mu > 0$ consists of (151) and

$$\lambda(A^*A + \mu DD^*)u + \mu Dv + Dw = \lambda A^* f \tag{166}$$

$$\partial F(v) + \mu v + \mu D^* u + w \ni 0. \tag{167}$$

In the case of total variation DD^* is the Laplacian, which can be inverted if the mean value or boundary values are fixed. Moreover, $\mu v + F(v)$ is maximally monotone and thus has a single-valued inverse, which can be computed explicitely by shrinkage formulas for the typical choices of M.

In its standard form the iteration first solve (166) and (167) for u and v, i.e.,

$$\lambda(A^*A + \mu DD^*)u^{k+1} + \mu Dv^{k+1} = \lambda A^* f - Dw^k \tag{168}$$

$$\partial F(v^{k+1}) + \mu v^{k+1} + \mu D^* u^{k+1} \ni -w^k, \tag{169}$$

and then perform a gradient ascent step for the dual variable w, i.e.,

$$w^{k+1} = w^k + \mu(D^* u^{k+1} + v^{k+1}). \tag{170}$$

A difficulty in the realization of this iteration is the fact that (168) and (169) are fully coupled, which is usually solved via an inner iteration alternating between these two. This idea can also be directly incorporated in the outer iteration, replacing (168) by

$$\lambda(A^*A + \mu DD^*)u^{k+1} = \lambda A^* f - \mu Dv^k - Dw^k, \tag{171}$$

yielding an approach equivalent to the alternating direction of multipliers method (ADMM, cf. [35, 158]). Variants can be found in [14, 85, 203].

10 Segmentation via TV Relaxation

While all above approaches were based on computing an image intensity as a function with values in $\mathbb{R}$ (or a closed subinterval), region-based segmentations rather look for a decomposition of Ω into subsets with different characteristics. The relation to functions can be made by two approaches, either via characteristic functions

$$\chi(x) = \begin{cases} 1 & x \in \Sigma \subset \Omega \\ 0 & \text{else} \end{cases} \tag{172}$$

or via level set functions (cf. [145])

$$\Sigma = \{ x \in \Omega \mid \phi(x) < 0 \} \tag{173}$$

with ϕ continuous. Modern region-based segmentation schemes are based on minimizing an energy depending on the image characteristics in the two subregions (some volume integrals of f) penalized by the perimeter of $\partial\Sigma$ in order to avoid a classification of noisy parts as small regions to be segmented. The perimeter provides the main link to total variation methods, since the coarea formula yields

$$TV(u) = \int_{\mathbb{R}} \mathrm{Per}(\partial\{u > \alpha\}) d\alpha, \tag{174}$$

where $\mathrm{Per}(M)$ denotes the perimeter of a set M (the $d-1$-dimensional Hausdorff-measure). For $u = \chi$ or $u = H(\phi)$ with χ and ϕ as above and H the Heaviside function we obtain $\{u > \alpha\} = \emptyset$ for $\alpha \geq 1$ and $\{u > \alpha\} = \Omega$ for $\alpha < 0$. In both cases the perimeter of $\partial\{u > \alpha\}$ is zero. Moreover, for $\alpha \in (0, 1)$ we have

$$\{u > \alpha\} = \Sigma.$$

Hence, the coarea formula implies

$$\mathrm{Per}(\partial\Sigma) = \int_0^1 \mathrm{Per}(\partial\Sigma) d\alpha = TV(u).$$

Thus, the minimization of shape functionals penalized by perimeter as encountered in image segmentation can be rewritten into problems with total variation minimization.

A very interesting connection has been established in a series of papers by Chan et al. [70, 71], who analyzed total variation problems with one-homogeneous fitting

$$\lambda \int_\Omega |u - f| \, dx + TV(u) \to \min_u, \tag{175}$$

as well as the regularized linear functional

$$\int_\Omega g \, u \, dx + TV(u) \to \min_u, \tag{176}$$

subject to bound constraints $0 \leq u \leq 1$. Based on the co-area formula and a layer-cake formula for the first integral a connection to shape optimization problems can be derived. We review this analysis in the case of (176). First of all, the existence of a minimizer of (176) with either the bound constraint

$$0 \leq u \leq 1 \qquad \text{a.e. in } \Omega \tag{177}$$

or the discrete constraint

$$u \in \{0, 1\} \qquad \text{a.e. in } \Omega \tag{178}$$

can be shown:

Lemma 10.1. *Let $g \in L^1(\Omega)$. Then there exists a minimizer of the constrained minimization problems* (176) *subject to* (177) *as well as of* (176) *subject to* (178).

Proof. The admissible set is a subset of the unit ball in $L^\infty(\Omega)$ and on sublevel sets of the objective functional

$$E(u) = \int_\Omega g\, u\, dx + TV(u)$$

we have

$$TV(u) \leq C - \int g\, u\, dx \leq C + \|g\|_{L^1}.$$

Hence, the sublevel sets of J intersected with the admissible set are bounded in $L^\infty(\Omega) \cap BV(\Omega)$ and consequently compact in the weak* topologies of both spaces.

As we have shown before the total variation is lower semicontinuous in the weak* topology of $BV(\Omega)$ and the first integral is a bounded linear functional (of g) on $L^1(\Omega)$ hence weak* continuous in $L^\infty(\Omega) = L^1(\Omega)^*$. Thus, the functional J is lower semicontinuous with compact sub level sets, which implies the existence of a minimizer. □

The major result is the following exact relaxation of the problem for characteristic functions:

Theorem 10.2. *Let u be a minimizer of* (176) *subject to* (177). *Then for almost all $\alpha \in (0, 1)$, the thresholded function*

$$\chi^\alpha(x) = \begin{cases} 1 \text{ if } u(x) > \alpha \\ 0 \text{ else} \end{cases} \tag{179}$$

is a solution of (176) *subject to* (178).

Proof. We first of all derive a layer-cake representation of the linear functional via

$$\begin{aligned} \int_\Omega g\, u\, dx &= \int_\Omega g(x) \int_0^{u(x)} d\alpha\, dx \\ &= \int_\Omega g(x) \int_0^1 \chi^\alpha(x)\, d\alpha\, dx \\ &= \int_0^1 \left[\int_\Omega g\, \chi^\alpha\, dx \right] d\alpha. \end{aligned}$$

For the total variation we have from the co-area formula

$$TV(u) = \int_0^1 \mathrm{Per}(\partial\{u > \alpha\})d\alpha = \int_0^1 TV(\chi^\alpha)\, dx.$$

Together we find that for u being a minimizer of (176) subject to (177)

$$E(u) = \int_0^1 E(\chi^\alpha)\, d\alpha \geq \int_0^1 E(u)\, d\alpha = E(u),$$

where we have used $E(\chi^\alpha) \geq E(u)$ for all α, since χ^α is admissible for the minimization problem (176) subject to (177) and u is the minimizer. The above relation is only possible if $E(\chi^\alpha) = E(u)$ for almost all $\alpha \in (0, 1)$. Hence, for almost every α the characteristic function χ^α is a minimizer of (176) subject to (177). On the other hand (177) is a relaxation of (178), thus χ^α needs to be a minimizer of (176) subject to (178). □

The exact relaxation result in Theorem 10.2 has two important consequences: First of all, the nonconvex $0-1$ minimization can be solved equivalently by minimizing a convex problem, which resolves issues with minimizers. Secondly, techniques from the analysis of total variation minimization, e.g. existence, asymptotic behaviour, and error estimates can be carried over.

A particularly interesting example is the relaxation of the Chan–Vese model [66], which can be written as the minimization of

$$J_{CV}(\chi, c_1, c_2) = \frac{\lambda}{2}\int_\Omega [\chi(c_1 - f)^2 + (1-\chi)(c_2 - f)^2]\, dx + TV(\chi) \tag{180}$$

subject to (178). The minimization can be realized in an alternating fashion, with respect to χ and with respect to the constants c_1 and c_2. For given χ, the minimizers c_1 and c_2 can be computed explicitely as mean values of f, namely

$$c_1 = \frac{\int_\Omega \chi\, f\, dx}{\int_\Omega \chi\, dx}, \qquad c_2 = \frac{\int_\Omega (1-\chi)\, f\, dx}{\int_\Omega (1-\chi)\, dx}. \tag{181}$$

For the minimization with respect to χ Theorem (10.2) can be applied with

$$g = \frac{\lambda}{2}[\chi(c_1 - f)^2 + (1-\chi)(c_2 - f)^2]. \tag{182}$$

Hence, the minimization of J_{CV} subject to (178) can be relaxed exactly to the minimization with (177).

Extensions of the relaxation to more general problems in topology optimization are considered in [46], and a region-based version of the Mumford-Shah model, which allows exact relaxation is introduced in [168, 198]. Cremers et al. [156] introduce a lifting method, which allows to treat some nonconvex problems in

imaging by convex relaxation after considering the graph and the associated 0-1 function in a higher dimension. Extensions to multi-label problems are considered in [124, 155], those relaxations are not exact however.

11 Applications

So far we have focused on the methodology of total variation whose development was heavily motivated by practical applications in the past. In this section and the following we finally take an opposite viewpoint and give a (certainly uncomplete) overview of the use of total variation methods in practices. We start in this section with applications in classical imaging tasks from computer vision and image analysis, before we proceed to applications in natural and life sciences or disciplines like the arts.

The most obvious application part is image analysis such as denoising, which was the original motivation for the ROF-model. Recent approaches rather use modified versions of total variations and in particular nonlocal versions (cf. e.g. [116, 129, 130, 166]), which are able to deal better with textures (assumed to be self-repeating patterns in the image). Another obvious application field of total variation techniques is segmentation as we have seen from the previous chapter. Other image analysis tasks where total variation methods have been applied successfully are zooming and superresolution (cf. e.g. [4, 199, 201]), colour enhancement (cf. e.g. [33, 74, 133, 142]), fusion (cf. e.g. [22, 157, 193]), flow estimation and video treatment (cf. e.g. [43, 76, 195, 196], inpainting (cf. [52, 64, 65, 84]), and image decomposition into a cartoon and texture part (cf. e.g. [12, 16, 17, 30, 40, 83, 134, 147, 186, 187]). In particular the latter two have led to an investigation of a variety of novel data fidelity terms, since standard L^2-terms are not appropriate. In particular the work by Meyer [134], who used the dual norm of the TV-norm as data fidelity, was quite influential and motivated several studies of the dual and predual space of BV. We also refer to [63] for total variation methods in image analysis.

We have already discussed deblurring and deconvolution above and here mention further work in blind deconvolution. The extension from deconvolution to blind deconvolution is rather straight-forward, one usually looks for the point-spread function in a parametric or non-parametric form as well as for the image. Since the point-spread function is usually smooth, the majority of approaches restricts the total variation regularization to the image. With alternating minimization approaches the numerical realization is quite straight-forward and analogous to standard deconvolution. An interesting question is whether total variation regularization on the image combined with classical smoothness priors on the point-spread function is suitable to overcome the non-uniqueness in blind deconvolution, which is confirmed by promising computational results, but has hardly been investigated by theoretical analysis (cf. e.g. [19, 31, 48, 67, 102].

Due their favourable properties, total variation methods have found access to many applied areas of biomedical imaging. Due to the enormous amount of

applications in the last two decades, we can only give a selection here and further links to literature:

- *Limited angle CT:* Due to the mild ill-posedness and the low noise levels, regularization is not a big topic in modern standard X-ray tomography. The situation changes in limited angle CT as e.g. appearing in modern C-Arm devices, where the ill-posedness of the reconstruction problem is much more severe. TV regularized reconstructions are usually obtained with the standard quadratic fidelity term and the forward operator being the Radon transform with limited angles, see e.g. [149, 153, 184]
- *Magnetic resonance (MR) imaging:* Image reconstruction in MR is mainly achieved by inverting a Fourier transform, which can be carried out efficiently and robustly if a sufficient number of Fourier coefficients, i.e. measurements, is available, which is usually the case. Regularization and thus also total variation approaches are mainly needed for special applications such as fast MR protocolls, which do not allow to measure enough frequencies. This motivation has also triggered a lot of research in compressed sensing, where for this application effectively a discretized total variation is used (cf. e.g. [132]). Improved approaches use a decomposition of total variation with a second functional, e.g. based on wavelets, which can take care of the typical slopes arising between edges in MR images, cf. [32, 103, 115, 118–120].
- *Emission tomography:* Emission tomography techniques (cf. [197]) used in nuclear medicine such as *Positron Emission Tomography* (PET) and *Single Photon Emission Computed Tomography* (SPECT) are a natural target for total variation regularization due to inherent high noise level, more precisely the Poisson statistics of the data. We refer to [23, 113, 150, 165–167].
- *Microscopy:* In modern fluorescence microscopy low photon counts are an issue in particular in live imaging and in high resolution imaging at nanoscopic scales. Here total variation methods with data terms appropriate for Poisson noise have been quite benefitial for denoising and (blind) deconvolution (cf. e.g. [42, 44, 77, 91, 159, 167, 180]).
- *Other modalities:* For certain highly ill-posed image reconstruction tasks with expected sharp edges the use of total variation as a regularization is a quite obvious choice, e.g. in optical tomography (cf. e.g. [1, 80, 152]) or in electron tomography (cf. e.g. [24, 97]). A less obvious application of total variation is the study of EEG/MEG reconstructions, where usually rather ℓ^1 priors are used since brain activity is often sparse. Adde et al. [3] however compared total variation methods to standard density reconstruction techniques and other nonlinear diffusions and found it clearly outperforms them. In ultrasound, denoising and segmentation is of interest, where main attention has to be focused on appropriate modelling of speckle noise in the variational problems (cf. e.g. [112, 168]).

Since imaging is nowadays of high importance, it is not surprising that total variation methods have found their place there as well. Examples are astronomy (cf. e.g. [151] and references therein) or geosciences, where hyperspectral imaging is becoming a standard approach (cf. e.g. [111, 136, 137, 202]). Less expected

applications appear in the tracking of sharp fronts in weather forecast (cf. [92]) and in the arts, where frescoes have been reconstructed by mathematical techniques instead of human experience (cf. [18, 89, 90]).

12 Open Questions

We finally provide a discussion of some problems related to the TV zoo, which may be particularly rewarding for future research. Again the choice of questions and viewpoints presented here are heavily motivated by our own research and may be subjective, moreover various yet unexpected questions due to the fast progress in this field the will certainly pop up in the next years:

- *Local choice of regularization parameters:* This issue receives growing interest in many recent investigations, motivated by different scales appearing in images locally as well as statistical models. Whereas the effects of locally large and small regularization parameters seem rather obvious in classical smoothing techniques (e.g. squared Sobolev norms), where small local regularization parameters allow high variation and approximate discontinuities (such as in the Ambrosio–Tortorelli model [5]), the effect in total variation is less clear and remains to be analyzed (cf. e.g. [106]).
- *Nonlinear inversion vs. two-step methods:* As we have discussed above, total variation methods can be incorporated into inversion methods and image reconstruction approaches. However, it is not obvious that the additional difficulties and increased computational effort is indeed necessary. In several cases results of a similar quality are achieved by simply using fast linear reconstruction methods with a subsequent application of the ROF model or variants thereof (cf. [166, 167]). A detailed analysis under which conditions the incorporation of total variation or similar methods into the reconstruction algorithms yields significantly superior results is still open.
- *Improved TV models:* Several improvements of total variation, in particular related to the staircasing effect, are currently a topic receiving increasing attention. A particularly timely topic seems to be the adaptive coupling of several TV-type functionals, which might concern anisotropy (cf. [30, 83]) or the combination of total variation and higher-order total variation. The latter has already been investigated via inf-convolution in [61], recently further improvements have been proposed, which yield promising results but also require a more detailed analysis (cf. [29, 36, 171, 172]).
- *4D regularization:* In many applications high-dimensional image structures appear, e.g. spectral or time-resolved images. In such cases it seems reasonable to apply total variation methods for the two or three spatial dimensions, while it makes less sense to favour piecewise constant structures in the additional directions. The appropriate combination of total variation with other regularization functionals related to the additional dimensions remains a challenging topic for future research.

- *Bayesian modelling:* While we have concentrated on variational techniques such as MAP estimation here, in wide areas of statistics it is much more common to use conditional mean (CM) estimates. In principle a prior probability related to total variation can be defined in discretizations in the form $p(u) \sim e^{-\alpha TV(u)}$ and the limit can be studied (cf. [121, 129]). However, convergence of CM estimates appears only with different scaling than in the case of MAP estimation and the limit is then a Gaussian measures, thus edges are not preserved. Appropriate models for edge preserving priors remain to be developed in a TV context. An interesting result in this direction is obtained in [75], who studies the choice $p(u) \sim e^{-\alpha TV(u)^2}$ instead.

Acknowledgements The work of MB has been supported by the German Science Foundation DFG through the project *Regularization with Singular Energies* and SFB 656 *Molecular Cardiovascular Imaging*, and by the German Ministery of Education and Research (BMBF) through the project *INVERS: Deconvolution problems with sparsity constraints in nanoscopy and mass spectrometry*.

The work of SO has been supported by NSF grants DMS0835863 and DMS0914561, ONR grant N000140910360 and ARO MURI subs from Rice University and the University of North Carolina.

The authors thank Alex Sawatzky (WWU Münster) for careful proofreading and comments to improve the manuscript.

References

1. J.F. Abascal, J. Chamorro-Servent, J. Aguirre, J.J. Vaquero, S. Arridge, T. Correia, J. Ripoll, M. Desco, Fluorescence diffuse optical tomography using the split Bregman method. Med. Phys. **38**, 6275 (2011)
2. R. Acar, C.R. Vogel, Analysis of total variation penalty methods. Inverse Probl. **10**, 1217–1229 (1994)
3. G. Adde, M. Clerc, R. Keriven, Imaging methods for MEG/EEG inverse problem. Int. J. Bioelectromagn. **7**, 111–114 (2005)
4. A. Almansa, V. Caselles, G. Haro B. Rouge, Restoration and zoom of irregularly sampled, blurred and noisy images by accurate total variation minimization with local constraints. Multiscale Mod. Simulat. **5**, 235–272 (2006)
5. L. Ambrosio, V. Tortorelli, Approximation of functionals depending on jumps by elliptic functionals via Γ- convergence. Comm. Pure Appl. Math. **43**, 999–1036 (1990)
6. L. Ambrosio, N. Fusco, D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems* (Oxford University Press, Oxford, 2000)
7. F. Andreu, C. Ballester, V. Caselles, J.M. Mazon, Minimizing total variation flow. Differ. Integr. Equat. **14**, 321–360 (2001)
8. F. Andreu, C. Ballester, V. Caselles, J.M. Mazon, The Dirichlet problem for the total variation flow. J. Funct. Anal. **180**, 347–403 (2001)
9. F. Andreu, V. Caselles, J.I. Diaz, J.M. Mazon, Some qualitative properties for the total variation flow. J. Funct. Anal. **188**, 516–547 (2002)
10. U. Ascher, E. Haber, *Computational Methods for Large Distributed Parameter Estimation Problems with Possible Discontinuities*. Symp. Inverse Problems, Design and Optimization, Rio, 2004

11. P. Athavale, E. Tadmor, Multiscale image representation using integro-differential equations. Inverse Probl. Imag. **3**, 693–710 (2009)
12. G. Aubert, J.F. Aujol, Modeling very oscillating signals, application to image processing. Appl. Math. Optim. **51**, 163–182 (2005)
13. G. Aubert, J.F. Aujol, A Variational approach to remove multiplicative noise. SIAM J. Appl. Math. **68**, 925–946 (2008)
14. J.F. Aujol, Some first-order algorithms for total variation based image restoration. J. Math. Imag. Vis. **34**, 307–327 (2009)
15. J.F. Aujol, A. Chambolle, Dual norms and image decomposition models. IJCV **63**, 85–104 (2005)
16. J.F. Aujol, G. Aubert, L. Blanc-Feraud, A. Chambolle, Image decomposition into a bounded variation component and an oscillating component. J. Math. Imag. Vis. **22**, 71–88 (2005)
17. J.F. Aujol, G. Gilboa, T. Chan, S. Osher, Structure-texture image decomposition - modeling, algorithms, and parameter selection. Int. J. Comput. Vis. **67**, 111–136 (2006)
18. W. Baatz, M. Fornasier, P. Markowich, C.B. Schönlieb, *Inpainting of ancient Austrian frescoes*. Conference Proceedings of Bridges 2008, Leeuwarden, 2008, pp. 150–156
19. S.D. Babacan, R. Molina, A.K. Katsaggelos, Variational Bayesian blind deconvolution using a total variation prior. IEEE Trans. Image Proc. **18**, 12–26 (2009)
20. M. Bachmayr, *Iterative Total Variation Methods for Nonlinear Inverse Problems*, Master Thesis (Johannes Kepler University, Linz, 2007)
21. M. Bachmayr, M. Burger, Iterative total variation methods for nonlinear inverse problems. Inverse Probl. **25**, 105004 (2009)
22. C. Ballester, V. Caselles, L. Igual, J. Verdera, B. Rouge, A variational model for P+XS image fusion. IJCV **69**, 43–58 (2006)
23. J. Bardsley, A. Luttman, Total variation-penalized Poisson likelihood estimation for ill-posed problems . Adv. Comp. Math. **31**, 35–59 (2009)
24. C. Bazan, *PDE-Based Image and Structure Enhancement for Electron Tomography of Mitochondria*, PhD-Thesis (San Diego State University, San Diego, 2009)
25. G. Bellettini, V. Caselles, The total variation flow in $\mathbf{R}^N$. J. Differ. Equat. **184**, 475–525 (2002)
26. G. Bellettini, V. Caselles, M. Novaga, Explicit solutions of the eigenvalue problem $-$ div $(Du/|Du|) = u$. SIAM J. Math. Anal. **36**, 1095–1129 (2005)
27. M. Benning, *Singular Regularization of Inverse Problems* (PhD Thesis, WWU Münster, 2011)
28. M. Benning, M. Burger, Error estimation with general fidelities. Electron. Trans. Numer. Anal. **38**, 44–68 (2011)
29. M. Benning, C. Brune, M. Burger, J. Müller, Higher-order TV methods—enhancement via Bregman iteration. J. Sci. Comput. **54**(2–3), 269–310 (2013)
30. B. Berkels, M. Burger, M. Droske, O. Nemitz, M. Rumpf, Cartoon extraction based on anisotropic image classification, in *Vision, Modeling, and Visualization 2006: Proceedings*, ed. by L. Kobbelt, T. Kuhlen, T. Aach, R. Westerman (IOS Press, Aachen, 2006)
31. J.M. Bioucas-Dias, M.A.T. Figueiredo, J.P. Oliveira, Total variation-based image deconvolution: a majorization-minimization approach, in *2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings*, vol. 2, pp. II, 14–19 May 2006
32. K.T. Block, M. Uecker, J. Frahm, Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn. Reson. Med. **57**, 1086–1098 (2007)
33. P. Blomgren, T. Chan, Color TV: Total variation methods for restoration of vector valued images. IEEE Trans. Image Proc. **7**, 304–309 (1998)
34. T. Bonesky, K.S. Kazimierski, P. Maass, F. Schöpfer, T. Schuster, Minimization of Tikhonov functionals in Banach spaces. Abstr. Appl. Anal. **19**, 192679 (2008)
35. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1–122 (2011)

36. K. Bredies, K. Kunisch, T. Pock, Total generalized variation. SIAM J. Imag. Sci. **3**, 492–526 (2010)
37. L.M. Bregman, 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. Math. Phys. **7**, 200–217 (1967)
38. M. Breuss, T. Brox, A. Bürgel, T. Sonar, J. Weickert, Numerical aspects of TV flow. Numer. Algorithms **41**, 79–101 (2006)
39. A. Briani, A. Chambolle, M. Novaga, G. Orlandi, On the gradient flow of a one-homogeneous functional, Preprint (SNS, Pisa, 2011)
40. T. Brox, J. Weickert, A TV flow based local scale measure for texture discrimination, in *Computer Vision - ECCV 2004*, ed. by T. Pajdla, J. Matas (Springer, Berlin, 2004), pp. 578–590
41. T. Brox, M. Welk, G. Steidl, J. Weickert, Equivalence results for TV diffusion and TV regularization, in *Scale Space Methods in Computer Vision*, ed. by L.D. Griffin, M. Lillholm (Springer, Berlin, 2003), pp. 86–100
42. C. Brune, A. Sawatzky, M. Burger, Bregman-EM-TV methods with application to optical nanoscopy, in *Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision*. LNC, vol. 5567 (Springer, Berlin, 2009), pp. 235–246
43. C. Brune, H. Maurer, M. Wagner, Detection of intensity and motion edges within optical flow via multidimensional control. SIAM J. Imag. Sci. **2**, 1190–1210 (2009)
44. C. Brune, A. Sawatzky, M. Burger, Primal and dual Bregman methods with application to optical nanoscopy. Int. J. Comput. Vis. **92**, 211–229 (2011)
45. A. Buades, B. Coll, J.M. Morel, A review of image denoising algorithms, with a new one. Multiscale Model. Simulat. **4**, 490–530 (2005)
46. M. Burger, M. Hintermüller, Projected gradient flows for BV / level set relaxation . PAMM **5**, 11–14 (2005)
47. M. Burger, S. Osher, Convergence rates of convex variational regularization. Inverse Probl. **20**, 1411–1421 (2004)
48. M. Burger, O. Scherzer, Regularization methods for blind deconvolution and blind source separation problems. Math. Contr. Signals Syst. **14**, 358–383 (2001)
49. M. Burger, G. Gilboa, S. Osher, J. Xu, Nonlinear inverse scale space methods for image restoration. Comm. Math. Sci. **4**, 179–212 (2006)
50. M. Burger, E. Resmerita, L. He, Error estimation for Bregman iterations and inverse scale space methods. Computing **81**, 109–135 (2007)
51. M. Burger, K. Frick, S. Osher, O. Scherzer, Inverse total variation flow. SIAM Multiscale Mod. Simul. **6**, 366–395 (2007)
52. M. Burger, L. He, C.B. Schönlieb, Cahn-Hilliard inpainting and a generalization for grayvalue images. SIAM J. Imag. Sci. **2**, 1129–1167 (2009)
53. M. Burger, M. Möller, M. Benning, S. Osher, An adaptive inverse scale space method for compressed sensing. Math. Comput. (2012, to appear)
54. M. Burger, M. Franek, C.-B. Schönlieb, Regularized regression and density estimation based on optimal transport. Appl. Math. Res. Express **2012**, 209–253 (2012)
55. J. Cai, S. Osher, Z. Shen, Linearized Bregman iterations for compressed sensing. Math. Comput. **78**, 1515–1536 (2008)
56. J. Cai, S. Osher, Z. Shen, Convergence of the linearized Bregman iteration for l_1-norm minimization. Math. Comput. **78**, 2127–2136 (2009)
57. J. Cai, S. Osher, Z. Shen, Linearized Bregman iteration for frame based image deblurring. SIAM J. Imag. Sci. **2**, 226–252 (2009)
58. V. Caselles, A. Chambolle, M. Novaga, The discontinuity set of solutions of the TV denoising problem and some extensions. MMS **6**, 879–894 (2007)
59. V. Caselles, A. Chambolle, M. Novaga, Regularity for solutions of the total variation denoising problem. Rev. Mat. Iberoamericana **27**, 233–252 (2011)
60. A. Chambolle, An algorithm for total variation regularization and denoising. J. Math. Imag. Vis. **20**, 89–97 (2004)

61. A. Chambolle, P.L. Lions, Image recovery via total variational minimization and related problems. Numer. Math. **76**, 167–188 (1997)
62. 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**, 319–335 (1998)
63. A. Chambolle, V. Caselles, D. Cremers, M. Novaga, T. Pock, An Introduction to total variation for image analysis, in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, ed. by M. Fornasier. Radon Series in Applied and Computational Mathematics (De Gruyter, Berlin, 2010)
64. T. Chan, J. Shen, Mathematical models for local non-texture inpainting. SIAM J. Appl. Math. **62**, 1019–1043 (2001)
65. T. Chan, J. Shen, *Image Processing and Analysis* (SIAM, Philadelphia, 2005)
66. T. Chan, L.A. Vese, Active contours without edges. IEEE Trans. Image Process. **10**, 266–277 (2001)
67. T. Chan, C.W. Wong, Total variation blind deconvolution. IEEE Trans. Imag. Proc. **7**, 370–375 (1998)
68. T. Chan, A. Marquina, P. Mulet, Second order differential functionals in total variation-based image restoration. CAM Report 98–35, UCLA, 1998
69. T. Chan, S.H. Kang, J. Shen, Total variation denoising and enhancement of color images based on the CB and HSV color models. J. Vis. Comm. Image Represent. **12**, 422–435 (2001)
70. T.F. Chan, S. Esedoglu, Aspects of total variation regularized L^1 function approximation. SIAM J. Appl. Math. **65**, 1817–1837 (2005)
71. T.F. Chan, S. Esedoglu, M. Nikolova, Algorithms for finding global minimizers of denoising and segmentation models. SIAM J. Appl. Math. **66**, 1632–1648 (2006)
72. G. Chavent, K. Kunisch, Regularization of linear least squares problems by total bounded variation. ESAIM Cont. Optim. Calc. Var. **2**, 359–376 (1997)
73. O. Christiansen, T.-M. Lee, J. Lie, U. Sinha, T.F. Chan, Total variation regularization of matrix-valued images. Int. J. Biomed. Imaging **2007**, Article ID 27432, 11 p. (2007). doi:10.1155/2007/27432
74. L. Cinque, G. Morrone, Retinex vombined with total variation for image illumination normalization, in *Image Analysis and Processing Ű ICIAP 2009*. LNCS, vol. 5716 (Springer, Berlin, 2009), pp. 958–964
75. S. Comelli, A Novel Class of Priors for Edge-Preserving Methods in Bayesian Inversion. Master thesis, University of Milano, 2011
76. R. Deriche, P. Kornprobst, G. Aubert, Optical flow estimation while preserving its discontinuities: A variational approach, in *Proceedings of the Asian Conference on Computer Vision, ACCVŠ95*, Singapore, 1995
77. 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. Microsc. Res. Tech. **69**, 260–266 (2006)
78. D. Dobson, O. Scherzer, Analysis of regularized total variation penalty methods for denoising. Inverse Probl. **12**, 601–617 (1996)
79. D. Donoho, I. Johnstone, Ideal spatial adaptation via wavelet shrinkage. Biometrika **81**, 425–455 (1994)
80. A. Douiri, M. Schweiger, J. Riley, S.R. Arridge, Local diffusion regularisation method for optical tomography reconstruction using robust statistics. Optic. Lett. **30**, 2439–2441 (2005)
81. I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*. Corrected Reprint Edition (SIAM, Philadelphia, 1999)
82. H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems* (Kluwer, Dordrecht, 1996)
83. S. Esedoglu, S.J. Osher, Decomposition of images by the anisotropic Rudin-Osher-Fatemi model. Comm. Pure Appl. Math. **57**, 1609–1626 (2004)
84. E. Esser, Primal Dual Algorithms for Convex Models and Applications to Image Restoration, Registration and Nonlocal Inpainting. PhD thesis, UCLA, 2010

85. E. Esser, X. Zhang, T. Chan, A general framework for a class of first order primal-dual algorithms for tv minimization. SIAM J. Imag. Sci. **3**, 1015–1046 (2010)
86. L.C. Evans, R.F. Gariepy, *Measure Theory and Fine Properties of Functions* (CRC Press, Boca Raton, 1992)
87. X. Feng, A. Prohl, Analysis of total variation flow and its finite element approximations. ESAIM: Math. Mod. Numer. Anal. **37**, 533–556 (2003)
88. J. Flemming, B. Hofmann, Convergence rates in constrained Tikhonov regularization: equivalence of projected source conditions and variational inequalities. Inverse Probl. **27**, 085001 (2011)
89. M. Fornasier, Mathematics enters the picture, in *Mathknow Mathematics*, ed. by M. Emmer, A. Quarteroni. Applied Sciences and Real Life (Springer, Milan, 2009), pp. 217–228
90. M. Fornasier, G. Teschke, R. Ramlau, A comparison of joint sparsity and total variation minimization algorithms in a real-life art restoration problem. Adv. Comput. Math. **31**, 301–329 (2009)
91. M. Freiberger, C. Clason, H. Scharfetter, Total variation regularization for nonlinear fluorescence tomography with an augmented Lagrangian splitting approach. Appl. Optic. **49**, 3741–3747 (2010)
92. M.A. Freitag, N.K. Nichols, C.J. Budd, Resolution of sharp fronts in the presence of model error in variational data assimilation, Preprint (University of Bath, 2010)
93. H.Y. Gao, A.G. Bruce, WaveShrink with firm shrinkage. Statist. Sinica **7**, 855–874 (1997)
94. S. Geman, D. Geman, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984)
95. E. Giusti, *Minimal Surfaces and Functions of Bounded Variation* (Birkhäuser, Boston, 1984)
96. T. Goldstein, S. Osher, The split Bregman method for L1 regularized problems. SIAM J. Imag. Sci. **2**, 323–343 (2009)
97. B. Goris, M.W. Van den Broek, K.J. Batenburg, H.H. Mezerji, S. Bals, Electron tomography based on a total variation minimization reconstruction technique. Ultramicroscopy **113**, 120–130 (2012)
98. M. Grasmair, Generalized Bregman distances and convergence rates for non-convex regularization methods. Inverse Probl. **26**, 115014 (2010)
99. M. Grasmair, Linear convergence rates for Tikhonov regularization with positively homogeneous functionals. Inverse Probl. **27**, 075014 (2011)
100. A. Haddad, Y. Meyer, Variational Methods in Image Processing. CAM-Report 04–52, UCLA, 2004
101. A. Haddad, Texture separation in $BV - G$ and $BV - L^1$ models. SIAM Multiscale Model. Simul. **6**, 273–286 (2007)
102. L. He, A. Marquina, S. Osher, Blind deconvolution using TV regularization and Bregman iteration. Int. J. Imag. Syst. Tech. **15**, 74–83 (2005)
103. L. He, T.C. Chung, S. Osher, T. Fang, P. Speier, MR Image Reconstruction by Using the Iterative Refinement Method and Nonlinear Inverse Scale Space Methods. CAM Report 06–35, UCLA, 2005
104. L. He, M. Burger, S. Osher, Iterative total variation regularization with non-quadratic fidelity. J. Math. Imag. Vis. **26**, 167–184 (2006)
105. W. Hinterberger, O. Scherzer, Variational methods on the space of functions of bounded Hessian for convexification and denoising. Computing **76**, 109–133 (2006)
106. M. Hintermüller, M. Monserrat Rincon-Camacho, Expected absolute value estimators for a spatially adapted regularization parameter choice rule in L1-TV-based image restoration. Inverse Probl. **26**, 085005 (2010)
107. B. Hofmann, B. Kaltenbacher, C. Pöschl, O. Scherzer, A Convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. Inverse Probl. **23**, 987–1010 (2007)
108. Y.M. Huang, M.K. Ng, Y.W. Wen, A new total variation method for multiplicative noise Removal. SIAM J. Imag. Sci. **2**, 20–40 (2009)
109. P.J. Huber, Robust estimation of a location parameter. Ann. Math. Stat. **35**, 73–101 (1964)

110. J. Idier, *Bayesian Approach to Inverse Problems* (Wiley, New York, 2008)
111. M.D. Iordache, J.M. Bioucas-Dias, A. Plaza, Total variation regulatization in sparse hyperspectral unmixing, in *3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing* (IEEE, New York, 2011), pp. 1–4
112. Z. Jin, X. Yang, A variational model to remove the multiplicative noise in ultrasound images. J. Math. Imag. Vis. **74**, 39–62 (2011)
113. E. Jonsson, S.C. Huang, T. Chan, Total Variation Regularization in Positron Emission Tomography. CAM Report 98-48, UCLA, 1998
114. J.P. Kaipio, E. Somersalo, *Statistical and Computational Inverse Problems* (Springer, New York, 2004)
115. S. Keeling, C. Clason, M. Hintermüller, F. Knoll, A. Laurain, G. Winckel, An image space approach to Cartesian based parallel MR imaging with total variation regularization. Med. Image Anal. **16**, 189–200 (2012)
116. S. Kindermann, S. Osher, P.W. Jones, Deblurring and denoising of images by nonlocal functionals. Multiscale Model. Simul. **4**, 1091–1115 (2005)
117. S. Kindermann, S. Osher, J. Xu, Denoising by BV-duality. J. Sci. Comp. **28**, 411–444 (2006)
118. F. Knoll, M. Unger, C. Clason, C. Diwoky, T. Pock, R. Stollberger, Fast reduction of undersampling artifacts in radial MR angiography with 3D total variation on graphics hardware. Magn. Reson. Mater. Phy. **23**, 103–114 (2010)
119. F. Knoll, K. Bredies, T. Pock, R. Stollberger, Second order total generalized variation (TGV) for MRI. Magn. Reson. Med. **65**, 480–491 (2011)
120. F. Knoll, C. Clason, K. Bredies, M. Uecker, R. Stollberger, Parallel imaging with nonlinear reconstruction using variational penalties. Magn. Reson. Med. **67**, 34–41 (2012)
121. M. Lassas, S. Siltanen, Can one use total variation prior for edge preserving Bayesian inversion. Inverse Probl. **20**, 1537–1564 (2004)
122. P. Lax, *Functional Analysis* (Wiley, New York, 2002)
123. T. Le, R. Chartrand, T.J. Asaki, A variational approach to reconstructing images corrupted by Poisson noise. J. Math. Imag. Vis. **27**, 257–263 (2007)
124. J. Lellman, J. Kappes, J. Yuan, F. Becker, C. Schnörr, Convex multi-class image labeling by simplex-constrained total variation Technical report (University of Heidelberg, 2008)
125. C. Lemarechal, C. Sagastizabal, Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries. SIAM J. Optim. **7** 367–385 (1997)
126. J. Lie, J.M. Nordbotten, Inverse scale spaces for nonlinear regularization. J. Math. Imag. Vis. **27**, 41–50 (2007)
127. D.A. Lorenz, Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. J. Inverse Ill-Posed Probl. **16**, 463–478 (2008)
128. Y. Lou, X. Zhang, S. Osher, A. Bertozzi, Image recovery via nonlocal operators. J. Sci. Comput. **42**, 185–197 (2010)
129. C. Louchet, Modeles variationnels et Bayesiens pour le d'ebruitage d'images: de la variation totale vers les moyennes non-locales. PhD thesis, University Paris-Descartes, 2008
130. C. Louchet, L. Moisan, Total Variation as a local filter. SIAM J. Imag. Sci. **4**, 651–694 (2011)
131. R. Luce, S. Perez, Parameter identification for an elliptic partial differential equation with distributed noisy data. Inverse Probl. **15**, 291–307 (1999)
132. M. Lustig, D.L. Donoho, J.M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn. Reson. Med. **58**, 1182–1195 (2007)
133. W. Ma, S. Osher, A TV Bregman Iterative Model of Retinex Theory. CAM-Report 10–13, UCLA, 2010
134. Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations* (AMS, Providence, 2001)
135. G. Mohler, A. Bertozzi, T. Goldstein, S. Osher, Fast TV regularization for 2D maximum penalized likelihood estimation. J. Comput. Graph. Stat. **20**, 479–491 (2011)
136. M. Möller, A Variational Approach for Sharpening High-Dimensional Images. Diploma thesis, WWU Münster, 2009

137. M. Möller, T. Wittman, A. Bertozzi, M. Burger, A variational approach for sharpening high-dimensional images. SIAM J. Imag. Sci. **5**, 150–178 (2012)
138. J.M. Morel, S. Solimini, *Variational Methods for Image Segmentation* (Birkhäuser, Boston, 1995)
139. J. Müller, Parallel Total Variation Minimization. Diploma thesis, WWU Münster, 2008
140. D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math. **42**, 577–685 (1989)
141. F. Natterer, F. Wübbeling, *Mathematical Methods in Image Reconstruction* (SIAM, Philadelphia, 2001)
142. M.K. Ng, W. Wang, A total variation model for retinex. SIAM J. Imag. Sci. **4**, 345–365 (2011)
143. A. Obereder, S. Osher, O. Scherzer, On the use of dual norms in bounded variation type regularization, in *Properties from Incomplete Data*, ed. by R. Klette et al. (Kluwer, Dordrecht, 2005), pp. 373–390
144. A. Obereder, O. Scherzer, A. Kovac, Bivariate density estimation using BV regularisation. Comput. Stat. Data Anal. **51**, 5622–5634 (2007)
145. S. Osher, R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces* (Springer, New York, 2002)
146. S. Osher, O. Scherzer, G-norm properties of bounded variation regularization. Comm. Math. Sci. **2**, 237–254 (2004)
147. S. Osher, A. Sole, L. Vese, Image decomposition and restoration using total variation minimization and the H^{-1} norm. SIAM Multiscale Model. Simul. **1**, 349–370 (2003)
148. S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation based image restoration. Multiscale Model. Simul. **4**, 460–489 (2005)
149. Y. Pan, R.T. Whitaker, A. Cheryauka, D. Ferguson, TV-regularized iterative image reconstruction on a mobile C-ARM CT, in *Proceedings of SPIE Medical Imaging 2010*, vol. 7622 (SPIE, San Diego, 2010)
150. V.Y. Panin, G.L. Zeng, G.T. Gullberg, Total variation regulated EM Algorithm. IEEE Trans. Nucl. Sci. **NS-46**, 2202–2010 (1999)
151. E. Pantin, J.L. Starck, F. Murtagh, Deconvolution and blind deconvolution in astronomy, in *Blind Image Deconvolution: Theory and Applications*, ed. by K. Egiazarian, P. Campisi (CRC Press, Boca Raton, 2007), pp. 277–317
152. K.D. Paulsen, H. Jiang, Enhanced frequency-domain optical image reconstruction in tissues through total-variation minimization. Appl. Opt. **35**, 3447–3458 (1996)
153. M. Persson, D. Bone, H. Elmqvist, Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. Phys. Med. Biol. **46**, 853–866 (2001)
154. G. Peyre, S. Bougleux, L. Cohen, Non-local regularization of inverse problems, in *Proceedings of the 10th European Conference on Computer Vision*. LNCS, vol. 5304 (Springer, Berlin, 2008), pp. 57–68
155. T. Pock, A. Chambolle, H. Bischof, D. Cremers, A convex relaxation approach for computing minimal partitions, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009
156. T. Pock, D. Cremers, H. Bischof, A. Chambolle, Global solutions of variational models with convex regularization. SIAM J. Imag. Sci. **3**, 1122–1145 (2010)
157. T. Pock, A. Chambolle, D. Cremers, H. Bischof, A convex relaxation approach for computing minimal partitions, in *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*, 20–25 June 2009 (2009), pp. 810–817. doi:10.1109/CVPR.2009.5206604
158. Z. Qin, D. Goldfarb, S. Ma, An Alternating Direction Method for Total Variation Denoising, Preprint (Columbia University, 2011)
159. S. Remmele, M. Seeland, J. Hesser, Fluorescence microscopy deconvolution based on Bregman iteration and Richardson-Lucy algorithm with TV regularization, in *Proceedings of the Workshop: Bildverarbeitung fi£¡ur die Medizin 2008* (Springer, Berlin, 2008), pp. 72–76
160. E. Resmerita, Regularization of ill-posed problems in Banach spaces: convergence rates. Inverse Probl. **21**, 1303–1314 (2005)

161. E. Resmerita, O. Scherzer, Error estimates for non-quadratic regularization and the relation to enhancing. Inverse Probl. **22**, 801–814 (2006)
162. W. Ring, Structural properties of solutions to total variation regularization problems. Math. Model. Numer. Anal. **34**, 799–810 (2000)
163. L.I. Rudin, S.J. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
164. L. Rudin, P.L. Lions, S. Osher, Multiplicative denoising and debluring: theory and algorithms, in *Geometric Level Sets in Imaging, Vision, and Graphics*, ed. by S. Osher, N. Paragios (Springer, New York, 2003), pp. 103–119
165. A. Sawatzky, C. Brune, F. Wübbeling, T. Kösters, K. Schäfers, M. Burger, Accurate EM-TV algorithm in PET with low SNR, 2008. IEEE Nuclear Science Symposium Conference Record. doi: 10.1109/NSSMIC.2008.4774392
166. A. Sawatzky, (Nonlocal) Total Variation in Medical Imaging. PhD thesis, WWU Münster, 2011
167. A. Sawatzky, C. Brune, J. Müller, M. Burger, Total variation processing of images with Poisson statistics, in *Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns*, LNCS, vol. 5702 (Springer, Berlin, 2009), pp. 533–540
168. A. Sawatzky, D. Tenbrinck, X. Jiang, M. Burger, A Variational Framework for Region-Based Segmentation Incorporating Physical Noise Models. CAM Report 11–81, UCLA, 2011
169. O. Scherzer, C. Groetsch, Inverse scale space theory for inverse problems, in *Scale-Space and Morphology in Computer Vision. Proceedings of the third International Conference Scale-space 2001*, ed. by M. Kerckhove (Springer, Berlin, 2001), pp. 317–325
170. F. Schoepfer, A.K. Louis, T. Schuster, Nonlinear iterative methods for linear ill-posed problems in Banach spaces. Inverse Probl. **22**, 311–329 (2006)
171. S. Setzer, G. Steidl, Variational methods with higher order derivatives in image processing, in *Approximation XII*, ed. by M. Neamtu, L.L.Schumaker (Nashboro Press, Brentwood, 2008), pp. 360–386
172. S. Setzer, G. Steidl, T. Teuber, Infimal convolution regularizations with discrete l1-type functionals. Comm. Math. Sci. **9**, 797–872 (2011)
173. D.L. Snyder, A.M. Hammoud, R.L. White, Image recovery from data acquired with a charge-coupled-device camera. J. Opt. Soc. Am. A **10**, 1014–1023 (1993)
174. D.L. Snyder, C.W. Helstrom, A.D. Lanterman, M. Faisal, R.L. White, Compensation for readout noise in CCD images. J. Opt. Soc. Am. A **12**, 272–283 (1995)
175. G. Steidl, A note on the dual treatment of higher order regularization functionals. Computing **76**, 135–148 (2006)
176. G. Steidl, S. Didas, J. Neumann, Relations between higher order TV regularization and support vector regression, in *Scale-Space and PDE Methods in Computer Vision*, ed. by R. Kimmel, N. Sochen, J. Weickert (Springer, Berlin, 2005), pp. 515–527
177. G. Steidl, S. Setzer, B. Popilka, B. Burgeth, Restoration of matrix fields by second order cone programming. Computing **81**, 161–178 (2007)
178. D. Strong, T. Chan, Exact Solutions to Total Variation Regularization Problems. CAM-Report 96-46, UCLA, 1996
179. D.M. Strong, J.F. Aujol, T. Chan, Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization. SIAM J. Multiscale Model. Simul. **5**, 273–303 (2006)
180. R. Stück, M. Burger, T. Hohage, The iteratively regularized GaussŰNewton method with convex constraints and applications in 4Pi microscopy. Inverse Probl. **28**, 015012 (2012)
181. E. Tadmor, S. Nezzar, L. Vese, A multiscale image representation using hierarchical (BV;L2) decompositions. Multiscale Model. Simul. **2**, 554–579 (2004)
182. E. Tadmor, S. Nezzar, L. Vese, Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation. Comm. Math. Sci. **6**, 281–307 (2008)
183. Y. Vardi, L.A. Shepp, L. Kaufman, A statistical model for Positron emission tomography. JASA **80**, 8–37 (1985)

184. J. Velikina, S. Leng, G.H. Chen, Limited view angle tomographic image reconstruction via total variation minimization. Proc. SPIE v6510 i1
185. L. Vese, L. Lieu, Image restoration and decomposition via bounded total variation and negative Hilbert-Sobolev spaces. Appl. Math. Optim. **58**, 167–193 (2008)
186. L. Vese, S. Osher, Modeling textures with total variation minimization and oscillating patterns in image processing. J. Sci. Comput. **19**, 533–572 (2003)
187. J. Velikina, S. Leng, G.-H. Chen, Limited view angle tomographic image reconstruction via total variation minimization, in *Proceedings of the SPIE 6510, Medical Imaging 2007: Physics of Medical Imaging, 651020*, 14 March, 2007. doi:10.1117/12.713750; http://dx.doi.org/10.1117/12.713750
188. K.R. Vixie, Some properties of minimizers for the Chan-Esedoglu L1TV functional. ArXiv 0710.3980 (2007)
189. K.R. Vixie, S.P. Morgan, L1TV computes the flat norm for boundaries. Abstr. Appl. Anal. (2007), Article ID 45153
190. C.R. Vogel, Nonsmooth Regularization, in *Inverse Problems in Geophysical Applications*, ed. by H.W. Engl et al. (SIAM, Philadelphia, 1997), pp. 1–11
191. C.R. Vogel, *Computational Methods for Inverse Problems* (SIAM, Philadelphia, 2002)
192. C.R. Vogel, M.E. Oman, Iterative methods for total variation denoising. SIAM J. Sci. Comput. **17**, 227–238 (1996)
193. W.W. Wang, P.L. Shui, X.C. Feng, Variational models for fusion and denoising of multifocus images. IEEE Signal Process. Lett. **15**, 65–68 (2008)
194. J. Weickert, *Anisotropic Diffusion in Image Processing* (Teubner, Stuttgart, 1998)
195. J. Weickert, C. Schnörr, A theoretical framework for convex regularizers in PDE-based computation of image motion. Int. J. Comp. Vis. **45**, 245–264 (2001)
196. M. Werlberger, T. Pock, M. Unger, H. Bischof, Optical flow guided TV-L1 video interpolation and restoration, in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, ed. by Y. Boykov, F. Kahl, V. Lempitsky, F.R. Schmidt. Lecture Notes in Computer Science , vol. 6819 (2011), pp. 273–286
197. M.N. Wernick, J.N. Aarsvold (eds.), *Emission Tomography: The Fundamentals of PET and SPECT* (Academic, San Diego, 2004)
198. D. Wirtz, SEGMEDIX: Development and Application of a Medical Image Segmentation Framework. Diploma thesis, WWU Münster, 2009
199. T. Wittman, Variational Approaches to Digital Image Zooming. PhD thesis, University of Minnesota, 2006
200. J. Xu, S. Osher, Iterative regularization and nonlinear inverse scale space applied to wavelet based denoising. IEEE Trans. Image Proc. **16**, 534–544 (2007)
201. Y.F. Yang, T.T. Wu, Z.F. Pang, Image-zooming technique based on Bregmanized nonlocal total variation regularization. Opt. Eng. **50**, 097008-097008-10 (2011)
202. Q. Zhang, X. Huang, L. Zhang, An energy-driven total variation model for segmentation and object-based classification of high spatial resolution remote-sensing imagery. IEEE Geosci. Rem. Sens. Lett. **10**, 125–129 (2013)
203. X. Zhang, M. Burger, S. Osher, A unified primal-dual algorithm framework based on Bregman iteration. J. Sci. Comp. **3** (2010). doi: 10.1007/s10915-010-9408-8
204. X. Zhang, M. Burger, X. Bresson, S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM J. Imag. Sci. **3**, 253–276 (2010)
205. M. Zhu, S.J. Wright, T.F. Chan, Duality-based algorithms for total variation image restoration. Comp. Optim. Appl. **47**, 377–400 (2010)

http://www.springer.com/978-3-319-01711-2

Level Set and PDE Based Reconstruction Methods in Imaging
Cetraro, Italy 2008, Editors: Martin Burger, Stanley Osher
Burger, M.; Mennucci, A.C.G.; Osher, S.; Rumpf, M.
2013, VII, 319 p. 88 illus., 40 illus. in color., Softcover
ISBN: 978-3-319-01711-2