Iterative Total Variation Schemes for Nonlinear Inverse Problems

Markus Bachmayr[*], Martin Burger[†]

Abstract

In this paper we discuss the construction, analysis, and implementation of iterative schemes for the solution of inverse problems based on total variation regularization. Via different approximations of the nonlinearity we derive three different schemes resembling three well-known methods for nonlinear inverse problems in Hilbert spaces, namely iterated Tikhonov, Levenberg-Marquardt, and Landweber. These methods can be set up such that all arising subproblems are convex optimization problems, analogous to those appearing in image denoising or deblurring.

We provide a detailed convergence analysis and appropriate stopping rules in presence of data noise. Moreover we discuss the implementation of the schemes and the application to distributed parameter estimation in elliptic partial differential equations.

AMS Subject classification: 65K10, 65J22, 65N21, 68U10, 47J30

1 Introduction

Variational methods based on penalization by total variation have become a popular and almost standard approach for the computation of discontinuous solutions of inverse problems (cf. [ROF92; AV94; V02; LP99]). Due to the properties of the total variation functional, the reconstructions exhibit a spatially sparse gradient, i.e. they consist of large constant regions and sharp edges. These properties are very desirable for many inverse problems, where the unknowns describe densities or material functions changing in different regions or objects. The total variation reconstructions allow in particular to separate objects clearly.

Besides their advantages total variation penalization methods suffer from several shortcomings. One of them is the difficulty of constructing efficient computational schemes for the minimization due to nonsmoothness of the total variation. Another is a loss of contrast in reconstructions that can be significant for ill-posed problems. Recently, a novel class of reconstruction schemes with a multi-scale nature has been proposed for total variation approaches in imaging (cf. [OBG$^+$05; BGO$^+$06; HBO06; HMO05]), which can overcome the aforementioned shortcomings. Instead of a single variational problem an iterative scheme (or in the limit a continuous flow in pseudo-time) is used with an appropriate stopping criterion dependent on data noise.

In this paper we investigate possible generalizations of this iterative approach to nonlinear inverse problems. For such nonlinear problems, iterative schemes are very natural, since some iterative approximation is usually needed in any case in order to deal

[*]RWTH Aachen, Aachen Institute for Advanced Study in Computational Engineering Sciences, Schinkelstr. 2, D 52056 Aachen, Germany (bachmayr@aices.rwth-aachen.de)

[†]Westfälische Wilhelms Universität (WWU) Münster, Institut für Numerische und Angewandte Mathematik, Einsteinstr. 62, D 48149 Münster, Germany (martin.burger@wwu.de)

with the nonlinearity. In the schemes we propose, the iterative approach to total variation reconstruction is directly combined with the approximation of the nonlinearity. The type of approximation will then distinguish three different methods, similar to three well-known schemes for nonlinear inverse problems (iterated Tikhonov, Levenberg-Marquardt, Landweber).

We mention that all the schemes discussed here are not restricted in their construction to total variation functionals. Indeed the schemes can be formulated for all common convex regularization functionals, including quadratic functionals, where the standard iterations are recovered, and other nonsmooth functionals such as the ones used in wavelet shrinkage or other sparsity approaches (cf. [DDD04; CDL^{+}98]). The convergence analysis is formulated here for the case of total variation schemes. However, the basic strategy of the proofs is not restricted to this case and can also be adapted to other convex functionals with suitable properties.

In [KSS09], a slightly different construction was used independently to obtain iterative regularization methods for nonlinear ill-posed problems in Banach spaces that are closely related to those considered in this work. However, the methods and convergence results in [KSS09] are formulated under the assumption of a smooth and uniformly convex regularization functional, and therefore do not apply to total variation regularization.

2 Iteration Schemes

Our basic setup in this paper is to consider ill-posed nonlinear operator equations of the form

$$F(x) = y \tag{1}$$

where $F\colon \mathcal{D}(F) \subset X \to H$, $y \in H$ for a Banach space X and a Hilbert space H. In practice, only noisy data $y^\delta \in H$ that are corrupted by numerical and measurement errors are available, where $\delta > 0$ denotes the noise level. We will assume the existence of $\bar{y} \in H$ with $\|y^\delta - \bar{y}\|_H \leq \delta$ and $F(\bar{x}) = \bar{y}$ for some $\bar{x} \in \mathcal{D}(F)$.

The iterative algorithms that will be introduced below are motivated by variational regularization methods, where the regularized solution is obtained as a global minimizer of

$$\frac{1}{2}\|F(x) - y^\delta\|_H^2 + \alpha J(x) \tag{2}$$

with a suitable convex regularization functional $J\colon X \to \mathbb{R} \cup \{+\infty\}$. We are especially interested in the case of total variation regularization, where J is the seminorm

$$J(x) = |x|_{BV(\Omega)} = \sup_{\substack{g \in C_0^\infty(\Omega,\mathbb{R}^n) \\ \|g\|_\infty \leq 1}} \int_\Omega x \operatorname{div} g \tag{3}$$

on the space $X = BV(\Omega)$ of functions of bounded variation on a domain Ω. Note that for functions in the Sobolev space $W^{1,1}(\Omega)$ the identity

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

holds.

In a similar spirit are sparse reconstruction techniques with respect to some orthonormal basis $\{b_k\}_{k=1}^\infty$ of X, which use an ℓ^1-norm for penalization, i.e.

$$J(x) = \sum_{k=1}^\infty |\langle x, b_k\rangle|. \tag{4}$$

A result of this choice is that almost all coefficients $\langle x, b_k \rangle$ will vanish. A typical example are wavelet coefficients, where the ℓ^1-norm is equivalent to the norm in an appropriate Besov space.

Note that the regularization functionals for total variation regularization and sparse reconstructions are nondifferentiable, and due to the nonlinearity of F the least-squares fitting term in (2) need not be convex, in particular for small α. Thus the numerical solution of the corresponding nonconvex and nondifferentiable minimization problem can be quite expensive for nonlinear inverse problems. This issue is addressed by the methods studied in the present work.

A key ingredient for those iterative methods is the *Bregman distance*, which was introduced in [Bre67] and can be interpreted as a generalization of the mean-square distance to more general functionals J. A generalized Bregman distance for J of $x, \tilde{x} \in X$ can be defined as

$$D^J_\xi(x, \tilde{x}) = J(x) - J(\tilde{x}) - \langle \xi,\, x - \tilde{x} \rangle$$

for a subgradient $\xi \in \partial J(\tilde{x})$. Note that for nonsmooth and not strictly convex functionals the Bregman distance is not a strict distance (i.e. it can be zero for $x \neq \tilde{x}$), and it can be multivalued (i.e. for each choice of a subgradient a different distance will be obtained). In our work this issue will however be of less importance, since we only use the Bregman distance for penalizations and all the methods will choose a particular subgradient.

Our starting point is the following iterative regularization method for linear inverse problems recently introduced in [OBG$^+$05],

$$x_{k+1} = \operatorname*{arg\,min}_{x \in BV(\Omega)} \left\{ \frac{1}{2} \|Kx - y^\delta\|^2_H + \alpha D^J_{\xi_k}(x, x_k) \right\}, \tag{5a}$$

$$\xi_{k+1} = \xi_k - \alpha_k^{-1} K^*(Kx_{k+1} - y^\delta), \tag{5b}$$

where in addition to our above assumptions $F(x) = Kx$ with a linear operator $K \in \mathcal{L}(X, H)$. Here $\alpha_k > 0$ can be chosen a priori and large, it is not the regularization parameter. The role of the actual regularization parameter is played by the stopping index k^*, determined by a modified discrepancy principle, at which the iteration is stopped. When the subdifferential of J is multivalued, which is the case for total variation regularization or sparse reconstructions, equation (5b) selects a specific subgradient $\xi_{k+1} \in \partial J(x_{k+1})$, which also lies in the range of the (smoothing) adjoint operator K^*.

In [OBG$^+$05], special attention was paid to the case $J = |\cdot|_{BV(\Omega)}$, $K = \mathrm{I}$, which leads to an iterative method for total variation denoising. For this particular case, several different motivations have been suggested, for instance as matching both grey level values and normal fields [OBG$^+$05] and as a combined denoising and enhancing method [RS06]. The iteration turns out to cure a major shortcoming of standard total variation denoising by considerably reducing its systematic error, i.e. the reduction of contrast in the image. The method was also applied to image deblurring [HMO05] and extended to non-quadratic fitting terms [HBO06], wavelet-based denoising [XO06] and MR imaging [HCO$^+$06]. For arbitrary K and J, the iteration (5) can also be regarded as a generalization of nonstationary iterated Tikhonov regularization. The latter is obtained by choosing J as the square of a Hilbert space norm, in which case (5a) and (5b) coincide (up to the Riesz isomorphism). This interpretation will be our starting point for the present work. We give three possible extensions of the idea to nonlinear operator equations, which can also be regarded as generalizations of certain well-known iterative regularization methods in a Hilbert space context.

In the case of linear operators, iterative methods as the ones considered in this paper have become popular in compressed sensing recently (cf. [CSO09]). Based on results to recover the sparsest solution (cf. [CT06]) it has become classical in this context to minimize the ℓ^1-norm of coefficients in a basis or the total variation subject to a constraint $Ax = y$, where $y \in \mathbb{R}^M$ with M typically rather small. Our approach yields a possibility to realize such an approach for nonlinear constraints, i.e. the minimization of the regularization functional subject to a low number of nonlinear equations.

2.1 Iterated Variational Method

The first method we consider for the nonlinear case can be regarded as a generalization of nonlinear iterated Tikhonov regularization. The iterates are defined analogously to (5) by

$$x_{k+1} \in \operatorname*{arg\,min}_{x \in \mathcal{D}(F)} \Big\{ \frac{1}{2} \|F(x) - y^\delta\|^2 + \alpha_k D^J_{\xi_k}(x, x_k) \Big\}, \tag{6a}$$

$$\xi_{k+1} = \xi_k - \alpha_k^{-1} F'(x_{k+1})^* \big(F(x_{k+1}) - y^\delta\big). \tag{6b}$$

Note that (6b) is an equation for a subgradient $\xi_{k+1} \in \partial J(x_{k+1})$ resulting from the first-order optimality condition corresponding to (6a).

Under standard assumptions (see also Section 3), well-definedness of the iterates, i.e. existence, uniqueness, and stability of the minimization problems to be solved in each step, can be verified using the same arguments as for (2), cf. [RS06].

At a first glance it is not obvious how this scheme provides any computational advantage compared to standard total variation regularization – contrary it seems that a single nonlinear variational problem is replaced by the solution of a sequence of problems of the same type. However, with the choice of an appropriate regularization functional and a sufficiently large α_k, the variational problem to be solved in each iteration can be made locally convex around x_k, so that the global minimum can indeed be computed by local descent methods. This property cannot be guaranteed by using the total variation functional only for penalization, but e.g. by adding a multiple of the squared L^2-norm, which however should not change the smoothing properties of the scheme. In our numerical experiments below we shall verify that this scheme also leads to improved results compared to the standard variational method.

If $J(x) = \frac{\kappa}{2}\|x\|^2 + J_1(x)$ for some Hilbert space norm, $\kappa > 0$, and some convex regularization functional J_1 (e.g. total variation), then (6a) can be rewritten equivalently as

$$x_{k+1} \in \operatorname*{arg\,min}_{x \in \mathcal{D}(F)} \Big\{ \frac{1}{2} \|F(x) - y^\delta\|^2 + \frac{\alpha_k \kappa}{2} \|x - \kappa^{-1}\xi_k\|^2 + \alpha_k J_1(x) \Big\}.$$

Hence, the problem in each iteration step can be written as a minimization with a standard regularization.

2.2 Levenberg-Marquardt-Type Method

In each step of the iterated variational scheme above, some approximation of F will be necessary in order to solve (6a). Hence, one could also consider variations of the scheme by approximating F directly in each iteration step. A first possibility is to approximate the operator by its linearization at the last iterate in each step, which leads to the familiar Levenberg-Marquardt method in a Hilbert space context. In our case we obtain the scheme

$$x_{k+1} = \underset{x \in \mathcal{D}(F)}{\arg\min} \Big\{ \frac{1}{2} \|F(x_k) + F'(x_k)(x - x_k) - y^\delta\|^2 + \alpha_k D^J_{\xi_k}(x, x_k) \Big\}\,, \tag{7a}$$

$$\xi_{k+1} = \xi_k - \alpha_k^{-1} F'(x_k)^* \big(F(x_k) + F'(x_k)(x_{k+1} - x_k) - y^\delta \big)\,. \tag{7b}$$

For this Levenberg-Marquardt-type method, a convex problem has to be solved in each step, where the only nonlinearity comes from the regularization functional. The convex problem in each step of the iteration is of the same type as in (5a), (5b) and therefore well-known and efficient numerical methods for these subproblems are available, e.g. methods based on duality in the case of total variation. Moreover, the well-posedness of the variational problem in (7a) follows with the same arguments as for (5a), cf. [OBG$^+$05].

If again $J(x) = \frac{\kappa}{2}\|x\|^2 + J_1(x)$, we can equivalently rewrite (7a) as

$$x_{k+1} = \underset{x \in \mathcal{D}(F)}{\arg\min} \Big\{ \frac{1}{2} \|F'(x_k)x - d_k\|^2 + \frac{\alpha_k \kappa}{2} \|x - \kappa^{-1}\xi_k\|^2 + \alpha_k J_1(x) \Big\}$$

with $d_k = y^\delta - F(x_k) + F'(x_k)x_k$.

2.3 Landweber-Type Method

A further simplification of each step can be achieved by linearization of the least squares functional, which leads to

$$x_{k+1} = \underset{x \in \mathcal{D}(F)}{\arg\min} \Big\{ \big\langle F'(x_k)^* \big(F(x_k) - y^\delta \big),\, x - x_k \big\rangle + \alpha_k D^J_{\xi_k}(x, x_k) \Big\}\,, \tag{8a}$$

$$\xi_{k+1} = \xi_k - \alpha_k^{-1} F'(x_k)^* \big(F(x_k) - y^\delta \big)\,. \tag{8b}$$

This method reduces to Landweber iteration in the Hilbert space case. If ∂J is single-valued, it is essentially the same as the algorithm described and analysed in [SLS06] for linear inverse problems and in [KSS09] for nonlinear problems, in both cases under the assumption that J is a norm on a smooth and uniformly convex Banach space.

Note that the well-posedness of the variational problem in (8a) follows by the same considerations as for image denoising with iterated total variation methods, cf. [OBG$^+$05].

Concerning implementation, the Landweber-type method is the most straightforward of the three schemes discussed. It can be realized in two subsequent steps: First, the update of the subgradient (8b) can be performed, which requires the same effort as the Landweber iteration in Hilbert spaces – only F and the adjoint of F' have to be evaluated. Subsequently (8a) can be solved, which is a problem similar to image denoising, independent of the operator F. In particular if again $J(x) = \frac{\kappa}{2}\|x\|^2 + J_1(x)$, we can equivalently rewrite (8a) as

$$x_{k+1} = \underset{x \in \mathcal{D}(F)}{\arg\min} \Big\{ \frac{\kappa}{2} \|x - \kappa^{-1} g_k\|^2 + J_1(x) \Big\}$$

with $g_k = \alpha_k^{-1} F'(x_k)^*(y^\delta - F(x_k)) + \xi_k$.

2.4 Stopping Rule

For noisy data, the methods have to be supplied with a suitable stopping rule. It turns out that, similarly to the corresponding methods in Hilbert spaces or to the case of linear operators (5), this can be achieved using modified versions of Morozov's discrepancy principle, i.e. we assume the iterative methods to be stopped at the index $k^*(\delta, y^\delta)$ defined by

$$k^* = \min\{k\colon \|F(x_k) - y^\delta\|_H \le \tau\delta\} \tag{9}$$

with a constant $\tau > 1$. We will formulate further conditions on τ for each method as part of our convergence results below. The regularized solution is given by the iterate x_{k^*}.

We finally mention that for each of these methods, given that $\operatorname{dom} J = \{x \in X\colon\ J(x) < \infty\} \subseteq \mathcal{D}(F)$, the sequence ξ_{k+1} generated by (6b), (7b) and (8b), respectively, satisfies $\xi_{k+1} \in \partial J(x_{k+1})$.

3 Convergence Analysis

To obtain a first convergence analysis, we restrict ourselves to the particular case $X = L^2(\Omega)$, $\Omega \subset \mathbb{R}^2$ a Lipschitz domain, and

$$J(x) = \frac{\kappa}{2}\|x\|^2_{L^2(\Omega)} + |x|_{BV(\Omega)} + \chi_{\mathcal{D}(F)}(x) \tag{10}$$

with some $\kappa > 0$, where we set

$$J_2(x) = \frac{\kappa}{2}\|x\|^2_{L^2(\Omega)}\,, \quad J_1(x) = |x|_{BV(\Omega)} + \chi_{\mathcal{D}(F)}(x)\,.$$

We assume $\mathcal{D}(F)$ to be convex, which ensures that J_1 and J are convex. By [ET99], any $\xi \in \partial J(x)$ can be decomposed as $\xi = \kappa x + p$, $p \in \partial J_1(x)$.

For simplicity, we set $\kappa = 1$. For any different choice of a $\kappa > 0$, the arguments remain valid with appropriate changes to constants.

Without further notice we shall make the following assumptions on the operator F in the sequel:

Assumptions 1. Let $F\colon \mathcal{D}(F) \subset L^2(\Omega) \to H$ be continuous, weakly sequentially closed, i.e. for any sequence $\{x_n\} \in \mathcal{D}(F)$, $x_n \rightharpoonup^* x$ in $BV(\Omega)$ and $F(x_n) \rightharpoonup y$ imply $x \in \mathcal{D}(F)$ and $F(x) = y$, and Fréchet differentiable with $F'(\cdot)$ locally bounded on the closed and convex set $\mathcal{D}(F)$.

To simplify notation, the norm on the image space H will be denoted by $\|\cdot\|$.

3.1 Monotonicity

In the following we first verify the monotonicity of the residuals in the three iteration schemes, which can be shown under very general conditions. A particularly straightforward case is the iterated Tikhonov-type method (6):

Proposition 1. *Let $x_k \in \mathcal{D}(F)$ and $\alpha_k > 0$. Then x_{k+1} defined by the iterated Tikhonov method* (6) *satisfies*

$$\|F(x_{k+1}) - y^\delta\|^2 - \|F(x_k) - y^\delta\|^2 \le -\alpha_k \|x_{k+1} - x_k\|^2_{L^2(\Omega)}\,.$$

Proof. The definition of x_{k+1} as a minimizer of the functional in (6a) in comparison with the functional value at x_k directly implies the assertion. □

In the case of the Levenberg-Marquardt method (7), monotonicity of the residual cannot be shown for small α_k, but with a lower bound:

Proposition 2. *Let $x_k \in \mathcal{D}(F)$ and let F and F' be Lipschitz continuous with respect to the L^2-norm in $B_R(x^k) \cap \mathcal{D}(F)$ for some $R > 0$. Then there exists $\alpha_k > 0$ such that x_{k+1} defined by the Levenberg-Marquardt method* (7) *satisfies*

$$\|F(x_{k+1}) - y^\delta\|^2 - \|F(x_k) - y^\delta\|^2 \le -\alpha_k \|x_{k+1} - x_k\|^2_{L^2(\Omega)} .$$

Proof. From the local Lipschitz continuity of F and F' we find that for some C_0, $C_1 > 0$, the estimates

$$\|F(x_{k+1}) - F(x_k)\| \le C_0 \|x_{k+1} - x_k\|_{L^2(\Omega)} , \tag{11a}$$
$$\|F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k)\| \le C_1 \|x_{k+1} - x_k\|^2_{L^2(\Omega)} \tag{11b}$$

hold. Let $\alpha_k \ge C_0(1 + C_0) + C_1\|F(x_k) - y^\delta\|$.

Comparing the value of the minimizer of x_{k+1} in (7a) with the functional value at x_k we find

$$\frac{1}{2}\|F(x_k) + F'(x_k)(x_{k+1} - x_k) - y^\delta\|^2 + \alpha_k D^J_{\xi_k}(x_{k+1}, x_k) \le \frac{1}{2}\|F(x_k) - y^\delta\|^2.$$

The first term can be estimated as

$$\begin{aligned}
&\|F(x_k) + F'(x_k)(x_{k+1} - x_k) - y^\delta\|^2 \\
&\quad = 2\langle F(x_{k+1}) - y^\delta, F(x_k) + F'(x_k)(x_{k+1} - x_k) - F(x_{k+1})\rangle \\
&\qquad + \|F(x_{k+1}) - y^\delta\|^2 + \|F(x_k) + F'(x_k)(x_{k+1} - x_k) - F(x_{k+1})\|^2 \\
&\quad \ge \|F(x_{k+1}) - y^\delta\|^2 - 2\|F(x_{k+1}) - F(x_k)\|^2 \\
&\qquad + 2\langle F(x_{k+1}) - F(x_k), F'(x_k)(x_{k+1} - x_k)\rangle \\
&\qquad\quad - 2\|F(x_k) - y^\delta\|\|F(x_k) + F'(x_k)(x_{k+1} - x_k) - F(x_{k+1})\| \\
&\quad \ge \|F(x_{k+1}) - y^\delta\|^2 - 2C_0(1 + C_0)\|x_{k+1} - x_k\|^2_{L^2(\Omega)} \\
&\qquad - 2C_1\|F(x_k) - y^\delta\|\|x_{k+1} - x_k\|^2_{L^2(\Omega)} .
\end{aligned}$$

Finally, the Bregman distance is estimated from below by

$$\alpha_k D^J_{\xi_k}(x_{k+1}, x_k) \ge \frac{\alpha_k}{2}\|x_{k+1} - x_k\|^2_{L^2(\Omega)} ,$$

and with the condition on α_k this implies the assertion. □

Finally we verify descent of the residual in the case of the Landweber-type method (8):

Proposition 3. *Let $x_k \in \mathcal{D}(F)$ and let F and F' be Lipschitz continuous with respect to the L^2-norm in $B_R(x^k) \cap \mathcal{D}(F)$ for some $R > 0$. Then there exists $\alpha_k > 0$ such that x_{k+1} defined by the Landweber method* (8) *satisfies*

$$\|F(x_{k+1}) - y^\delta\|^2 - \|F(x_k) - y^\delta\|^2 \le -\alpha_k \|x_{k+1} - x_k\|^2_{L^2(\Omega)} .$$

Proof. As above, we have the estimates (11). Let $\alpha_k \ge C_0^2 + 2C_1\|F(x_k) - y^\delta\|$ and $\xi_k = \kappa x_k + p_k$ with $p_k \in \partial J_1(x_k)$. By convexity of J_1, $\langle x_{k+1} - x_k, p_{k+1} - p_k\rangle \ge 0$, and combining this with the first-order optimality condition for (8a) we obtain

$$\langle x_{k+1} - x_k, F'(x_k)^*(F(x_k) - y^\delta)\rangle \le -\alpha_k\|x_{k+1} - x_k\|^2 .$$

Furthermore, we have

$$\begin{aligned}\|F(x_{k+1})-y^\delta\|^2 - \|F(x_k)-y^\delta\|^2 \\ &= \|F(x_{k+1})-F(x_k)\|^2 + 2\langle F(x_k)-y^\delta, F(x_{k+1})-F(x_k)\rangle \\ &\le C_0^2\|x_{k+1}-x_k\|^2 + 2\langle F'(x_k)^*(F(x_k)-y^\delta),\, x_{k+1}-x_k\rangle \\ &\quad + 2\|F(x_k)-y^\delta\|\,\|F(x_{k+1})-F(x_k)-F'(x_k)(x_{k+1}-x_k)\|\end{aligned}$$

and hence

$$\|F(x_{k+1})-y^\delta\|^2 - \|F(x_k)-y^\delta\|^2 \le \left(-2\alpha_k + C_0^2 + 2C_1\|F(x_k)-y^\delta\|\right)\|x_{k+1}-x_k\|^2_{L^2(\Omega)}\,.$$

which implies the assertion. □

We will use the descent of the residual as a motivation for a heuristic selection criterion for α_k in our numerical examples for the method (8).

3.2 Basic Properties

In the following we discuss some preliminary results needed in the further convergence analysis of the three iterative schemes. We will use the following identity for Bregman distances, which was also employed for convergence analysis of iterative methods e.g. in [CT93] and [OBG+05]: Let $x, \tilde{x}, \hat{x} \in X$, $\tilde{\xi} \in \partial J(\tilde{x})$, $\hat{\xi} \in \partial J(\hat{x})$, then

$$D^J_{\tilde{\xi}}(x,\tilde{x}) - D^J_{\hat{\xi}}(x,\hat{x}) + D^J_{\hat{\xi}}(\tilde{x},\hat{x}) = \langle \tilde{\xi}-\hat{\xi},\, \tilde{x}-x\rangle\,. \quad (12)$$

We denote the Bregman distance corresponding to J by $D_\xi(x,\tilde{x})$ in what follows.

For any $y \in H$, let $\mathcal{S}(y) = \{x \in \mathcal{D}(F) \colon F(x) = y\}$. We make assumptions on the nonlinear operator F that are rather common in the convergence analysis of iterative regularization methods.

Assumptions 2. We assume that F satisfies a nonlinearity condition of the form

$$\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x}-x)\| \le \eta\|x-\tilde{x}\|_{L^2(\Omega)}\,\|F(x)-F(\tilde{x})\|\,, \qquad x,\tilde{x} \in B_\rho(\bar{x}) \cap \mathcal{D}(F)\,, \quad (13)$$

for some $\eta, \rho > 0$, where $\bar{x} \in \mathcal{S}(\bar{y}) \cap \operatorname{dom} J$ and $B_\rho(\bar{x})$ denotes the open ball around $\bar{x}$ of radius ρ in $L^2(\Omega)$.

It has to be mentioned that the condition (13), restricting the nonlinearity of F, is a rather severe one, see [EHN96] for further details. Although there are a number of examples for which it can be verified, such as distributed parameter identification problems, it remains open for many problems of practical interest, e.g. parameter identification from boundary measurements.

As usual for nonlinear problems we can only expect local convergence of the above algorithms, hence the starting values x_0 (also in relation with ξ_0) will need to be close enough to $\bar{x}$ in an appropriate sense. In our convergence analysis, it will turn out that the Bregman distance $D_{\xi_0}(\bar{x}, x_0)$ has to be small. In the following Lemma we make sure that indeed starting values with arbitrarily small Bregman distance to $\bar{x}$ exist.

Lemma 4. *Let $\bar{x} \in BV(\Omega) \cap \mathcal{D}(F)$. For $\alpha > 0$, let $x^\alpha \in BV(\Omega) \cap \mathcal{D}(F)$ be defined by*

$$x^\alpha = \arg\min_{x \in \mathcal{D}(F)} \big\{ \alpha |x|_{BV(\Omega)} + \|x - \bar{x}\|^2_{L^2(\Omega)} \big\} .$$

Then $x^\alpha \to \bar{x}$ in $L^2(\Omega)$ as $\alpha \to 0$ and for any $\gamma > 0$, there is an $\alpha > 0$ and $\xi^\alpha \in \partial J(x^\alpha)$ such that $D^J_{\xi^\alpha}(\bar{x}, x^\alpha) < \gamma$.

Proof. By definition of x^α, for any $\alpha > 0$ we have

$$\alpha |x^\alpha|_{BV(\Omega)} + \|x^\alpha - \bar{x}\|^2_{L^2(\Omega)} \leq \alpha |\bar{x}|_{BV(\Omega)} . \tag{14}$$

This implies $x^\alpha \to \bar{x}$ in $L^2(\Omega)$ and $\limsup_{\alpha \to 0} |x^\alpha|_{BV(\Omega)} \leq |\bar{x}|_{BV(\Omega)}$. Let $\alpha_k \to 0$ and $x_k := x^{\alpha_k}$, then by lower semicontinuity of the BV seminorm we have

$$|\bar{x}|_{BV(\Omega)} \leq \liminf_k |x_k|_{BV(\Omega)} \leq \limsup_k |x_k|_{BV(\Omega)} \leq |\bar{x}|_{BV(\Omega)} ,$$

i.e. $|x_k|_{BV(\Omega)} \to |\bar{x}|_{BV(\Omega)}$ as $k \to \infty$. Together with (14), this implies $\alpha_k^{-1} \|x_k - \bar{x}\|^2_{L^2(\Omega)} \to 0$. For any k we have a $p_k \in \partial J_1(x_k)$ such that $\alpha_k p_k + 2(x_k - \bar{x}) = 0$. Hence we obtain a subgradient

$$\xi_k = p_k + x_k = -2\alpha_k^{-1}(x_k - \bar{x}) + x_k \in \partial J(x_k).$$

Combining this with the decay of $x_k - \bar{x}$ in $L^2(\Omega)$, we obtain

$$\begin{aligned} D^J_{\xi_k}(\bar{x}, x_k) =& |\bar{x}|_{BV(\Omega)} - |x_k|_{BV(\Omega)} + \frac{1}{2}\big(\|\bar{x}\|^2_{L^2(\Omega)} - \|x_k\|^2_{L^2(\Omega)}\big) \\ & - 2\alpha_k^{-1} \|x_k - \bar{x}\|^2_{L^2(\Omega)} - \langle x_k, \bar{x} - x_k \rangle \to 0 , \end{aligned}$$

which proves the assertion. □

Note that the sequence of regularization parameters $\{\alpha_k\}$, as well as the sequence of iterates $\{x_k\}$, depend on δ. We shall use the abbreviations $y_k := F(x_k)$, $K_k := F'(x_k)$, $r^\delta_k := F(x_k) - y^\delta$, $\bar{r}_k := F(x_k) - \bar{y}$ where appropriate to simplify notation.

Under the assumptions stated above, we show weak* convergence in $BV(\Omega)$ as $\delta \to 0$ of the methods (6), (7) and (8). In all three cases, the basic strategy is similar to the one in [OBG+05]. We restrict ourselves to results on semiconvergence for $\delta > 0$ under the above stopping rule, for "exact data" with $\delta = 0$ one can show convergence of the full sequence of iterates by basically the same techniques.

These results should rather be regarded as a first step, because we have to make use of the Hilbert space structure of $L^2(\Omega)$ in dealing with the nonlinearity of F, the methods themselves being applicable in a more general Banach space setting. On the other hand, we obtain a much stronger type of convergence than convergence in $L^2(\Omega)$.

3.3 Convergence of the Iterated Variational Method

We begin with (6); in this case our assumptions are rather restrictive in comparison to the ones necessary for the stationary case (2) (if $x_0 = 0$, the first step of the method actually concides with (2)). To the best of our knowledge, no analogous result for nonlinear iterated Tikhonov regularization in a Hilbert space setting is available in the literature, where this method is usually considered for a fixed number of steps and variable regularization parameters, which allows for weaker assumptions in the convergence analysis.

We start with a fundamental monotonicity result for the error:

Lemma 5. *If for given iterates x_k, ξ_k a minimizer x_{k+1} for* (6a) *satisfies*

$$\|y^\delta - F(x_{k+1}) - F'(x_{k+1})(\bar{x} - x_{k+1})\| \leq \beta \, \|y^\delta - F(x_{k+1})\|\,, \quad 0 < \beta < 1\,, \tag{15}$$

we have $\|y^\delta - F(x_{k+1})\| \leq \|y^\delta - F(x_k)\|$ as well as

$$D_{\xi_{k+1}}(\bar{x}, x_{k+1}) - D_{\xi_k}(\bar{x}, x_k) \leq -\frac{1-\beta}{\alpha_k}\|y^\delta - F(x_{k+1})\|^2\,.$$

Proof. Monotonicity of residuals follows directly from the definition of the method. By (12),

$$D_{\xi_{k+1}}(\bar{x}, x_{k+1}) - D_{\xi_k}(\bar{x}, x_k) + D_{\xi_k}(x_{k+1}, x_k) = \langle \xi_{k+1} - \xi_k,\, x_{k+1} - \bar{x}\rangle\,.$$

Using (6b) we obtain

$$\begin{aligned}\langle \xi_{k+1} - \xi_k,\, x_{k+1} - \bar{x}\rangle &= \alpha_k^{-1}\langle r_{k+1}^\delta,\, K_{k+1}(\bar{x} - x_{k+1})\rangle \\ &= -\alpha_k^{-1}\|r_{k+1}^\delta\|^2 + \alpha_k^{-1}\langle r_{k+1}^\delta,\, r_{k+1}^\delta + K_{k+1}(\bar{x} - x_{k+1})\rangle \\ &\leq -\alpha_k^{-1}\|r_{k+1}^\delta\|\big(\|r_{k+1}^\delta\| - \|r_{k+1}^\delta + K_{k+1}(\bar{x} - x_{k+1})\|\big)\,.\end{aligned}$$

By assumption (15), $\langle \xi_{k+1} - \xi_k,\, x_{k+1} - \bar{x}\rangle \leq -\alpha_k^{-1}(1-\beta)\|r_{k+1}^\delta\|^2$. □

The main result of this section is the (semi-)convergence of iterated variational methods:

Theorem 6. *Let $\gamma < \min\{1/\eta, \rho/2\}$, $0 < \underline{\alpha}(\delta) \leq \alpha_k \leq \overline{\alpha}$, where $\delta^2/\underline{\alpha}(\delta) < 3\gamma^2/4$, and the starting values $x_0 \in \mathcal{D}(F) \cap BV(\Omega)$, $\xi_0 \in L^2(\Omega)$ satisfy $D_{\xi_0}(\bar{x}, x_0) < \gamma^2/8$. Let $\delta_m > 0$, $\{\delta_m\} \to 0$ with corresponding stopping indices $\{k_m^*\}$, where*

$$\tau > (1 + \eta\gamma)/(1 - \eta\gamma)\,, \tag{16}$$

then for every δ_m, the stopping index is finite and every subsequence of $\{x_{k_m^*}\}$ has a subsequence converging to an $x \in \mathcal{S}(\bar{y})$ in the weak topology of $BV(\Omega)$. If furthermore $\mathcal{S}(\bar{y}) \cap \overline{B_\rho(\bar{x})} = \{\bar{x}\}$, $x_{k_m^*} \overset{*}{\rightharpoonup} \bar{x}$ in $BV(\Omega)$.*

Proof. Take $\delta > 0$ arbitrary but fixed and let k^* be the corresponding stopping index, which at this point can possibly be infinite.

Assume $k < k^* - 1$ and $D_{\xi_k}(\bar{x}, x_k) < \gamma^2/8$. By definition of the iterates,

$$\begin{aligned}\frac{1}{2}\tau^2\delta^2 + \alpha_k D_{\xi_k}(x_{k+1}, x_k) &< \frac{1}{2}\|r_{k+1}^\delta\|^2 + \alpha_k D_{\xi_k}(x_{k+1}, x_k) \\ &\leq \frac{1}{2}\delta^2 + \alpha_k D_{\xi_k}(\bar{x}, x_k)\,.\end{aligned}$$

Hence $D_{\xi_k}(x_{k+1}, x_k) \leq D_{\xi_k}(\bar{x}, x_k)$, and in particular $\|x_{k+1} - x_k\|_{L^2(\Omega)} \leq \sqrt{2D_{\xi_k}(\bar{x}, x_k)}$, which combined with the same estimate for $\|\bar{x} - x_k\|$ yields

$$\|\bar{x} - x_{k+1}\|_{L^2(\Omega)} < 2\sqrt{2D_{\xi_k}(\bar{x}, x_k)} < \gamma\,.$$

Thus by (13) we can apply Lemma 5 to obtain $D_{\xi_{k+1}}(\bar{x}, x_{k+1}) \leq D_{\xi_k}(\bar{x}, x_k)$, which by induction implies $\|\bar{x} - x_{k+1}\|_{L^2(\Omega)} < \gamma$ for any $k < k^* - 1$.

Using (13) and $\|r_{k+1}^\delta\| \geq \tau\delta$, we can verify the required nonlinearity condition (15) for noisy data for all $k < k^* - 1$:

$$\begin{aligned}\|r_{k+1}^\delta + K_{k+1}(\bar{x} - x_{k+1})\| &\leq \delta + \|\bar{r}_{k+1} + K_{k+1}(\bar{x} - x_{k+1})\| \\ &\leq \delta + \eta\gamma\|\bar{r}_{k+1}\| \leq (1 + \eta\gamma)\delta + \eta\gamma\|r_{k+1}^\delta\|\,, \\ &\leq \Big(\frac{1}{\tau}(1 + \eta\gamma) + \eta\gamma\Big)\|r_{k+1}^\delta\|\,,\end{aligned}$$

where $\beta := \tau^{-1}(1+\eta\gamma) + \eta\gamma < 1$ by our choice of τ.

Hence for τ as in (16), the assumption (15) of Lemma 5 is satisfied for $k < k^* - 1$. By Lemma 5 we obtain

$$D_{\xi_{k^*-1}}(\bar{x}, x_{k^*-1}) + \sum_{k=0}^{k^*-2} \frac{1-\beta}{\alpha_k} \|r_{k+1}^\delta\|^2 \leq D_{\xi_0}(\bar{x}, x_0)\,.$$

Now for given δ, k^* is necessarily finite because

$$\frac{(k^*-1)\tau^2\delta^2}{\max_{k\leq k^*-2}\alpha_k} \leq \sum_{k=0}^{k^*-2} \frac{1}{\alpha_k}\|r_{k+1}^\delta\|^2 \leq \frac{D_{\xi_0}(\bar{x}, x_0)}{1-\beta}\,. \tag{17}$$

Again by definition of the iterates,

$$\alpha_{k^*-1} D_{\xi_{k^*-1}}(x_{k^*}, x_{k^*-1}) \leq \frac{1}{2}\delta^2 + \alpha_{k^*-1} D_{\xi_{k^*-1}}(\bar{x}, x_{k^*-1})$$

and hence

$$\|x_{k^*} - x_{k^*-1}\|_{L^2(\Omega)} \leq \Big(\frac{\delta^2}{\alpha_{k^*-1}} + \frac{\gamma^2}{4}\Big)^{\frac{1}{2}}\,.$$

Since $\delta^2/\underline{\alpha}(\delta) < 3\gamma^2/4$, this implies $\|x_{k^*} - \bar{x}\| \leq 2\gamma < \rho$. Using convexity of J and expanding the definition of ξ_{k^*},

$$J(x_{k^*}) \leq J(\bar{x}) + \sum_{k=1}^{k^*} \frac{1}{\alpha_{k-1}} \big|\langle r_k^\delta, K_k(\bar{x} - x_{k^*})\rangle\big| + \rho\|\xi_0\|_{L^2(\Omega)}\,.$$

For $k < k^*$, by (15), (13) and monotonicity of $\|r_k^\delta\|$ we have

$$\begin{aligned}
\big|\langle r_k^\delta, K_k(\bar{x} - x_{k^*})\rangle\big| &\leq \|r_k^\delta\|\big(\|K_k(\bar{x} - x_k)\| + \|K_k(x_{k^*} - x_k)\|\big) \\
&\leq \|r_k^\delta\|\big((1+\beta)\|r_k^\delta\| + (1+3\eta\gamma)\|y_k - y_{k^*}\|\big) \\
&\leq (3+\beta+3\eta\rho)\|r_k^\delta\|^2\,.
\end{aligned}$$

For the remaining term in the sum ($k = k^*$), we have

$$\big|\langle r_{k^*}^\delta, K_{k^*}(\bar{x} - x_{k^*})\rangle\big| \leq \tau\delta(1+\eta\rho)\|\bar{r}_{k^*}\| \leq \tau\delta^2(1+\tau)(1+\eta\rho)\,.$$

Combining the above, we obtain

$$J(x_{k^*}) \leq J(\bar{x}) + \frac{3+\beta+3\eta\rho}{1-\beta} D_{\xi_0}(\bar{x}, x_0) + \tau\delta^2(1+\tau)(1+\eta\rho) + \rho\|\xi_0\|_{L^2(\Omega)}$$

and thus $J(x_{k^*})$ is uniformly bounded for small δ.

We choose a sequence $\{\delta_m\}$ with corresponding stopping indices $\{k_m^*\}$ as in our assumption. We have $\|F(x_{k_m^*}) - y^{\delta_m}\| \to 0$, and hence $\|F(x_{k_m^*}) - \bar{y}\| \to 0$ as $\delta_m \to 0$ by definition of the stopping index.

As $J(x_{k_m^*})$ is uniformly bounded and F is weakly sequentially closed, we obtain weak* convergence in $BV(\Omega)$ and weak convergence in $L^2(\Omega)$ of a subsequence of any subsequence of $\{x_{k_m^*}\}$ to an $x \in \mathcal{S}(\bar{y})$.

If the solution is unique in $\overline{B_\rho(\bar{x})}$, a subsequence-of-subsequence argument gives convergence of $x_{k_m^*}$ to $\bar{x}$ in the same sense. □

3.4 Convergence of the Levenberg-Marquardt-type Method

The following analysis for (7) uses ideas from [Han97], where the Levenberg-Marquardt method in a Hilbert space setting was analysed as a regularization method. Again we start with a monotonicity result on the Bregman distance:

Lemma 7. *Let the parameter* α_k *in* (7) *be chosen such that for some* $0 < \mu < 1$,

$$\mu \|y^\delta - F(x_k)\| \le \|y^\delta - F(x_k) - F'(x_k)(x_{k+1} - x_k)\| \le \|y^\delta - F(x_k)\| \,. \tag{18}$$

Additionally we assume that for a $\nu > 1$,

$$\|y^\delta - F(x_k) - F'(x_k)(\bar{x} - x_k)\| \le \frac{\mu}{\nu} \|y^\delta - F(x_k)\| \,. \tag{19}$$

Then the iterates for the scheme (7) *satisfy*

$$D_{\xi_{k+1}}(\bar{x}, x_{k+1}) - D_{\xi_k}(\bar{x}, x_k) \le -\frac{\mu^2(\nu-1)}{\alpha_k \nu} \|y^\delta - F(x_k)\|^2 \,. \tag{20}$$

Proof. By (12),

$$D_{\xi_{k+1}}(\bar{x}, x_{k+1}) - D_{\xi_k}(\bar{x}, x_k) + D_{\xi_k}(x_{k+1}, x_k) = \langle \xi_{k+1} - \xi_k,\, x_{k+1} - \bar{x} \rangle \,.$$

Using (7b) we obtain

$$\begin{aligned} \langle \xi_{k+1} - \xi_k,\, x_{k+1} - \bar{x} \rangle &= -\alpha_k^{-1} \langle r_k^\delta + K_k(x_{k+1} - x_k),\, K_k(x_{k+1} - \bar{x}) \rangle \\ &= -\alpha_k^{-1} \big\langle r_k^\delta + K_k(x_{k+1} - x_k), \\ &\qquad\qquad r_k^\delta + K_k(x_{k+1} - x_k) - r_k^\delta - K_k(\bar{x} - x_k) \big\rangle \\ &\le -\alpha_k^{-1} \|r_k^\delta + K_k(x_{k+1} - x_k)\| \\ &\qquad \big(\|r_k^\delta + K_k(x_{k+1} - x_k)\| - \|r_k^\delta + K_k(\bar{x} - x_k)\| \big) \,. \end{aligned}$$

Combined with (18) and (19), this yields

$$\langle \xi_{k+1} - \xi_k,\, x_{k+1} - \bar{x} \rangle \le -\frac{\mu^2(\nu-1)}{\alpha_k \nu} \|r_k^\delta\|^2 \,.$$

□

In order to obtain a consistent convergence analysis, we make sure that for appropriate parameter choice, condition (18) can indeed be fulfilled:

Lemma 8. *Let* $x_k \in \mathcal{D}(F) \cap BV(\Omega)$, $\xi_k \in L^2(\Omega)$ *where* $\xi_k = x_k + p_k$ *with* $p_k \in \partial J_1(x)$. *For given* $0 < \mu < 1$, *the condition* (18) *is satisfied if* $\alpha_k > 0$ *is chosen such that*

$$\alpha_k \ge \frac{\|F'(x_k)\|}{1-\mu} \Big(q_k^\delta + \sqrt{q_k^\delta \big[(1-\mu)\|F'(x_k)\| + q_k^\delta \big]} \Big) \tag{21}$$

where $q_k^\delta = \|F'(x_k)^*(F(x_k) - y^\delta)\|_{L^2(\Omega)} / \|F(x_k) - y^\delta\|$.

Proof. By convexity of J_1, $\langle x_{k+1} - x_k,\, p_{k+1} - p_k \rangle \ge 0$. Substituting the optimality condition for (7a), which reads

$$\alpha_k(p_{k+1} - p_k) + \alpha_k(x_{k+1} - x_k) + K_k^*\big(r_k^\delta + K_k(x_{k+1} - x_k)\big) = 0 \,,$$

we obtain

$$\big\langle (K_k^* K_k + \alpha_k\, \mathrm{I})^{-1}(-K_k^* r_k^\delta - \alpha_k(p_{k+1} - p_k)),\, p_{k+1} - p_k \big\rangle \geq 0\,.$$

Using continuity of $(K_k^* K_k + \alpha_k\, \mathrm{I})^{-1}$ we may conclude

$$\|\alpha_k(p_{k+1} - p_k)\|_{L^2(\Omega)} \leq \Big(\frac{\alpha_k + \|K_k\|^2}{\alpha_k}\Big)\|K_k^* r_k^\delta\|_{L^2(\Omega)}\,.$$

Again using the optimality condition for (7a) to solve for $x_{k+1} - x_k$, this yields the estimate

$$\|x_{k+1} - x_k\|_{L^2(\Omega)} \leq \alpha_k^{-1}\Big(1 + \frac{\alpha_k + \|K_k\|^2}{\alpha_k}\Big)\|K_k^* r_k^\delta\|_{L^2(\Omega)}\,.$$

Finally, by the second triangle inequality,

$$\begin{aligned}\|r_k^\delta + K_k(x_{k+1} - x_k)\| &\geq \|r_k^\delta\| - \|K_k\|\,\|x_{k+1} - x_k\|_{L^2(\Omega)}\\ &\geq \|r_k^\delta\|\Big(1 - q_k^\delta \frac{\|K_k\|}{\alpha_k}\Big(1 + \frac{\alpha_k + \|K_k\|^2}{\alpha_k}\Big)\Big)\,,\end{aligned}$$

and with

$$\alpha_k \geq \frac{\|K_k\|}{1-\mu}\Big(q_k^\delta + \sqrt{q_k^\delta\big[(1-\mu)\|K_k\| + q_k^\delta\big]}\Big)$$

this yields the first inequality in (18). The second one follows directly by comparing x_{k+1} to x_k in the objective functional for (7a). □

With the above ingredients we can also prove (semi-)convergence of the Levenberg-Marquardt-type method:

Theorem 9. *Let $0 < \gamma < \min\{\mu/\eta, \rho\}$, $x_0 \in \mathcal{D}(F) \cap BV(\Omega)$, $\xi_0 \in L^2(\Omega)$ such that $D_{\xi_0}(\bar{x}, x_0) < \gamma^2/2$, α_k satisfy (21) and $\alpha_k \leq \overline{\alpha}$, and let the stopping index be chosen with a τ such that*

$$\tau > (1 + \eta\gamma)/(\mu - \eta\gamma)\,. \tag{22}$$

Then for given $\delta > 0$, the iterates for (7) are well-defined for $k \leq k^*$, where k^* is finite. If $\delta_m > 0$, $\{\delta_m\} \to 0$ with corresponding stopping indices k_m^*, then every subsequence of $\{x_{k_m^*}\}$ has a subsequence converging to an $x \in \mathcal{S}(\bar{y})$ in the weak topology of $BV(\Omega)$. If $\mathcal{S}(\bar{y}) \cap \overline{B_\rho(\bar{x})} = \{\bar{x}\}$, then $x_{k_m^*} \overset{*}{\rightharpoonup} \bar{x}$ in $BV(\Omega)$.*

Proof. If $\|\bar{x} - x_k\| < \gamma$ and $k < k^*$, i.e. $\tau\delta < \|r_k^\delta\|$, we have

$$\begin{aligned}\|r_k^\delta + K_k^*(\bar{x} - x_k)\| &\leq \delta + \eta\gamma\|\bar{r}_k\| \leq (1 + \eta\gamma)\delta + \eta\gamma\|r_k^\delta\|\,,\\ &\leq \Big(\frac{1 + \eta\gamma}{\tau} + \eta\gamma\Big)\|r_k^\delta\|\,.\end{aligned}$$

As a consequence, (19) holds with $\nu = \mu\tau/(1 + (1+\tau)\eta\gamma) > 1$.

Again by induction, Lemma 7 applies for any $k < k^*$, which gives $\|\bar{x} - x_k\|_{L^2(\Omega)} < \gamma$ for any $k \leq k^*$. Summing the inequalities (20),

$$D_{\xi_{k^*}}(\bar{x}, x_{k^*}) + \sum_{k=0}^{k^*-1} \frac{\mu^2(\nu - 1)}{\alpha_k \nu}\|r_k^\delta\|^2 \leq D_{\xi_0}(\bar{x}, x_0)$$

and hence for some S independent of δ,

$$\sum_{k=0}^{k^*-1} \frac{1}{\alpha_k}\|r_k^\delta\|^2 \leq S\,.$$

It follows analogously to (17) that k^* is finite for given $\delta > 0$.

To use a compactness argument, an estimate for $J(x_{k^*})$ independent of δ is required. Proceeding similarly to Theorem 6 by expanding the definition of ξ_{k^*},

$$\big|\langle \xi_{k^*}, x_{k^*} - \bar{x}\rangle\big| \le \sum_{l=0}^{k^*-1} \frac{1}{\alpha_l}\big|\langle r_l^\delta + K_l(x_{l+1} - x_l), K_l(x_{k^*} - \bar{x})\rangle\big| + \rho\|\xi_0\|_{L^2(\Omega)}\,.$$

For each $0 \le l \le k^* - 1$, using that $\|r_{k^*}^\delta\| < \|r_l^\delta\|$ by definition of the stopping index,

$$\begin{aligned}\big|\langle r_l^\delta + K_l(x_{l+1} - x_l), K_l(x_{k^*} - \bar{x})\rangle\big| &\le \|r_l^\delta\|\big(\|K_l(x_{k^*} - x_l)\| + \|K_l(\bar{x} - x_l)\|\big)\\ &\le \|r_l^\delta\|\big((1 + 2\eta\gamma)\|y_{k^*} - y_l\|\\ &\qquad + (1 + \mu/\nu)\|r_l^\delta\|\big)\,,\\ &\le (3 + 4\eta\gamma + \mu/\nu)\|r_l^\delta\|^2\,.\end{aligned}$$

As a consequence,

$$J(x_{k^*}) \le J(\bar{x}) + (3 + 4\eta\gamma + \mu/\nu)S + \rho\|\xi_0\|_{L^2(\Omega)}\,.$$

Due to the stopping rule we have $\|\bar{r}_{k_m^*}\| \to 0$, thus the statement follows as in the proof of Theorem 6. □

3.5 Convergence of the Landweber-type Method

We finally turn our attention to the Landweber-type method (8). The following results rely on estimates that are quite similar to the analysis of Landweber iteration in a Hilbert space context given in [HNS95].

Lemma 10. *Let $x_k \in \mathcal{D}(F) \cap BV(\Omega)$, $\xi_k \in L^2(\Omega)$ where $\xi_k = x_k + p_k$ with $p_k \in \partial J_1(x)$. Then* (8a) *has a unique minimizer x_{k+1}, and if $\|\bar{x} - x_k\|_{L^2(\Omega)} < \gamma < \rho$ and α_k is chosen such that $\alpha_k \ge (2\|F'(x_k)\|)^2$, then*

$$\begin{aligned}&D_{\xi_{k+1}}(\bar{x}, x_{k+1}) - D_{\xi_k}(\bar{x}, x_k)\\ &\qquad\qquad \le -(2\alpha_k)^{-1}\|F(x_k) - y^\delta\|\big((1 - 2\eta\gamma)\|F(x_k) - y^\delta\| - 2(1 + \eta\gamma)\delta\big)\,.\end{aligned}$$

Proof. Similarly to Lemma 8 we use the optimality condition for (8a), which reads

$$\alpha_k(p_{k+1} - p_k) + \alpha_k(x_{k+1} - x_k) + K_k^* r_k^\delta = 0\,,$$

as well as $\langle x_{k+1} - x_k,\, p_{k+1} - p_k\rangle \ge 0$ to obtain the estimate

$$\|p_{k+1} - p_k\|_{L^2(\Omega)} \le \alpha_k^{-1}\big\|K_k^* r_k^\delta\big\|_{L^2(\Omega)}\,. \tag{23}$$

By (12) we have

$$\begin{aligned}D_{\xi_{k+1}}(\bar{x}, x_{k+1}) - D_{\xi_k}(\bar{x}, x_k) &\le \langle \xi_{k+1} - \xi_k,\, x_{k+1} - \bar{x}\rangle\\ &= \langle \xi_{k+1} - \xi_k,\, x_{k+1} - x_k\rangle + \langle \xi_{k+1} - \xi_k,\, x_k - \bar{x}\rangle\,.\end{aligned}$$

Using (23), for the first term we obtain the bound

$$\begin{aligned}\langle \xi_{k+1} - \xi_k,\, x_{k+1} - x_k\rangle &= \langle \alpha_k^{-1} K_k^* r_k^\delta,\, (p_{k+1} - p_k) + \alpha_k^{-1} K_k^* r_k^\delta\rangle\\ &\le \frac{2\|K_k\|^2}{\alpha_k^2}\|r_k^\delta\|^2\,.\end{aligned}$$

Employing the nonlinearity condition (13), the second term can be estimated by

$$\begin{aligned}\langle \xi_{k+1}-\xi_k,\, x_k-\bar{x}\rangle &= -\alpha_k^{-1}\big\langle r_k^\delta,\, r_k^\delta - r_k^\delta - K_k(\bar{x}-x_k)\big\rangle \\ &\le -\alpha_k^{-1}\|r_k^\delta\|^2 + \alpha_k^{-1}\|r_k^\delta\|\big(\delta + \eta\gamma\|\bar{r}_k\|\big) \\ &\le -\alpha_k^{-1}\|r_k^\delta\|\big((1-\eta\gamma)\|r_k^\delta\| - (1+\eta\gamma)\delta\big)\,.\end{aligned}$$

Combining the two estimates, we arrive at the assertion. □

Theorem 11. *Let* $0<\gamma<\min\{1/(2\eta),\rho\}$, $x_0\in\mathcal{D}(F)\cap BV(\Omega)$, $\xi_0\in L^2(\Omega)$ *such that* $D_{\xi_0}(\bar{x},x_0)<\gamma^2/2$, α_k *satisfy* $0<\underline{\alpha}\le\alpha_k\le\overline{\alpha}$ *and* $\alpha_k\ge(2\|F'(x_k)\|)^2$, *and the stopping index be chosen with* τ *such that*

$$\tau > 2(1+\eta\gamma)/(1-2\eta\gamma)\,. \tag{24}$$

Then for given $\delta>0$, *the iterates for* (8) *are well-defined for* $k\le k^*$, *where* k^* *is finite. If* $\delta_m>0$, $\{\delta_m\}\to 0$ *with corresponding stopping indices* k_m^*, *then every subsequence of* $\{x_{k_m^*}\}$ *has a subsequence converging to an* $x\in\mathcal{S}(\bar{y})$ *in the weak* topology of* $BV(\Omega)$. *If* $\mathcal{S}(\bar{y})\cap\overline{B_\rho(\bar{x})}=\{\bar{x}\}$, $x_{k_m^*}\overset{*}{\rightharpoonup}\bar{x}$ *in* $BV(\Omega)$.

Proof. Using Lemma 10, we again inductively obtain $\|\bar{x}-x_k\|_{L^2(\Omega)}<\gamma$ for any $k\le k^*$ and

$$D_{\xi_{k^*}}(\bar{x},x_{k^*}) + \sum_{k=0}^{k^*-1}\frac{(1-2\eta\gamma)\tau-2(1+\eta\gamma)}{2\alpha_k\tau}\|r_k^\delta\|^2 \le D_{\xi_0}(\bar{x},x_0)\,,$$

where $(1-2\eta\gamma)\tau-2(1+\eta\gamma)>0$ by (24). Now the statement follows analogously to the proof of Theorem 9. □

4 Application to Parameter Identification

In the following we shall discuss the application of the total variation methods to parameter identification problems. In these problems one often seeks parameters that are close to piecewise constants (with unknown numbers of constants on unknown numbers of subdomains), with constants modelling e.g. material parameters in regions of different composition. Here we will investigate two particular identification problems in elliptic partial differential equations with distributed measurements, in one case additionally with boundary measurements.

4.1 Identification of a Reaction Coefficient

Our first test problem can be shown to satisfy the assumptions of our convergence analysis. The problem consists in recovering q from an observation u^δ of a true solution $u\in H^1(\Omega)$ of

$$-\Delta u + qu = f \ \text{ in }\Omega, \quad u=g \ \text{ on }\partial\Omega, \tag{25}$$

where $\Omega\subset\mathbb{R}^2$ is a bounded domain with smooth boundary or a rectangle, $f\in L^2(\Omega)$ and $g\in H^{3/2}(\partial\Omega)$. The nonlinear operator $F\colon\mathcal{D}(F)\subset L^2(\Omega)\to L^2(\Omega)$ is defined as $F(q)=u(q)$, where $u(q)$ is the solution of (25) for parameter q.

This example is taken from [HNS95]. It can be shown that for some $\omega>0$, F is Fréchet differentiable with locally bounded derivative and weakly sequentially closed on

$$\mathcal{D}(F)=\{q\in L^2(\Omega)\colon\ \|q-\bar{q}\|_{L^2(\Omega)}\le\omega \text{ for a } \bar{q}\in L^2(\Omega),\ \bar{q}\ge 0 \text{ a.e.}\}\,,$$

and that $u(q)\in H^2(\Omega)$ for $q\in\mathcal{D}(F)$.

Lemma 12. *Let $q \in \mathcal{D}(F)$ with $\|q\|_{L^2(\Omega)} \le M$ for some $M > 0$. Then there exists an $\eta > 0$ depending only on Ω and M such that*

$$\|F(\tilde{q}) - F(q) - F'(q)(\tilde{q} - q)\|_{L^2(\Omega)} \le \eta \|q - \tilde{q}\|_{L^2(\Omega)} \, \|F(q) - F(\tilde{q})\|_{L^2(\Omega)}$$

for any $\tilde{q} \in \mathcal{D}(F)$.

Proof. Let $q \in \mathcal{D}(F)$ with $\|q\|_{L^2(\Omega)} \le M$. Let $A(q)\colon H^2(\Omega) \cap H_0^1(\Omega) \to L^2(\Omega)$, $u \mapsto -\Delta u + qu$. For $\psi \in L^2(\Omega)$, let $\nu := A(q)^{-1}\psi$. Since $q\nu \in L^{3/2}(\Omega)$, the imbeddings $H_0^1(\Omega) \subset L^6(\Omega)$ and $W_0^{2,3/2}(\Omega) \subset L^\infty(\Omega)$ in addition to regularity for the Poisson equation give

$$\begin{aligned} \|\nu\|_{L^\infty(\Omega)} \lesssim \|\nu\|_{W_0^{2,3/2}(\Omega)} &\lesssim \|\psi - q\nu\|_{L^{3/2}(\Omega)} \\ &\lesssim \|\psi\|_{L^2(\Omega)} + \|q\|_{L^2(\Omega)} \|\nu\|_{L^6(\Omega)} \le C_M \|\psi\|_{L^2(\Omega)} \,. \end{aligned}$$

Let $\tilde{q} \in \mathcal{D}(F)$ and $h := \tilde{q} - q$. With $w \in H^2(\Omega) \cap H_0^1(\Omega)$ defined by

$$w := -A(q)^{-1}\big(hu(q+h) - hu(q)\big)$$

we have $w = F(q+h) - F(q) - F'(q)h$. For any $\varphi \in L^2(\Omega)$ with $\|\varphi\|_{L^2(\Omega)} = 1$,

$$\begin{aligned} \langle w, \varphi \rangle_{L^2(\Omega)} &\le |\langle hu(q+h) - hu(q), A(q)^{-1}\varphi \rangle_{L^2(\Omega)}| \\ &\le \|hu(q+h) - hu(q)\|_{L^1(\Omega)} \|A(q)^{-1}\varphi\|_{L^\infty(\Omega)} \\ &\le C_M \|h\|_{L^2(\Omega)} \|u(q+h) - u(q)\|_{L^2(\Omega)} \,. \end{aligned}$$

Consequently, $\|w\|_{L^2(\Omega)} \le \eta \|q - \tilde{q}\|_{L^2(\Omega)} \|F(q) - F(\tilde{q})\|_{L^2(\Omega)}$ with $\eta = C_M$. □

This implies condition (13). In summary, the convergence results of Section 3 apply to problem (25). The results also hold for the simpler case of the analogue of (25) on an interval $[a,b] \subset \mathbb{R}$ with Dirichlet boundary conditions.

We finally mention that convergence rates for the standard variational method (2) in the Bregman distance corresponding to the regularization functional have been obtained in [RS06]. In particular, the authors demonstrated the results to be applicable to (25) with J as in Section 3.

As a second test problem, we consider a one-dimensional version of (25) with boundary measurements, i.e. identification of a parameter $q\colon [0,1] \to \mathbb{R}$ from boundary values $u_i(1)$ of $u_i \in H^1([0,1])$, $i = 1, \dots, n_s$ solving

$$-u_i'' + qu_i = f_i \ \text{ in } (0,1), \quad u_i(0) = u_0 \in \mathbb{R}\,, \ u_i'(1) = 0 \tag{26}$$

where $f_i \in L^2([0,1])$ are given sources. Though this is an instance of a problem where condition (13) cannot be verified, it is nonetheless of interest how the schemes behave. In particular we shall also take care of the case of small number of measurements n_s (corresponding to few sources) to test a nonlinear version of compressed sensing.

4.2 Identification of a Diffusion Coefficient

We now turn to a problem of identification of a coefficient of a higher-order term, which leads to additional complications because the regularity of the coefficient has more impact on the regularity of the solution.

Here we want to reconstruct q from a solution $u \in H_0^1(\Omega)$ of

$$-\operatorname{div}(q\nabla u) = f \,, \tag{27}$$

where $\Omega \subset \mathbb{R}^2$ is convex and $f \in L^2(\Omega)$.

It can be shown that the parameter-to-solution map $F(q) = u(q)$ is continuous and Fréchet differentiable with locally bounded derivative on $\{q \in L^\infty(\Omega) : q \geq \underline{q} \text{ a.e.}\}$ for any $\underline{q} > 0$. Furthermore, with the additional constraint

$$\mathcal{D}(F) = \{q \in L^\infty(\Omega) : \overline{q} \geq q \geq \underline{q} \text{ a.e.}\}$$

with fixed $\overline{q} > \underline{q} > 0$ it can be shown that with J as in Section 3 – here including the possibility $\kappa = 0$ – the minimization problem (2), and similarly the subproblems arising in (6), (7) and (8) are well-posed. For further details and proofs, we refer to [CKP98] and [B07].

The convergence results of Section 3 do not carry over to (27) directly, because unless additional smoothness assumptions such as $q \in H^1(\Omega)$ are made, which of course is not of interest in our context, the problem cannot be formulated in a Hilbert space framework and suitable nonlinearity conditions are not available. However, our numerical experiments show that the methods still give good results for this problem.

4.3 Results

As outlined above, the numerical implementation of the iterative methods (6), (7) and (8) reduces to a sequence of standard regularization problems in each case, the main computational challenge being the non-differentiability of the BV seminorm. A standard way of dealing with this is to replace $|q|_{BV}$ by a differentiable approximation, e.g.

$$|q|_{BV} \approx \int_\Omega \sqrt{|Dq|^2 + \varepsilon^2} \tag{28}$$

with small $\varepsilon > 0$, cf. [AV94].

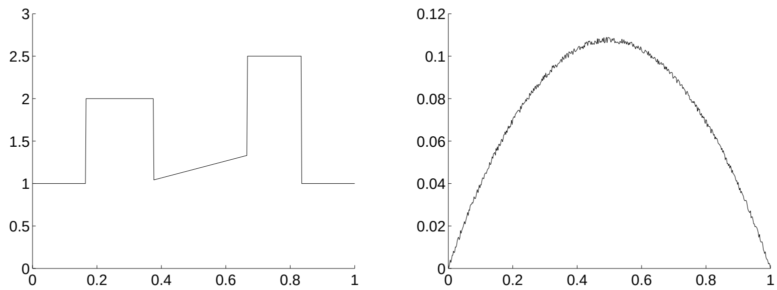
For methods (7) and (8), the minimization problems are convex, and hence methods based on duality, which do not necessarily involve a smoothing of the BV seminorm, become applicable. Examples are the projected gradient descent method of [C04] or the semismooth Newton method given in [HK04]; we have used the latter in our numerical tests.

In the case of the method (6) and for comparison to standard BV regularization, we have to resort to the approximation (28), where we use locally linearly convergent lagged diffusivity iteration [V098] as a minimization method. Using a slightly different differentiable approximation, this method was also used for total variation regularization in the form (2) of nonlinear problems, including the test problem (27), in [AHH06]. To have a direct comparison of the methods, we also use the differentiable approximation (28) within methods (7) and (8) for the example (27), where the minimization problems can be solved by a more efficient primal-dual Newton method [CGM99].

As the regularization functional for the iterative methods, we always use (10). In general, it will be necessary to take care of constraints defining $\mathcal{D}(F)$; for simplicity, our examples are chosen such that these constraints remain inactive.

4.3.1 One-dimensional example with distributed data

As a first example, we consider the one-dimensional case of (25) on the unit interval with boundary conditions $u(0) = u(1) = 0$. We give results for the Levenberg-Marquardt-type method (7), with starting value $q_0 \equiv 1.5$, $\kappa = 0.1$ and α_k some appropriate constant. To better illustrate the convergence behaviour, we choose α_k larger than necessary for this



(a) (b)

Figure 1: (a) Exact parameter $\bar{q}$, (b) distributed input data u^δ with 1% noise for one-dimensional examples.

example. The exact parameter $\bar{q}$ and the corresponding input data u^δ with 1% noise are shown in Figure 1.

We remark that the Landweber-type method (8) gives very similar results, but requires a much higher number of iterations if κ is chosen small.

The subproblems were solved using the semismooth Newton method from [HK04], which in the one-dimensional case can easily be implemented without any artificial smoothing of the BV seminorm. For the discretization, we use piecewise constant elements for q and piecewise linear elements for u and the dual variable arising in the optimization method. The given results were obtained using 2000 elements.

Figure 2 shows the convergence history of the Levenberg-Marquardt-type scheme with 1% and 5% noise and the reconstructions obtained using the discrepancy principle (9); since the parameter τ it involves is not explicitly known, we use the first iterate with residual below the noise level (for larger τ the discrepancy principle tends to oversmooth anyway in the general experience and also in our tests). In fact, the monotonicity of the Bregman distance to the exact solution predicted by the theoretical results is observed to hold also for later iterates, whereas the L^2-error does not show any monotonicity. The contour plots of Figure 3 show the evolution of q_k and ξ_k during the iteration at 1% noise in more detail.

In Figure 4 we compare the reconstruction to the one obtained by the Levenberg-Marquardt method in L^2, corresponding to $J(q) = \frac{\kappa}{2}\|q\|_{L^2}^2$, where for both methods the iterate with minimal L^2-error is shown. As a further illustration of the differences between the two types of regularization, Figure 5 gives results for data without noise. Whereas BV regularization reproduces piecewise constants extremely well, in smoothly varying regions one observes the characteristic “staircasing”. In contrast, L^2 regularization leads to persistent Gibbs oscillations at jumps and to problems at the boundary.

4.3.2 One-dimensional example with boundary data

In a similar setting, we compare to the problem (26) with boundary measurements. We use the boundary conditions $u_i(0) = 1$, $u_i'(1) = 0$ and measurements of $u_i(1)$ for the sources

$$f_i(x) = 10 \exp\left[-10\big(x - i/(1+n_s)\big)^2\right], \quad i = 1, \dots, n_s\,.$$

The implementation of the Levenberg-Marquardt-type method is done as above, in this example using 500 elements.

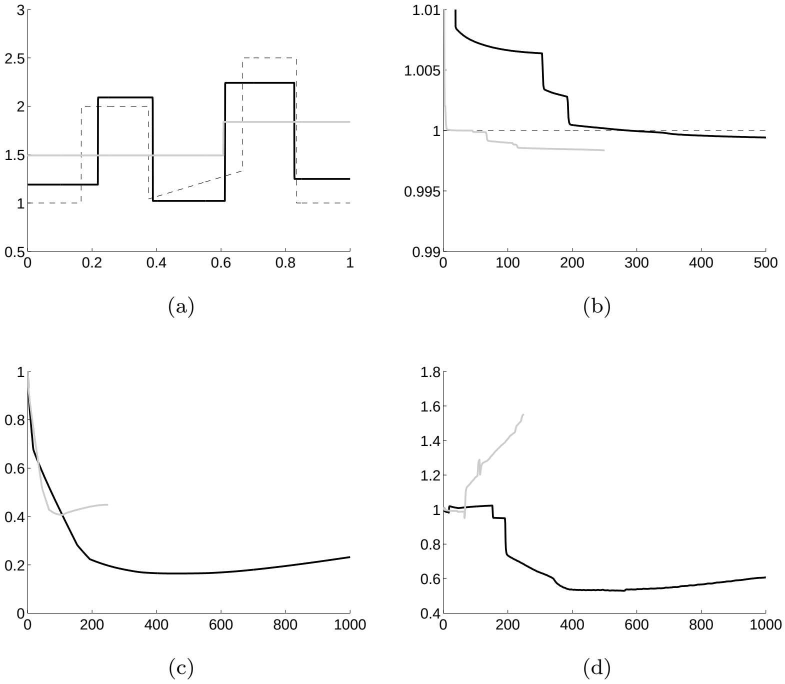


Figure 2: Iteration for 1% (black) and 5% (gray) noise, both with $\alpha_k = 5\cdot 10^{-7}$: (a) reconstructions (discrepancy principle with $\tau = 1$, iterates 287 resp. 25), (b) residuals relative to noise level, (c) relative Bregman distance $D_{\xi_k}(\bar{q}, q_k)/D_{\xi_0}(\bar{q}, q_0)$, (d) relative L^2-error $\|\bar{q} - q_k\|_{L^2}/\|\bar{q} - q_0\|_{L^2}$

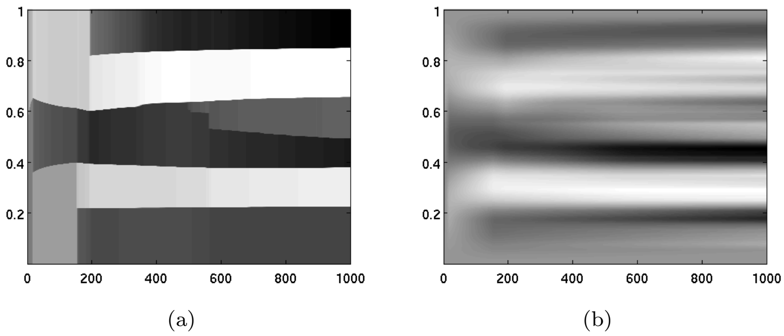


Figure 3: Contour plot of iterates $k = 0, \ldots, 1000$ of (a) q_k and (b) ξ_k with 1% noise as in Figure 2.

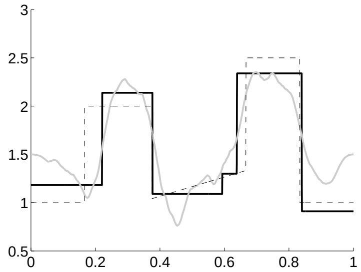
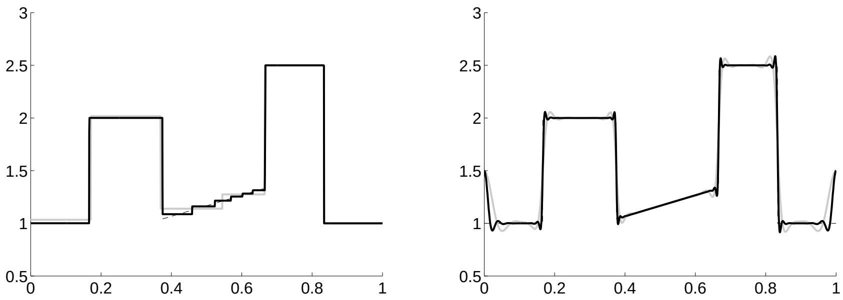


Figure 4: Reconstructions of minimal L^2-error with 1% noise (black) as in Figure 2, compared to that obtained by L^2 regularization (gray).



(a) (b)

Figure 5: Reconstructions of Levenberg-Marquardt-type method for noise-free data, (a) with BV regularization and (b) L^2 regularization only, both with $\kappa = 0.1$, $\alpha_k = 10^{-9}$, iterates 10 (gray), 1000 (black).

Figure 6 shows the results of the Levenberg-Marquardt-type method with $n_s = 50$. Again, the Landweber-type method leads to similar results, but requires a high number of iterations in this case due to rather severe restrictions on the regularization parameters.

As noted above we can also regard our schemes as methods for solving nonlinear compressed sensing problems, the standard approach to which consists in minimizing the regularization functional subject to an underdetermined equation constraint (a nonlinear one in our case). Hence we tested our approach for a very low number of measurements, as an example in Figure 7 we compare the reconstruction of a piecewise constant parameter obtained for $n_s = 5$ without noise to that of the Levenberg-Marquardt method in L^2. The position of the peaks and the edges can be reconstructed reasonably well already with these measurements, but there remains quite some error on the amplitude. Similar experiences have been gained in other computational tests with few measurements.

4.3.3 Two-dimensional example with distributed data

Finally, we consider (27) with $\Omega = B_1(0) \subset \mathbb{R}^2$ and $f \equiv 1$. To obtain a direct comparison of all methods, we use a differentiable approximation of the BV seminorm as described above. For the discretization, we use an unstructured mesh with 8648 triangles and 4421 nodes, where piecewise linear nodal elements are used for both q and u. In all cases, we take $q_0 \equiv 1.5$ and, unless stated otherwise, $\kappa = 0.1$. Here we prescribe for α_k a fixed geometrically decreasing sequence for iterated Tikhonov and Levenberg-Marquardt, whereas for Landweber α_k is chosen by an ad-hoc backtracking scheme that ensures nonincreasing residuals.

Figure 8 shows the exact parameter $\bar{q}$ and the corresponding input data with 5% noise. In Figure (9), results for the iterated Tikhonov- and Levenberg-Marquardt-type methods are compared to those of standard BV regularization, i.e. the stationary method (2) with $J(q) = \frac{\kappa}{2}\|q - q_0\|^2_{L^2(\Omega)} + |q|_{BV(\Omega)}$. In this example, relatively small regularization parameters are prescribed for the two iterative schemes. For comparison, the regularization parameters in (2) are chosen such that the residuals of the iterated Tikhonov method are reproduced; note that this requires a much smaller regularization parameter in the case of the stationary method.

As can be seen from Figure 10, the regularization parameters $\{\alpha_k\}$ that can be used for the Landweber-type method depend strongly on κ.

In Figure 11, the reconstructions of all four methods are compared, all corresponding

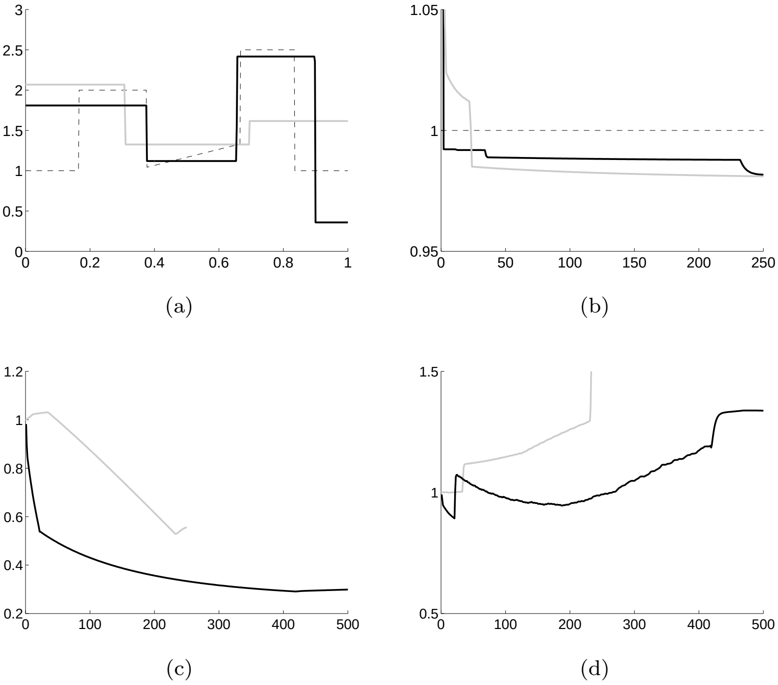


Figure 6: $n_s = 50$ with 1% noise (black, $\alpha_k = 5 \cdot 10^{-3}$), and with 0.1% noise (gray, $\alpha_k = 10^{-4}$): (a) reconstructions (iterates 200 resp. 100), (b) residuals relative to noise level, (c) relative Bregman distance $D_{\xi_k}(\bar{q}, q_k)/D_{\xi_0}(\bar{q}, q_0)$, (d) relative L^2-error $\|\bar{q} - q_k\|_{L^2}/\|\bar{q} - q_0\|_{L^2}$.

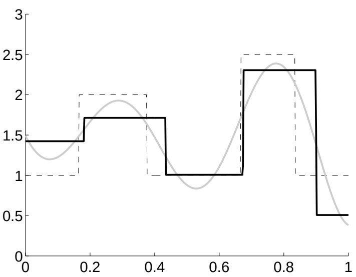


Figure 7: $n_s = 5$ without noise, reconstructions of a piecewise constant parameter (dashed) with BV regularization (black) and L^2 regularization only (gray).

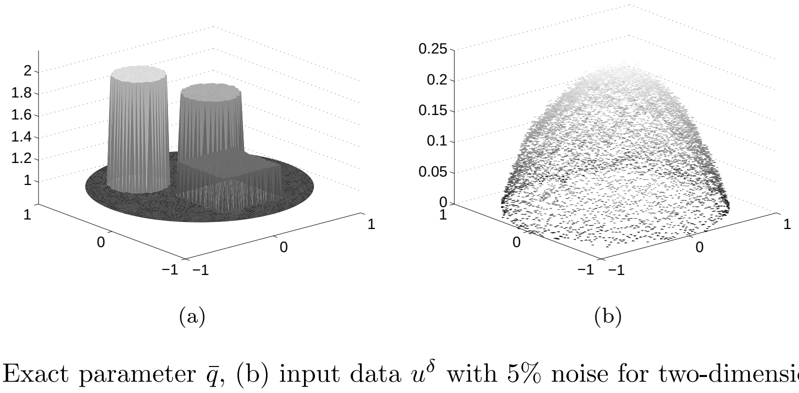


Figure 8: (a) Exact parameter $\bar{q}$, (b) input data u^δ with 5% noise for two-dimensional example.

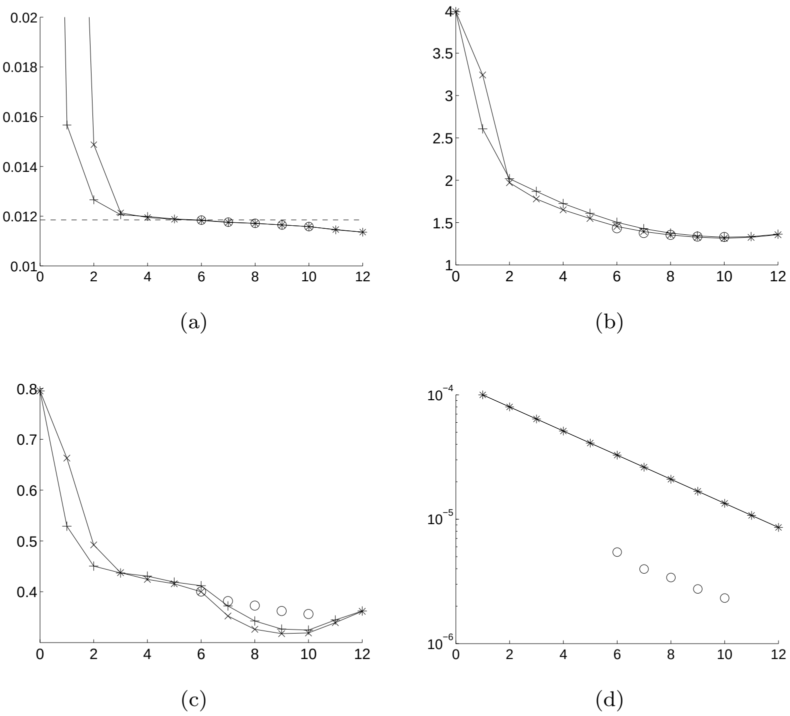


Figure 9: Iterations of Iterated Tikhonov-type method (+), Levenberg-Marquardt-type method ($\times$), compared to stationary BV regularization ($\circ$): (a) residual, (b) Bregman distance $D_{\xi_k}(\bar{q}, q_k)$, (c) L^2-error $\|\bar{q} - q_k\|$, (d) regularization parameters α_k.

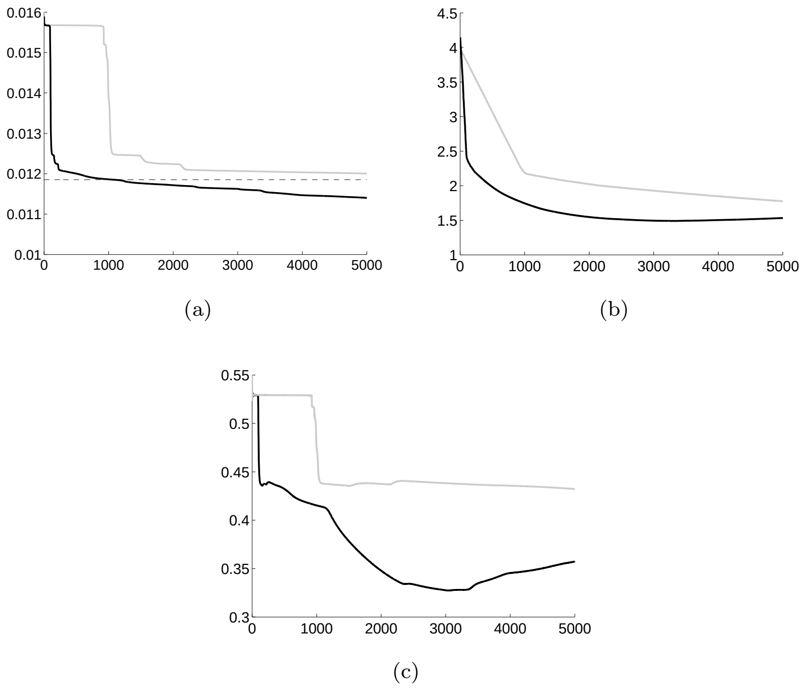
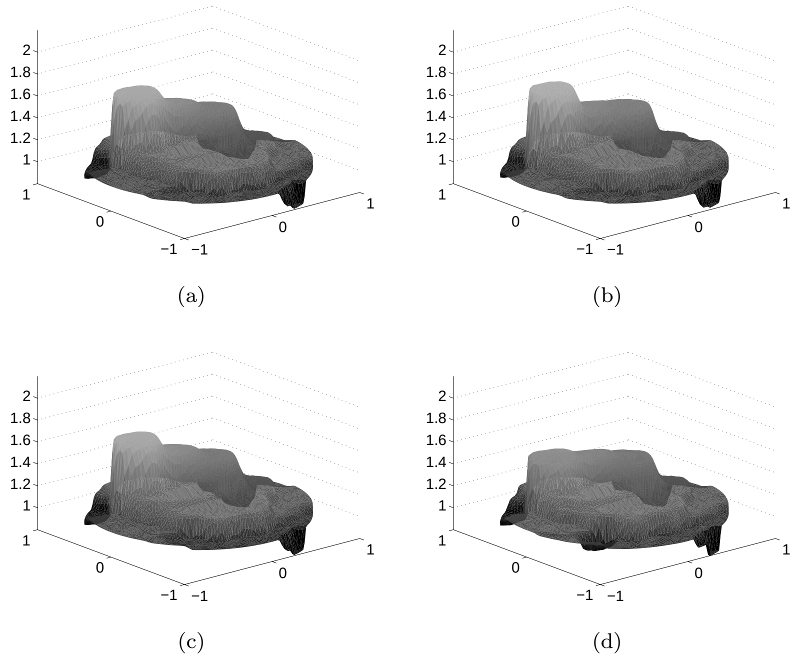


Figure 10: Landweber-type method with $\kappa = 1$ (black) and $\kappa = 0.1$ (gray): (a) residual, (b) Bregman distance $D_{\xi_k}(\bar{q}, q_k)$, (c) L^2-error $\|\bar{q} - q_k\|$; for $\kappa = 1$, $\alpha_k \in [5.36e-3, 1.70e-2]$, for $\kappa = 0.1$, $\alpha_k \in [4.28e-2, 1.62e-1]$.



(a) (b) (c) (d)

Figure 11: Reconstructions of (a) Iterated Tikhonov- (with $\kappa = 0.1$, iterate 8), (b) Levenberg-Marquardt- (with $\kappa = 0.1$, iterate 8), (c) Landweber-type ($\kappa = 1$, iterate 2040) methods and (d) stationary BV regularization (at the same residual, compare Figure 9).

to the same residual slightly below the noise level (compare Figures 9 and 10 for the corresponding regularization parameters).

For further numerical results for this problem and for the two-dimensional case of (25), we refer to [B07].

5 Conclusions

We have described the construction of three iterative methods for total variation regularization of ill-posed nonlinear operator equations and have analysed their convergence under a standard condition on the nonlinearity of the operator. Illustrative applications to parameter identification in elliptic partial differential equations have been given, as well as numerical results that demonstrate the usefulness of the schemes and the improvement compared to standard variational schemes.

An important problem for future research is the construction of efficient methods for the minimization subproblems to be solved in each step of the total variation schemes, in particular for Levenberg-Marquardt and iterated Tikhonov. Indeed the availability or non-availability of efficient schemes for the subproblems in specific applications might be the decisive fact upon choosing one of the three schemes.

The schemes we presented in this paper can actually be applied for more general regularization functionals than just total variation, the main necessary ingredient being convexity of the regularization. Possible examples are several kinds of regularizations enforcing sparsity (ℓ^1-penalization in some basis), entropy functionals, or higher-order total variation functionals recently investigated in imaging applications. Most of the convergence analysis carries over if ones has a Banach space with suitable embedding into a

Hilbert space, on which the operator satisfies appropriate conditions (as the one used in this paper). Some details in the convergence proofs still rely on specific properties of the spaces and regularizations in connection with properties of the operators F. Hence, we suggest that the right conditions should be tuned to the specific problem, keeping our analysis as a guide line for different applications. From a practical point of view, the specific statements of the conditions for convergence might have less impact for iterated Tikhonov and Levenberg-Marquardt methods, whereas they can be crucial for the success of the Landweber iteration.

Acknowledgements

This research was performed when the second author was with the Industrial Mathematics Institute at JKU Linz and the Johann Radon Institute for Computational and Applied Mathematics (Austrian Academy of Sciences), and was financially supported by the Austrian National Science Foundation FWF through project SFB F 013 / 08. Part of this work was completed while the first author was supported by Deutsche Forschungsgemeinschaft DFG grant GSC 111 and the second author by the DFG through the project *Regularization with singular energies.* The authors would like to thank Stanley Osher for stimulating discussions on the methods and Benjamin Hackl and Andreas Neubauer for useful hints concerning the example problems.

References

[AHH06] U. Ascher, E. Haber, and H. Huang. On effective methods for implicit piecewise smooth surface recovery. *SIAM J. Scient. Comput.*, 28:339–358, 2006.

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

[B07] M. Bachmayr. Iterative total variation methods for nonlinear inverse problems. *Master thesis*, Johannes Kepler Universität Linz, 2007.

[Bre67] L. Bregman. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *U.S.S.R. Comput. Math. and Math. Phys.*, 7:200–217, 1967.

[BGO+06] M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. *Comm. Math. Sci.* 4:179–212, 2006.

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

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

[CGM99] T.-F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Scient. Comput.*, 20:1964–1977, 1999.

[CKP98] E. Casas, K. Kunisch, C. Pola. Some applications of BV functions in optimal control and calculus of variations. *ESAIM: Proceedings* 4:83–96, 1998.

[CSO09] J. F. Cai, S. Osher, Z. Shen Linearized Bregman iterations for compressed sensing Math. Comp. 2009, to appear.

[CT93] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim.*, 3:538–543, 1993.

[CT06] E. Candes, T. Tao Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inform. Theory 52: 5406-5425, 2006

[DDD04] I. Daubechies, M. Defrise, and C. DeMol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.* 57:1413-1457, 2004.

[EHN96] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and Its Applications*. Kluwer Academic Publishers, 1996.

[ET99] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*, volume 28 of *Classics in Applied Mathematics*. SIAM, Philadelphia, 1999.

[Han97] M. Hanke. A regularizing Levenberg-Marquardt scheme with applications to inverse groundwater filtration problems. *Inverse Problems*, 13:79–95, 1997.

[HNS95] M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numerische Mathematik*, 72:21–37, 1995.

[HBO06] L. He, M. Burger, and S. J. Osher. Iterative total variation regularization with non-quadratic fidelity. *Journal of Mathematical Imaging and Vision*, 26:167–184, 2006.

[HCO$^+$06] L. He, T.-Ch. Chang, S. J. Osher, T. Fang, and P. Speier. MR image reconstruction by using the iterative refinement method and nonlinear inverse scale space methods. CAM Report 06-35 (UCLA), 2006.

[HMO05] L. He, A. Marquina, and S. J. Osher. Blind deconvolution using TV regularization and Bregman iteration. *International Journal of Imaging Systems and Technology*, 15:74–83, 2005.

[HK04] M. Hintermüller and K. Kunisch, Total bounded variation regularization as bilaterally constrained optimization problem. *SIAM J. Appl. Math.* 64: 1311-1333, 2004.

[KSS09] B. Kaltenbacher and F. Schöpfer and T. Schuster. Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems. *Inverse Problems*, to appear.

[LP99] R. Luce and S. Perez. Parameter identification for an elliptic partial differential equation with distributed noisy data. *Inverse Problems* 15:291-307, 1999.

[OBG$^+$05] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.*, 4(2):460–489, 2005.

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

[RS06] E. Resmerita and O. Scherzer. Error estimates for non-quadratic regularization and the relation to enhancing. *Inverse Problems*, 22:801–814, 2006.

[SLS06] F. Schöpfer, A. K. Louis, and T. Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems*, 22:311–329, 2006.

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

[V098] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Scient. Comput.*, 17:227–238, 1996.

[XO06] J. Xu and S. J. Osher. Iterative regularization and nonlinear inverse scale space applied to wavelet based denoising. CAM Report 06-11 (UCLA), 2006.