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 subproblems to be solved 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 also suffer from several shortcomings. One of them is the difficulty to construct efficient

[*]Research performed at: Institut für Industriemathematik, Johannes Kepler Universität Linz. Current address: Freistädterstr. 430, A 4040 Linz, Austria (markus.bachmayr@students.jku.at)

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

computational schemes for the minimization due to nonsmoothness of the total variation. Another one 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 these shortcomings. Instead of a single variational problem an iterative scheme (or in the limit a continuous flow in pseudo-time) is used with appropriate stopping criterion dependent on data noise.

In this paper we shall 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 anyway in order to deal 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 formulated in a more general way than just for total variation functionals. Indeed the schemes can be constructed for all common convex regularization functionals, including quadratic functionals, where the standard iterations are recovered, strictly convex functionals as considered in [SLS06], or 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, the basic strategy of the proofs, however, is not restricted to this case and can also be adapted to other convex functionals with suitable properties.

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 typical applications, 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 a $\bar{y} \in H$ with $\|y^\delta - \bar{y}\|_H \leq \delta$ and $F(\bar{x}) = \bar{y}$ for a $\bar{x} \in \mathcal{D}(F)$.

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

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

on the space $X = BV(\Omega)$ of functions of bounded variation on the domain Ω. Note that for functions in 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

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

for a subgradient $\xi \in \partial J(\tilde{x})$. Note that for nonsmooth and not strictly 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 ist the following iterative regularization method for linear inverse problems recently introduced in [OBG+05]

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

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

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.

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

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

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

Note that (6b) is an equation for a subgradient $\xi_{k+1} \in \partial J(x_{k+1})$, it can then be interpreted as the first-order optimality condition corresponding to (6a).

Under standard assumptions (see also Section 3), well-definedness of the iterates, i.e. existence, uniqueness, 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 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.

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

$$x_{k+1} = \operatorname*{arg\,min}_{x \in \mathcal{D}(F)} \left\{ \frac{1}{2} \| F(x_k) + F'(x_k)(x - x_k) - y^\delta \|^2 + \alpha_k D^J_{\xi_k}(x, x_k) \right\}, \quad \text{(7a)}$$

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

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 the same 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].

2.3 Landweber-Type Method

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

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

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

This method reduces to Landweber iteration in the Hilbert space case. If ∂J is single-valued, it is essentially the same as the algorithm for linear inverse problems described and analysed in [SLS06] 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 straight-forward of the three schemes discussed, it can be realized in two subsequent steps. First of all, 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.

2.4 Stopping Rule

For noisy data, the methods have to be supplied with a suitable stopping rule. It turns out that, similarly to 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 $\operatorname{dom} J \subseteq \mathcal{D}(F)$, 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,

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

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

$$J_2(x) = \frac{\kappa}{2}\|x\|^2_{L^2(\Omega)}, \quad J_1(x) = |x|_{BV(\Omega)} + \chi_{\mathcal{D}(F)}(x). \tag{11}$$

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 $\xi = \kappa x + p$, $p \in \partial J_1(x)$.

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

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

We denote the Bregman distance corresponding to J by $D_\xi(x,\tilde{x})$ in what follows. For simplicity, we set $\kappa = 1$. For any different choice of a $\kappa > 0$, the arguments remain valid with appropriate changes to constants.

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 for iterative regularization methods.

Assumptions 1. Let $F\colon \mathcal{D}(F) \subset L^2(\Omega) \to H$ be continuous, weakly sequentially closed, i.e. for any sequence $\{x_n\} \in \mathcal{D}(F)$, $x_n \rightharpoonup x$ 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 $\mathcal{D}(F)$. Furthermore, let F satisfy a nonlinearity condition of the form

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

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 exist.

Lemma 1. *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*

$$x^\alpha = \arg\min_{x \in \mathcal{D}(F)} \left\{ \alpha |x|_{BV(\Omega)} + \|x - \bar{x}\|^2_{L^2(\Omega)} \right\} \tag{14}$$

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

$$\alpha |x^\alpha|_{BV(\Omega)} + \|x^\alpha - \bar{x}\|^2_{L^2(\Omega)} \le \alpha |\bar{x}|_{BV(\Omega)} \,. \tag{15}$$

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

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

i.e. $|x_k|_{BV(\Omega)} \to |\bar{x}|_{BV(\Omega)}$ as $k \to \infty$. Together with (15), 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

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

which proves the assertion. □

Note that the sequence of regularization parameters $\{\alpha_k\}$, as well as the sequence of iterates $\{x_k\}$, can depend on δ. In order not to make our notation too complicated, we will not make this explicit in what follows.

We shall use the abbreviations $y_k := F(x_k)$, $K_k := F'(x_k)$, $r_k^\delta := F(x_k) - y^\delta$, $\bar{r}_k := F(x_k) - \bar{y}$ where appropriate to simplify notation. The norm on the image space H will be denoted by $\|\cdot\|$.

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 repeated 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 $L^2(\Omega)$.

3.1 Convergence of Iterated Tikhonov Methods

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

$$\|y^\delta - F(x_{k+1}) - F'(x_{k+1})(\bar{x} - x_{k+1})\| \leq \beta \, \|y^\delta - F(x_{k+1})\| \,, \quad 0 < \beta < 1 \,, \tag{16}$$

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

$$D_{\xi_{k+1}}(\bar{x}, x_{k+1}) - D_{\xi_k}(\bar{x}, x_k) \leq -\frac{1-\beta}{\alpha_k} \|y^\delta - F(x_{k+1})\|^2 \,. \tag{17}$$

Proof. Monotonicity of residuals follows directly from the definition of the method. By Proposition 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 (16), $\langle \xi_{k+1} - \xi_k, \, x_{k+1} - \bar{x} \rangle \leq -\alpha_k^{-1}(1-\beta) \|r_{k+1}^\delta\|^2$. □

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

Theorem 3. *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*

$$\tau > (1 + \eta\gamma)/(1 - \eta\gamma) \,, \tag{18}$$

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,

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

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 $\|\bar{x} - x_k\|$ gives

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

Thus by (13) we can apply Lemma 2 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 $\|r^\delta_{k+1}\| \geq \tau\delta$, we can verify the required nonlinearity condition (16) for noisy data for all $k < k^* - 1$:

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

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

Hence for τ as in (18), the assumption (16) of Lemma 2 is satisfied for $k < k^* - 1$. By Lemma 2 we obtain

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

Now for given δ, k^* has to be finite because

$$\frac{(k^*-1)\tau^2\delta^2}{\max_{k\leq k^*-2}\alpha_k} \leq \sum_{k=0}^{k^*-2} \frac{1}{\alpha_k}\|r^\delta_{k+1}\|^2 \leq \frac{D_{\xi_0}(\bar{x}, x_0)}{1-\beta}\,. \tag{19}$$

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

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

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 (16), (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 summand $k = k^*$, 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 this, we get

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

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. □

3.2 Convergence of Levenberg-Marquardt Methods

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:

Lemma 4. *Let the parameter α_k in (7) be chosen such that for some $0 < \mu < 1$,*

$$\mu\|y^\delta - F(x_k)\| \leq \|y^\delta - F(x_k) - F'(x_k)(x_{k+1} - x_k)\| \leq \|y^\delta - F(x_k)\| . \quad (20)$$

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

$$\|y^\delta - F(x_k) - F'(x_k)(\bar{x} - x_k)\| \leq \frac{\mu}{\nu}\|y^\delta - F(x_k)\| . \quad (21)$$

Then the iterates for the scheme (7) satisfy

$$D_{\xi_{k+1}}(\bar{x}, x_{k+1}) - D_{\xi_k}(\bar{x}, x_k) \leq -\frac{\mu^2(\nu-1)}{\alpha_k\nu}\|y^\delta - F(x_k)\|^2 . \quad (22)$$

Proof. By Proposition 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

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

Combined with (20) and (21), this yields

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

□

In order to obtain a consistent convergence analysis, we make sure that for appropriate parameter choice, condition (20) can indeed be fulfilled:

Lemma 5. *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 (20) is satisfied if $\alpha_k > 0$ is chosen such that*

$$\alpha_k \geq \frac{\|F'(x_k)\|}{1 - \mu} \Big(q_k^\delta + \sqrt{q_k^\delta \big[(1 - \mu) \|F'(x_k)\| + q_k^\delta \big]} \Big) \tag{23}$$

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

$$\alpha_k(p_{k+1} - p_k) + \alpha_k(x_{k+1} - x_k) + K_k^* \big(r_k^\delta + K_k(x_{k+1} - x_k) \big) = 0\,, \tag{24}$$

we obtain

$$\big\langle (K_k^* K_k + \alpha_k \,\mathrm{I})^{-1} (-K_k^* r_k^\delta - \alpha_k(p_{k+1} - p_k)),\, p_{k+1} - p_k \big\rangle \geq 0\,. \tag{25}$$

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

$$\|\alpha_k(p_{k+1} - p_k)\|_{L^2(\Omega)} \leq \Big(\frac{\alpha_k + \|K_k\|^2}{\alpha_k} \Big) \|K_k^* r_k^\delta\|_{L^2(\Omega)}\,. \tag{26}$$

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

$$\|x_{k+1} - x_k\|_{L^2(\Omega)} \leq \alpha_k^{-1} \Big(1 + \frac{\alpha_k + \|K_k\|^2}{\alpha_k} \Big) \|K_k^* r_k^\delta\|_{L^2(\Omega)}\,. \tag{27}$$

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 (20). 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 method:

Theorem 6. *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 (23) and $\alpha_k \leq \overline{\alpha}$, and the stopping index be chosen with a τ such that*

$$
\tau > (1 + \eta\gamma)/(\mu - \eta\gamma) . \tag{28}
$$

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 $\mathcal{S}(\bar{y}) \cap \overline{B_\rho(\bar{x})} = \{\bar{x}\}$, $x_{k_m^*} \overset{*}{\rightharpoonup} \bar{x}$ 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, (21) holds with $\nu = \mu\tau/(1 + (1 + \tau)\eta\gamma) > 1$.

Again by induction, Lemma 4 applies for any $k < k^*$, which gives $\|\bar{x} - x_k\|_{L^2(\Omega)} < \gamma$ for any $k \leq k^*$. Summing the inequalities (22),

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

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 (19) 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 3 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 $\|\bar{r}_{k^*_m}\| \to 0$, thus the statement follows as in the proof of Theorem 3. □

3.3 Convergence of Landweber Methods

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

Lemma 7. *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*

$$\begin{aligned}&D_{\xi_{k+1}}(\bar{x}, x_{k+1}) - D_{\xi_k}(\bar{x}, x_k)\\ &\quad \le -(2\alpha_k)^{-1}\|F(x_k) - y^\delta\|\big((1 - 2\eta\gamma)\|F(x_k) - y^\delta\| - 2(1 + \eta\gamma)\delta\big)\,. \qquad (29)\end{aligned}$$

Proof. Similarly to Lemma 5 we use the optimality condition for (8a), which reads

$$\alpha_k(p_{k+1} - p_k) + \alpha_k(x_{k+1} - x_k) + K_k^* r_k^\delta = 0\,, \qquad (30)$$

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

$$\|p_{k+1} - p_k\|_{L^2(\Omega)} \le \alpha_k^{-1}\big\|K_k^* r_k^\delta\big\|_{L^2(\Omega)}\,. \qquad (31)$$

By Proposition 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 (31), 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 8. *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*

$$\tau > 2(1 + \eta\gamma)/(1 - 2\eta\gamma) \,. \tag{32}$$

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* $\mathcal{S}(\bar{y}) \cap \overline{B_\rho(\bar{x})} = \{\bar{x}\}$, $x_{k_m^*} \xrightharpoonup{*} \bar{x}$ *in* $BV(\Omega)$.

Proof. Using Lemma 7, we again inductively obtain $\|\bar{x} - x_k\|_{L^2(\Omega)} < \gamma$ for any $k \le k^*$ and

$$D_{\xi_{k^*}}(\bar{x}, x_{k^*}) + \sum_{k=0}^{k^*-1} \frac{(1 - 2\eta\gamma)\tau - 2(1 + \eta\gamma)}{2\alpha_k \tau} \|r_k^\delta\|^2 \le D_{\xi_0}(\bar{x}, x_0) \,, \tag{33}$$

where $(1 - 2\eta\gamma)\tau - 2(1 + \eta\gamma) > 0$ by (32). Now the statement follows analogously to the proof of Theorem 6. □

Proposition 9. *Let the assumptions of Lemma 7 hold and assume that for some* C_0, $C_1 > 0$,

$$\begin{aligned}\|F(x_{k+1}) - F(x_k)\| &\le C_0 \|x_{k+1} - x_k\|_{L^2(\Omega)} \,, \\ \|F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k)\| &\le C_1 \|x_{k+1} - x_k\|_{L^2(\Omega)}^2 \,.\end{aligned}$$

Then we have

$$\begin{aligned}&\|F(x_{k+1}) - y^\delta\|^2 - \|F(x_k) - y^\delta\|^2 \\ &\qquad \le \big(-2\alpha_k + C_0^2 + 2C_1 \|F(x_k) - y^\delta\|\big) \|x_{k+1} - x_k\|_{L^2(\Omega)}^2 \,.\end{aligned}$$

Proof. Using $\langle x_{k+1} - x_k,\, p_{k+1} - p_k\rangle \geq 0$ as above, we obtain

$$\langle x_{k+1} - x_k, K_k^* r_k^\delta\rangle \leq -\alpha_k \|x_{k+1} - x_k\|^2 .$$

Furthermore, we have

$$\begin{aligned} \|r_{k+1}^\delta\|^2 - \|r_k^\delta\|^2 &= \|y_{k+1} - y_k\|^2 + 2\langle r_k^\delta, y_{k+1} - y_k\rangle \\ &\leq C_0^2 \|x_{k+1} - x_k\|^2 + 2\|r_k^\delta\|\,\|y_{k+1} - y_k - K_k(x_{k+1} - x_k)\| \\ &\quad + 2\langle K_k^* r_k^\delta,\, x_{k+1} - x_k\rangle \,, \end{aligned}$$

which implies the assertion. □

This shows that for a given previous iterate x_k, one can choose α_k large enough to ensure $\|F(x_{k+1}) - y^\delta\| \leq \|F(x_k) - y^\delta\|$, and furthermore that there is a threshold $\tilde{\alpha} > 0$ such that $\alpha_k \geq \tilde{\alpha}$ for all k implies $\|F(x_{k+1}) - y^\delta\| \leq \|F(x_k) - y^\delta\|$ for all k. We will use this as motivation for a heuristic selection criterion for α_k in our numerical examples. Note that due to the nonlinear dependence of x_{k+1} on α_k, Proposition 9 does not imply $\|F(x_{k+1}) - y^\delta\| < \|F(x_k) - y^\delta\|$.

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 being 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.

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

$$-\Delta u + qu = f \ \text{ in } \Omega, \quad u = g \ \text{ on } \partial\Omega, \tag{34}$$

where $\Omega \subset \mathbb{R}^n$, $n = 2, 3$ is a bounded domain with $C^{1,1}$ boundary, $f \in L^2(\Omega)$ 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 (34) for parameter q.

To gain some insight into the type of ill-posedness and nonlinearity in this example, we can formally solve for q in (34) directly to obtain

$$q = \frac{f + \Delta u}{u} \,. \tag{35}$$

This hints at two causes of instability: On the one hand, u is differentiated twice; on the other hand, the expression can become unbounded as u tends to zero, which reflects the fact that we have no information on q at points where u vanishes unless we make additional assumptions.

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

$$\mathcal{D}(F) = \{q \in L^2(\Omega) \colon \|q - \bar{q}\|_{L^2(\Omega)} \leq \omega \text{ for a } \bar{q} \in L^2(\Omega),\, \bar{q} \geq 0 \text{ a.e.}\}\,. \quad (36)$$

The nonlinearity condition required for the convergence analysis in Section 3 can be verified for this problem. The following result, under slightly rephrased assumptions, is given in [HNS95].

Lemma 10. *Let $\bar{q} \in \operatorname{int}\mathcal{D}(F)$, and assume there is a $\rho < \omega$ such that for some $\kappa > 0$, $F(q) \geq \kappa$ a.e. for all $q \in B_\rho(\bar{q}) \subset \mathcal{D}(F)$, then there exists an $\eta > 0$ such that*

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

for $q, \tilde{q} \in B_\rho(\bar{q})$.

The assumption of $F(q)$ being uniformly bounded away from zero is a rather strong one, considering that according to the heuristic solution (35) it completely rules out a potential source of instability. However, in the two-dimensional case that will be of interest to us in our numerical tests we can drop this assumption.

Lemma 11. *Let $n = 2$. Let $q \in \mathcal{D}(F)$ with $\|q\|_{L^2(\Omega)} \leq C$ for some $C > 0$. Then there exists an $\eta > 0$ depending only on Ω and C such that*

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

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

As there is no restriction on the size of C, this contains condition (13). For a proof, we refer to [B07].

In summary, this shows that the convergence results of Section 3 apply to problem (34). We finally mention that convergence rates for the 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 (34) with J as in Section 3.

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

$$- \operatorname{div}(q \nabla u) = f\,, \tag{37}$$

where $\Omega \subset \mathbb{R}^2$ is convex or a parallelepiped 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 $\{q \in L^\infty(\Omega) \colon q \geq \underline{q} \text{ a.e.}\}$ for any $\underline{q} > 0$. Furthermore, with the additional constraint

$$\mathcal{D}(F) = \{q \in L^\infty(\Omega) \colon \overline{q} \geq q \geq \underline{q} \text{ a.e.}\} \tag{38}$$

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 (37), 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 Numerical Implementation

In the implementation of the iterative methods (6), (7) and (8), we need to handle the nondifferentiability of the BV seminorm in some way. A variety of approaches has been suggested, but as we are dealing with several different classes of total variation minimization problems, the common smoothed approximation of the total variation (3) of q by

$$\tilde{J}_\varepsilon(q) = \int_\Omega \sqrt{|Dq|^2 + \varepsilon^2} \tag{39}$$

with small $\varepsilon > 0$ seems to be most suitable due to its versatility. Note that for several recently proposed approaches without smoothing, the main issue is the application to the subproblems in the iterated Tikhonov method (6), whereas the subproblems in the other methods could also be solved by any scheme for total-variation deblurring or deconvolution, respectively (cf. e.g. [C04; HK04; WYZ07]).

The locally superlinearly convergent *primal-dual Newton method* [CGM99] can be adapted to the minimization problems (7a) and (8a). In the case of the potentially nonconvex problem (6a), however, the scheme can possibly converge to a local maximizer. The latter issue can be avoided by using a *lagged diffusivity iteration* [V098], which is a robust descent method, but shows only locally linear convergence. This method was also used for total variation regularization in the form (2) of nonlinear problems, including the test problem (37), in [AHH06].

The discretization for the test problems (34) and (37) is performed using the finite element method on an unstructured grid, where we use standard P^1 elements for both parameter q and state u.

Each step of the primal-dual Newton method or of the lagged diffusivity iteration requires solving a large linear system of equations. For the primal-dual Newton method applied to (8), one obtains a sparse positive definite system, and using the same method for (7) or using the lagged diffusivity iteration for (6) the respective systems can be written as sparse saddle point problems. Solving these systems in the large-scale case is a strong challenge because the occurring derivatives of the total variation lead to complications in the construction of efficient preconditioners. This point is in fact the most severe limitiation in the implementation of total variation methods for more involved inverse problems or identification in 3D. For our purposes, however, it turns out that employing a sparse direct solver is still reasonably efficient. A further investigation of efficient preconditioning is however an important task for future research.

The numerical methods were implemented in MATLAB with external C modules, using the PDE Toolbox for mesh generation and refinement and the built-in sparse direct solver based on UMFPACK for solving linear systems of equations.

4.4 Results

In our numerical tests, we focus on problem (37), which is of higher practical interest and more challenging from the computational point of view. Our numerical results demonstrate that the methods can be used succesfully also for this kind of problem, not covered by the results of Section 3 due to missing regularity. We again use the regularization functional (10), where our examples are chosen such that the constraints on $\mathcal{D}(F)$ remain inactive. For further numerical experiments, also for the example (34), we refer to [B07].

To arrive at more practicable algorithms, we slightly deviate from the assumptions used in our convergence analysis in some points, for instance in using simplified criteria for selecting the regularization parameters $\{\alpha_k\}$. For (6) and (7), the regularization parameters are chosen as prescribed decreasing geometric sequences. In the case of (8), where convergence of the iteration turns out to depend much more strongly on the choice of regularization parameters, an ad-hoc backtracking scheme motivated by Proposition 9 is used to ensure that the sequence of residuals is nonincreasing. For both examples (34) and (37) this leads to sequences $\{\alpha_k\}$ that remain between certain upper and lower bounds in our numerical experiments.

The used domain $\Omega \subset \mathbb{R}^2$ is the unit disc, on which a mesh of 8648

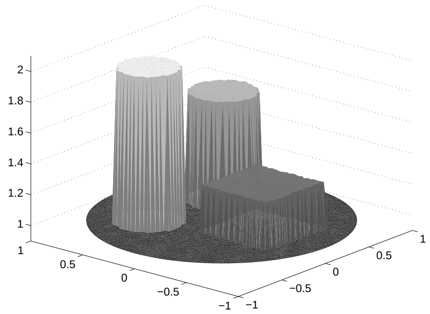


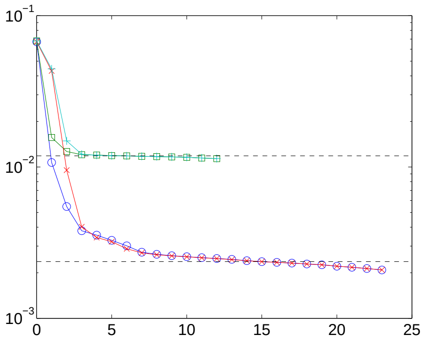
Figure 1: Parameter $\bar{q}$ used to generate sample data.

triangles and 4421 nodes is used in the reconstruction. Right hand side and boundary data are chosen as $f \equiv 1$ and $g \equiv 0$, respectively, for both examples. The parameter $\bar{q}$ used to generate sample data, i.e. the exact solution of the inverse problem, is shown in Figure 1. The data u^δ are generated on an independent, finer grid to avoid "inverse crimes".

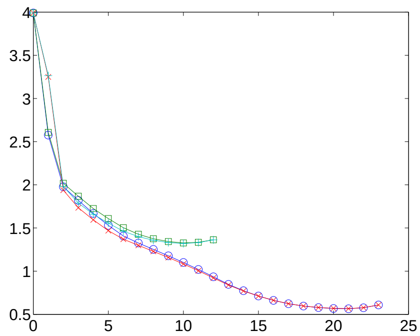
Convergence behaviour and characteristics of reconstructions for (34) and (37) are quite similar; we present results for (37) with 1% and 5% noise in what follows and refer to [B07] for further results. Figure 2 shows the iteration histories of methods (6) and (7) with $\alpha_k = 10^{-4} \cdot 0.8^k$ and $\kappa = 0.1$, which illustrate the expected semiconvergence property. The obtained results are almost identical, but computationally the Levenberg-Marquardt-type scheme (7) is clearly advantageous, since simpler problems need to be solved in each iteration step.

Figure 3 shows a comparison of the first iterate of the Levenberg-Marquardt-type scheme that has a residual below the noise level to stationary total variation regularization (2) with α tuned to the same residual. Typical features of the iterative schemes, which were also observed in all further numerical experiments, can be discerned here, namely that the reconstructions share the basic characteristics of the results of the corresponding stationary method, but have a slightly lower systematic error, i.e. better contrast.

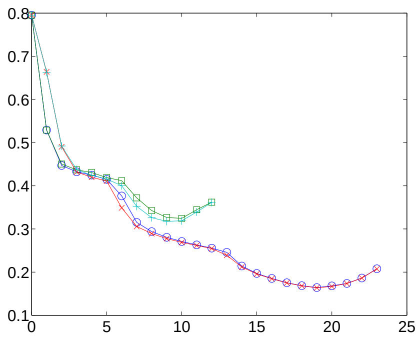
The Landweber-type method (8) leads to qualitatively very similar results, but has different convergence properties comparable to those of Landweber iteration in a Hilbert space. For this reason, it is suitable especially when the minimization problems in each step of the methods (6) and (7) become prohibitively expensive to solve, or for high noise levels, but can be rather inefficient for low noise levels. As to be expected the Landweber iteration turns out to be much more dependent on the choice of κ. The size of admissible parameters α_k is approximately inversely proportional to κ, i.e. a smaller κ requires smaller steps and hence leads to slower convergence. Results for this method are summarized in Table 1.



(a) Residual $\|F(q_k) - u^\delta\|$

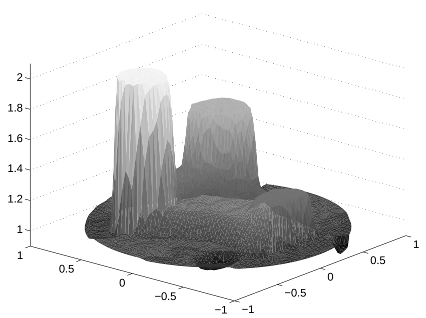


(b) Bregman distance $D^J_{\xi_k}(\bar{q}, q_k)$

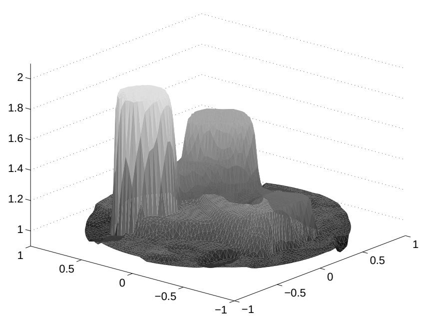


(c) L^2 error $\|\bar{q} - q_k\|$

Figure 2: Results for example (37), with 1% noise ($\circ$ method (6), $\times$ method (7)) and with 5% noise ($\square$ method (6), $+$ method (7)), where the dashed lines in (a) mark the respective noise levels.



(a)



(b)

Figure 3: (a) Iterate 16 of (7) as in Figure 2, data with 1% noise, where $\min_{k\leq 16} \alpha_k = 3.51\text{e–}6$, (b) reconstruction obtained using (2) with the same J and $\alpha = 4.47\text{e–}7$, adjusted to lead to the same residual.

k	Res.	$\kappa = 0.1$ D^J	L^2	Res.	$\kappa = 1$ D^J	L^2
0	$6.807e{-}2$	$3.991e{+}0$	$7.952e{-}1$	$6.807e{-}2$	$4.275e{+}0$	$7.952e{-}1$
1	$3.840e{-}2$	$3.969e{+}0$	$6.388e{-}1$	$1.589e{-}2$	$4.023e{+}0$	$5.350e{-}1$
100	$1.568e{-}2$	$3.793e{+}0$	$5.293e{-}1$	$1.425e{-}2$	$2.302e{+}0$	$4.861e{-}1$
500	$1.567e{-}2$	$3.054e{+}0$	$5.293e{-}1$	$1.200e{-}2$	$1.859e{+}0$	$4.323e{-}1$
1000	$1.391e{-}2$	$2.175e{+}0$	$4.744e{-}1$	$1.186e{-}2$	$1.625e{+}0$	$4.153e{-}1$
5000	$1.200e{-}2$	$1.763e{+}0$	$4.322e{-}1$	*1.140e–2*	*1.418e+0*	*3.574e–1*
6000	$1.195e{-}2$	$1.702e{+}0$	$4.257e{-}1$	*1.129e–2*	*1.446e+0*	*3.784e–1*
10000	$1.186e{-}2$	$1.539e{+}0$	$4.143e{-}1$	–	–	–

Table 1: Results for (37) and the Landweber-type method (8), data with 5% noise, which here means $\delta = 1.185e{-}2$. For $\kappa = 0.1$ one obtains $4.28e{-}2 \leq \alpha_k \leq 1.62e{-}1$, whereas for $\kappa = 1$, $5.36e{-}3 \leq \alpha_k \leq 1.70e{-}2$.

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 also higher-order total variation functionals recently investigated in imaging applications. Most of the convergence analysis carries over if one 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 main guide line for different applications. We mention that from a practical point of view the specific statement 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 (see also our investigations on the importance of adding an L^2-term to the regularization functional).

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. The authors would like to thank Stanley Osher (UCLA) for stimulating discussions on the methods and Benjamin Hackl (MathConsult, Linz) and Andreas Neubauer (JKU Linz) for useful hints concerning the example problems.

References

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

[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.

[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.

[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.

[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.

[CS05] T.-F. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, Philadelphia, 2005.

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

[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.

[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.

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

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

[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.

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

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

[WYZ07] Y. Wang, W. Yin, Y.Zhang. A fast algorithm for image deblurring with total variation regularization. CAM Report 07-22 (UCLA), 2007.

[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.