Electronic Transactions on Numerical Analysis.
Volume 39, pp. 437-463, 2012.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

GRADIENT DESCENT FOR TIKHONOV FUNCTIONALS WITH SPARSITY CONSTRAINTS: THEORY AND NUMERICAL COMPARISON OF STEP SIZE RULES*

DIRK A. LORENZ†, PETER MAASS‡, AND PHAM Q. MUOI‡

Abstract. In this paper, we analyze gradient methods for minimization problems arising in the regularization of nonlinear inverse problems with sparsity constraints. In particular, we study a gradient method based on the subsequent minimization of quadratic approximations in Hilbert spaces, which is motivated by a recently proposed equivalent method in a finite-dimensional setting. We prove convergence of this method employing assumptions on the operator which are different compared to other approaches. We also discuss accelerated gradient methods with step size control and present a numerical comparison of different step size selection criteria for a parameter identification problem for an elliptic partial differential equation.

Key words. nonlinear inverse problems, sparsity constraints, gradient descent, iterated soft shrinkage, accelerated gradient method

AMS subject classifications. 65K10, 46N10, 65M32, 90C48

1. Introduction. We consider operator equations

$$K(u) = f, \tag{1.1}$$

where $K : \mathcal{H}_1 \to \mathcal{H}_2$ is a nonlinear operator between Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. The related inverse problem involves the computation of an approximation to the solution of this operator equation from given noisy data f^δ with

$$\|f - f^\delta\|_{\mathcal{H}_2} \leqslant \delta. \tag{1.2}$$

We are particularly interested in the case of ill-posed equations, which need a stabilization by regularization methods for computing stable approximations.

In this paper, we focus on inverse problems where the solution u has a sparse series expansion $u = \sum_{k\in\Lambda} u_k\varphi_k$ with respect to an orthonormal basis $\{\varphi_k\}_{k\in\Lambda} \subset \mathcal{H}_1$, i.e., the series expansion of u has only a small number of non-vanishing coefficients u_k. Exploiting this sparsity property for a stabilization of the inverse problem, (1.1)–(1.2) leads us to consider the following minimization problem (Tikhonov regularization with sparsity constraint): for a positive regularization parameter α, weights $\omega_k \geq \omega_{\min} > 0$, and an exponent $p \in [1, 2]$, consider

$$\min_{u\in\mathcal{H}_1} \frac{1}{2}\|K(u) - f^\delta\|^2_{\mathcal{H}_2} + \alpha \sum_{k\in\Lambda} \omega_k |\langle u, \varphi_k\rangle|^p. \tag{1.3}$$

Such an approach yields sparse minimizers of (1.3) for $p = 1$. For $1 < p < 2$, this approach is said to promote sparsity [12]. For most of the paper it is convenient to consider the more general class of minimization problems

$$\min_{u\in\mathcal{H}_1} F(u) + \Phi(u), \tag{1.4}$$

* Received December 15, 2011. Accepted August 29, 2012. Published online on November 26, 2012. Recommended by R. Ramlau.

†Institute for Analysis and Algebra, TU Braunschweig, Pockelsstr. 14, D-38118 Braunschweig, Germany (d.lorenz@tu-braunschweig.de).

‡Center for Industrial Mathematics, University of Bremen, Bibliothekstr. 1, D-28334 Bremen, Germany ({pmaass,pham}@math.uni-bremen.de).

where $F(u) := S(K(u), f^\delta)$ is a discrepancy functional that measures the difference between $K(u)$ and f^δ, and $\Phi(u)$ is some regularizing penalty term. Obviously, (1.3) and (1.4) coincide for $F(u) = \frac{1}{2}\|K(u) - f^\delta\|^2_{\mathcal{H}_2}$ and $\Phi(u) = \alpha\Phi_p(u)$ with

$$\Phi_p(u) = \sum_{k\in\Lambda} \omega_k |\langle u, \varphi_k\rangle|^p. \tag{1.5}$$

The problem whether such functionals yield regularizations of the underlying inverse problem (i.e., whether minimizers of (1.3) converge to a solution of (1.1) as $\alpha, \delta \to 0$) has been analyzed intensively for linear and nonlinear settings over the last years; see, e.g., [12, 20, 24, 32, 37]. Recent research has concentrated on developing algorithms for computing minimizers of (1.3). Starting with the pioneering paper [12], where convergence of the iterated soft shrinkage algorithm was proven for linear operator equations, several extensions and generalizations to the case of nonlinear operators have been considered; see, e.g., [6, 7, 38]. Most of these algorithms are known to have a linear convergence rate in theory and are quite slow in practice. The present paper aims at proving convergence results for accelerated gradient methods for nonlinear and ill-posed operator equations as well as at comparing numerically different step size selection criteria.

The motivation for the present paper originates in the results of Bredies et al. [6, 8], Beck and Teboulle [5], and Nesterov [35]. In [5, 35], an efficient scheme for computing a minimizer of the problem (1.4) in the case of a general convex F and specific "simple" Φ is proposed. Although those papers consider the problem in finite-dimensional spaces $\mathbb{R}^d$, the proofs carry over to the Hilbert space setting. Nesterov [35] and Beck and Teboulle [5] also introduced accelerated versions of the gradient method and proved that the objective functional decreases with rate $O(\frac{1}{n^2})$ where n is the iteration counter. These gradient methods are closely related to the generalized conditional gradient method [8] and the generalized projected gradient method [7]. Convergence of this method was proved under fairly general assumptions on F and Φ and a linear convergence rate was obtained in [7] for the case of $F(u) = \frac{1}{2}\|K(u) - f^\delta\|^2_{\mathcal{H}_2}$ with a linear operator K.

In this paper, we combine the algorithmic approach of [35] with the analytic tools developed in [8]. We consider the problem (1.4) where F can be non-convex, i.e., the problem (1.4) includes regularization of nonlinear, ill-posed problems. The gradient method as introduced in [5, 35] as well as some accelerated versions are investigated in a Hilbert space setting. We prove strong convergence of the minimizing sequence generated by the gradient method for the special case of $\Phi = \alpha\Phi_p$ with Φ_p defined by (1.5). We want to emphasize that the assumptions on F needed in the proof of convergence are different from those employed in [8].

The remaining part of this paper is organized as follows: in Section 2, we survey different approaches for deriving first order methods for minimizing functionals of type (1.4). Section 3 is devoted to the convergence analysis of a gradient method derived from successive minimization of quadratic approximations, and Section 4 contains a discussion of the choice of step sizes. In Section 5, we analyze two accelerated versions for the case of convex F. Finally, the algorithms are implemented and analyzed for a parameter identification problem for an elliptic partial differential equation in Section 6.

2. The basic motivations for gradient descent methods.

In this section we summarize several well known approaches for introducing gradient descent methods for the minimization of (1.3) or its generalized version (1.4), respectively. We start by introducing some basic notation.

2.1. Proximal mappings and shrinkage operators.

We will frequently need the notion of the proximal mapping, which is a generalization of the orthogonal projection P_C onto

ETNA
Kent State University
http://etna.math.kent.edu

closed convex sets $C \subset \mathcal{H}$: orthogonal projections are defined as solutions of the minimization problem

$$P_C(v) = \underset{u \in C}{\operatorname{argmin}} \frac{\|u-v\|^2}{2}.$$

Using the indicator function I_C, which takes zero values for $u \in C$ and infinity otherwise, one can rephrase the projection operator as an unconstrained minimization problem $(\lambda > 0)$

$$P_C(v) = \underset{u \in \mathcal{H}}{\operatorname{argmin}} \left(\frac{\|u-v\|^2}{2} + \lambda I_C(u) \right).$$

We now replace I_C by a general convex, coercive, and lower semi-continuous penalty functional Φ and define the generalized projection operator, which is called the proximal mapping of Φ, by

$$P_{\lambda\Phi}(v) = \underset{u \in \mathcal{H}}{\operatorname{argmin}} \left(\frac{\|u-v\|^2}{2} + \lambda \Phi(u) \right).$$

This minimizer u can be characterized using the subdifferential of Φ; it has to satisfy

$$0 \in u - v + \lambda \partial\Phi(u) \;\text{ or }\; v \in (I + \lambda\partial\Phi)(u).$$

Hence, we obtain a well studied equivalence, see [11],

$$P_{\lambda\Phi}(v) = (I + \lambda\partial\Phi)^{-1}(v).$$

The proximal mapping has an explicit expression in terms of shrinkage operators for penalty functionals of the type Φ_p from (1.5). For $1 \le p < \infty$ and $\tau > 0$, define the real valued shrinkage function $S_{\tau,p} : \mathbb{R} \to \mathbb{R}$ by

$$S_{\tau,p}(x) = \begin{cases} \operatorname{sgn}(x)\max(|x| - \tau, 0) & \text{for } p = 1 \\ G_{\tau,p}^{-1}(x) & \text{for } p \in (1,2] \end{cases}, \tag{2.1}$$

where

$$G_{\tau,p}(x) = x + \tau p \operatorname{sgn}(x)|x|^{p-1} \text{ for } 1 < p \leqslant 2. \tag{2.2}$$

DEFINITION 2.1. *Denote $\omega = \{\omega_k\}_{k\in\Lambda}$, with $\omega_k \geqslant \omega_{min} > 0$ for all k, and assume that $\{\varphi_k\}_{k\in\Lambda}$ is an orthonormal basis of $\mathcal{H}$. Let $S_{\omega_k,p}$ denote the shrinkage functions as given in (2.1). The soft shrinkage operator $\mathbb{S}_{\omega,p} : \mathcal{H} \to \mathcal{H}$ is then defined as*

$$\mathbb{S}_{\omega,p}(v) = \sum_{k\in\Lambda} S_{\omega_k,p}(\langle v, \varphi_k \rangle)\varphi_k.$$

For penalty functionals of type (1.5), we obtain the well known equivalence, see, e.g., [11],

$$P_{\alpha\Phi_p}(v) = \mathbb{S}_{\alpha\omega,p}(v). \tag{2.3}$$

We now state different motivations for gradient type methods for the minimization of (1.3) and (1.4).

ETNA
Kent State University
http://etna.math.kent.edu

2.2. First order optimality conditions and gradient descent methods. The classical approach for designing gradient descent methods is based on the first order optimality conditions. Let

$$\Theta(u) := \frac{1}{2}\|K(u) - f^\delta\|^2 + \alpha\Phi_p(u)$$

with Φ_p as in (1.5). The first order optimality condition for a minimizer u is given by

$$0 \in \partial\Theta(u) = K'(u)^*(K(u) - f^\delta) + \alpha\partial\Phi_p(u).$$

Multiplying by λ and adding u on both sides yields a fixed point relation which has to be satisfied for a minimizer u and for all $\lambda \in \mathbb{R}$

$$u \in u + \lambda K'(u)^* \left(K(u) - f^\delta\right) + \lambda\alpha\partial\Phi_p(u).$$

Turning this into an iteration and choosing $s = -\lambda$ yields the classical gradient descent method. However, the convergence analysis of this method relies on higher order smoothness properties for K and Φ, which are not met for sparsity constraints. Hence, in our case it is more appropriate to study iteration methods which are obtained in a slightly different way. Reordering the fixed point relation yields

$$u - \lambda K'(u)^* \left(K(u) - f^\delta\right) \in u + \lambda\alpha\partial\Phi_p(u) = \left(I + \lambda\alpha\partial\Phi_p\right)(u).$$

We can turn this into an iteration by demanding that

$$u^k - \lambda K'(u^k)^* \left(K(u^k) - f^\delta\right) \in u^{k+1} + \lambda\alpha\partial\Phi_p(u^{k+1}) = \left(I + \lambda\alpha\partial\Phi_p\right)(u^{k+1}).$$

The expression on the right-hand side is inverted by the proximal mapping for Φ_p, and hence (2.3) yields the iteration

$$u^{k+1} = \mathbb{S}_{\lambda\alpha\omega,p}\left(u^k - \lambda K'(u^k)^* \left(K(u^k) - f^\delta\right)\right). \tag{2.4}$$

This iteration is the most widely used iterated soft shrinkage algorithm as analyzed in [12] for linear operators and, e.g., in [6, 8] for nonlinear operators. This procedure can be interpreted as first taking a gradient descent step with respect to $\frac{1}{2}\|K(u) - f^\delta\|^2$, i.e., computing $v^k = u^k - \lambda K'(u^k)^* \left(K(x^k) - f^\delta\right)$, and then taking care of the penalty term by determining the shrinkage

$$u^{k+1} = S_{\lambda\alpha\omega,p}(v^k).$$

We could combine some of the parameters ω, α, and λ in order to reduce notation. However, these parameters have different meanings: ω allows to model weighted ℓ_p-spaces and is chosen a priorily, α is a regularization parameter, which has to be chosen carefully, and λ is a step size parameter.

2.3. The generalized gradient projection method. We follow the approach described in [7]. Constrained optimization procedures for solving

$$\min_{u\in C} F(u),$$

where $C \subset \mathcal{H}$ denotes a convex set, are well established; see, e.g., [13, 14, 15, 16, 19, 36]. Projected gradient methods for solving such a problem generate a sequence $\{u^k\}$ by first performing a gradient descent step with respect to F followed by a projection P_C onto the set C, i.e.,

$$z^k = u^k - s_k F'(u^k) \quad \text{and} \quad u^{k+1} = P_C(z^k)$$

with some suitable step size s_k.

We can rephrase the constrained optimization problem as an unconstrained problem by choosing an arbitrary $\lambda > 0$ and using the indicator function I_C as before. Replacing the indicator function by a general convex penalty functional Φ and replacing P_C by the proximal mapping $P_{\lambda\Phi}$ yields the following algorithm:

ALGORITHM 2.2 (Generalized gradient projection method).
Choose u^0 and iterate for $k > 0$
1. *determine a value λ_k, e.g., $\lambda_k = \lambda$ constant for all k*
2. *determine $z^k = u^k - \lambda_k F'(u^k)$*
3. *determine $u^{k+1} = P_{\lambda_k\Phi}(z^k) = \operatorname{argmin}_{u\in\mathcal{H}} \left(\frac{\|u-z^k\|^2}{2} + \lambda_k \Phi(x) \right)$.*

We observe that for the special case $\Phi = \alpha\Phi_p$, the proximal mapping coincides with the shrinkage operator, and hence by inserting $F(u) = \|K(u) - f^\delta\|^2/2$, we obtain the familiar iterated soft shrinkage algorithm

$$u^{k+1} = S_{\lambda_k\alpha\omega,p}\left(u^k - \lambda_k K'(u^k)^*(K(u^k) - f^\delta)\right).$$

The convergence properties of the generalized projected gradient method has been analyzed in [7] for convex F. In particular, a linear convergence rate was shown for linear operators K under additional assumptions.

2.4. The quadratic approximation. Another approach rests on constructing a quadratic approximation of $\Theta = F + \Phi$ at u^k and determining the next iterate as the minimizer of this quadratic approximation. This approach, including some clever step size selection criteria and several generalizations, has been studied in [35] in the finite-dimensional case. We will now formulate this approach in a Hilbert space setting.

In this approach, one chooses a $\lambda_k > 0$ and defines the quadratic approximation by

$$\Theta_\lambda(u, u^k) = F(u^k) + \langle F'(u^k), u - u^k\rangle_{\mathcal{H}} + \frac{\lambda_k}{2}\|u - u^k\|^2 + \Phi(u).$$

By completing squares we obtain

$$\Theta_\lambda(u, u^k) = c(u^k) + \frac{\lambda_k}{2}\|u - u^k + \frac{1}{\lambda_k}F'(u^k)\|^2 + \Phi(u) \tag{2.5}$$

with a constant $c(u^k)$ not depending on u. The minimizer u of this quadratic approximation is again obtained from the first order optimality condition, which states

$$0 \in \frac{1}{\lambda_k}F'(u^k) + (u - u^k) + \frac{1}{\lambda_k}\partial\Phi(u).$$

Choosing this minimizer as the next iterate yields, again by using the proximal mapping of Φ and (2.5), the following algorithm:

ALGORITHM 2.3 (Quadratic approximation).
Choose u^0 and iterate for $k > 0$
1. *determine a value λ_k, e.g., $\lambda_k = \lambda$ constant for all k*
2. *determine $z^k = u^k - \frac{1}{\lambda_k}F'(u^k)$*
3. *determine $u^{k+1} = P_{\frac{1}{\lambda_k}\Phi}(z^k) = \operatorname{argmin}_{u\in\mathcal{H}} \left(\frac{1}{2}\|u - u^k + \frac{1}{\lambda_k}F'(u^k)\|^2 + \frac{1}{\lambda_k}\Phi(u) \right)$.*

ETNA
Kent State University
http://etna.math.kent.edu

We directly see that this iteration coincides with the generalized gradient projection method if λ is identified with $\frac{1}{\lambda}$. We want to emphasize that the main achievement of [5, 35] is the introduction of a clever rule for choosing λ_k, which on the one hand guarantees a λ as small as possible (thus allowing for large gradient steps in Step 2 of the algorithm). On the other hand it is ensured that $\lambda_k \geq L$, where L is the Lipschitz constant of F', i.e., this ensures that the quadratic approximation always satisfies

$$\Theta_\lambda(u, u^k) \geq \Theta(u).$$

Also, several accelerated versions of this basic scheme are presented there, e.g., one variant constructs two sequences $\{u^k\}$ and $\{z^k\}$ which are related as follows

1. $u^k = z^k + t_k(z^k - z^{k-1})$ is a convex combination of z^k and z^{k-1}
2. $z^{k+1} = P_{\frac{1}{\lambda}\Phi}\left(u^k - \frac{1}{\lambda}F'(u^k)\right)$,

and z^k is shown to approximate the minimizer of the functional.

2.5. The generalized conditional gradient method. The starting point for motivating this iteration is a generalized version of a first order optimality condition; see [8]. This characterizes a minimizer u for $\Theta = F + \Phi$ by

$$\min_{z\in\mathcal{H}} \langle F'(u), z\rangle_{\mathcal{H}} + \Phi(z) = \langle F'(u), u\rangle_{\mathcal{H}} + \Phi(u).$$

In other words, if u^k is not a stationary point of Θ, then

$$\langle F'(u^k), u^k\rangle + \Phi(u^k) \;>\; \min_{z\in\mathcal{H}} \langle F'(u^k), z\rangle + \Phi(z).$$

This characterization motivates the following gradient method, which is called generalized conditional gradient method.

ALGORITHM 2.4 (Generalized conditional gradient method).
Choose u^0 with $F(u^0) + \Phi(u^0) < \infty$.
Compute $\{u^k | k > 0\}$ by

1. *determine* $z^k = \operatorname{argmin}_{z\in\mathcal{H}} \langle F'(u^k), z\rangle_{\mathcal{H}} + \Phi(z)$
2. *determine* $s_k = \operatorname{argmin}_{s\in[0,1]} \Theta(u^k + s(z^k - u^k))$ *or set* $s_k = \bar{s}$ *constant for all* k
3. $u^{k+1} = u^k + s_k(z^k - u^k)$.

Again, we can specify the above algorithm for the case defined in (1.1) and obtain a familiar expression by splitting the Tikhonov functional as follows

$$\Theta_\alpha(u) = \left(\frac{1}{2}\|K(u) - f^\delta\|^2 - \frac{\lambda}{2}\|u\|^2\right) + \left(\frac{\lambda}{2}\|u\|^2 + \alpha\Phi_p(u)\right).$$

In this case we have

$$F(u) = \frac{1}{2}\|K(u) - f^\delta\|^2 - \frac{\lambda}{2}\|u\|^2 \;\text{ and }\; \Phi(u) = \frac{\lambda}{2}\|u\|^2 + \alpha\Phi_p(u).$$

The minimizer in the first step of the algorithm can now be obtained by considering the first order optimality condition

$$K'(u^k)^*(K(u^k) - f^\delta) - \lambda u^k + (\lambda z + \alpha\partial\Phi_p(z)) = 0.$$

Hence, the minimizer z is given by

$$z = \mathbb{S}_{(\alpha/\lambda)\omega,p}\left(u^k - \frac{1}{\lambda}K'(u^k)^*(K(u^k) - f^\delta)\right)$$

and

$$u^{k+1} = u^k + s_k \left(\mathbb{S}_{(\alpha/\lambda)\omega,p} \left(u^k - \frac{1}{\lambda} K'(u^k)^*(K(u^k) - f^\delta) \right) - u^k \right),$$

which reduces to the iterated soft shrinkage algorithm as in (2.4) for $s_k = 1$.

The convergence properties of the generalized conditional gradient method applied to nonlinear operator equations was studied in detail in [6]. In particular, convergence for a fixed value of $s = 1$ was shown if λ is chosen large enough. However, the assumptions imposed in that paper are different from the ones we are using in the next section.

2.6. Surrogate functional approach. For motivating this approach, we start with the Tikhonov functional

$$\Theta_\alpha(u) = \frac{1}{2}\|K(u) - y^\delta\|^2 + \alpha\Phi_p(u).$$

The pioneering paper [12], which introduced sparsity constrained regularization techniques to the field of inverse problems, suggested to define a surrogate functional in order to decouple the analytic difficulties stemming from the operator and from the non-standard penalty term. This approach has be extended to nonlinear inverse problems by [38]. The main idea is to introduce

$$\Theta_\alpha^s(u, a) = \frac{1}{2}\|K(u) - f^\delta\|^2 + \frac{\lambda}{2}\|u - a\|^2 - \frac{1}{2}\|K(u) - K(a)\|^2 + \alpha\Phi_p(u).$$

This reduces to the original Tikhonov functional for $a = u$. The minimization of $\Theta_\alpha^s(x, a)$ with respect to u for a fixed a is assumed to be much easier, since—as can be seen after expanding the norms into scalar products—the quadratic term involving $K(u)$ cancels. The iteration based on this idea suggests the following algorithm:

ALGORITHM 2.5 (Surrogate functional).
Choose u^0 with $\frac{1}{2}\|K(u^0) - f^\delta\|^2 + \Phi_p(u^0) < \infty$ and λ sufficiently large.
For $k > 0$ determine

$$u^{k+1} = \operatorname*{argmin}_{u \in \mathcal{H}} \Theta_\alpha^s(u, u^k).$$

For linear operators, the minimization step can be performed explicitly and leads to a soft shrinkage iteration. For nonlinear operators, the minimizer cannot be computed explicitly in general, and the authors of [38] suggest to use a fixed point iteration based on the first order optimality condition of the surrogate functional. For fixed u^k, the first order optimality condition of $\Theta_\alpha^s(z, u^k)$ with respect to z reads as

$$0 \in K'(z)^*(K'(z) - f^\delta) + \lambda(z - u^k) - K'(z)^*(K(z) - K(u^k)) + \alpha\partial\Phi_p(z).$$

The term $K'(z)^*K(z)$ cancels and we obtain the more familiar expression after reordering

$$u^k - \lambda^{-1}K'(z)^*(K(u^k) - y^\delta) \in (I + \frac{\alpha}{\lambda}\partial\Phi_p)(z).$$

Hence, the inner iteration, where we need to find the fixed point defined by the minimization step in the algorithm, is a modified soft shrinkage iteration:

1. choose $z^0 = u^k$
2. iterate $z^{\ell+1} = \mathbb{S}_{(\alpha/\lambda)\omega,p}\left(u^k - \lambda^{-1}K'(z^\ell)^*(K(u^k) - f^\delta)\right)$ until convergence
3. put u^{k+1} equal to the last iterate of $\{z^\ell\}$.

ETNA
Kent State University
http://etna.math.kent.edu

The authors prove convergence of a subsequence to a stationary point for

$$\lambda \geq 2\max\left\{(\sup_{u\in M}\|K'(x)\|)^2, L\sqrt{\|K(u^0)-f^\delta\|^2+2\alpha\Phi_p(u^0)}\right\},$$

where $M=\{u\in\mathcal{H}:\Phi_p(u)\leq\Phi_p(u^0)\}$ and L is a Lipschitz constant for K'.

Let us note that the inner iterations also need an evaluation of K', hence their numerical costs is of the same order as an iteration step of the conditional gradient projection method. However, the condition on λ can be checked more easily. For a numerical comparison of these methods; see [6].

2.7. A comparison of the different gradient descent methods. As we have seen, all previous motivations for introducing an iteration method for minimizing Tikhonov functionals have been—up to different strategies for the selection of the step sizes λ and s—identical (except for the surrogate approach, which has some kind of implicit gradient step and hence has to use an additional inner fixed point iteration). They all reduce to a version of the iterated soft shrinkage algorithm when applied to functionals of type Θ_α with an ℓ_p-penalty term.

However, they have merits on their own. For instance, the approach via the generalized gradient projection method paves the way to incorporating additional constraints, i.e.,

$$\min_{u\in C}\frac{1}{2}\|K(u)-f^\delta\|^2+\alpha\Phi(u)=\min_{u\in\mathcal{H}}\Theta_\alpha(u)+tI_C(u).$$

For first steps in this direction; see [33].

The quadratic approximation method instead allows to analyze accelerated versions by considering convex combinations and step size selection criteria. Several approaches for linear operator equations have been analyzed so far; see, e.g., [5]. For a comparison of different minimization schemes for linear operator equations; see [34]. However, a thorough analysis of such accelerated versions of the iterated soft shrinkage algorithm for nonlinear operator equations is still missing.

Also, it is not surprising that the respective convergence analysis for these different algorithms use different analytic assumptions. In the following section we will extend the convergence results for the quadratic approximation method.

3. The quadratic approximation method for nonlinear operator equations in Hilbert spaces. The starting point for our investigation is a quadratic approximation method as proposed in [5, 35] for convex optimization problems (1.4) in $\mathbb{R}^n$. In this section, we analyze the convergence properties of this method in a general Hilbert space setting, moreover we discuss different step size selection criteria in the next section.

We examine the following general minimization problem

$$\min_{u\in\mathcal{H}}\left\{\Theta(u):=F(u)+\Phi(u)\right\}, \tag{3.1}$$

with $F:\mathcal{H}\to\mathbb{R}$ and $\Phi:\mathcal{H}\to\mathbb{R}$ under the following assumptions:

ASSUMPTION 1 (Assumptions on H and Φ).

1. *$\mathcal{H}$ is a Hilbert space.*
2. *$\Phi:\mathcal{H}\to\mathbb{R}$ is proper, convex, weakly lower semi-continuous, and weakly coercive.*

We assume that Assumption 1 holds throughout the paper. Most of the following analysis considers the general problem as stated in (3.1). However, we want to emphasize that we are especially interested in the case $\Phi=\alpha\Phi_p$ with Φ_p from (1.5). Accordingly, we will specialize and extend our general results to this particular choice of a penalty functional, e.g., in Lemma 3.2 and 3.8.

ETNA
Kent State University
http://etna.math.kent.edu

ASSUMPTION 2 (Assumptions on F and Θ).

1. *Problem (3.1) has at least one minimizer.*
2. *F is bounded from below. We may assume $F(u) \geq 0$, $\forall u \in \mathcal{H}$, without loss of generality.*
3. *F has a Lipschitz continuous Fréchet derivative, i.e., there exists a constant L such that*

$$\|F'(u) - F'(u')\| \leqslant L\|u - u'\|,\ \forall u, u' \in \mathcal{H}.$$

4. *If u^n converges weakly to u such that $\Theta(u^n)$ is monotonically decreasing, then there exists a subsequence $\{u^{n_j}\}$ such that*

$$F'(u^{n_j}) \to F'(u).$$

As discussed in the previous section, several methods have been proposed and investigated recently for minimizing functionals of this type (3.1) or more specifically for dealing with Tikhonov regularization for linear and nonlinear inverse problems such as (1.3); see, e.g., [8]. Each of these methods requires particular assumptions for proving its convergence.

REMARK 3.1. We discuss the role of the different parts of Assumption 2.

1. Condition 1 of Assumption 2 can be guaranteed if F is bounded below and weakly lower semi-continuous. Another sufficient condition for Condition 1 is given in [8, Lemma 3].
2. Condition 2 of Assumption 2 together with the weak coercivity of Φ implies the weak coercivity of $F + \Phi$, i.e., $F(u) + \Phi(u) \to \infty$ as $\|u\| \to \infty$. It is used to obtain the boundedness of the sequence generated by the gradient method; see Lemma 3.6. Note that this condition is weaker than the coercivity required in [8], i.e., $(F(u) + \Phi(u))/\|u\| \to \infty$ as $\|u\| \to \infty$.
3. Condition 3 of Assumption 2 is used to obtain Lemma 3.4 and the existence of step sizes in the gradient method and its accelerated versions; see Lemma 3.6. From this condition, we have

$$|F(v) - F(u) - \langle F'(u), v - u\rangle| \leqslant \frac{L}{2}\|v - u\|^2,\ \forall v, u \in \mathcal{H}.$$

4. Condition 4 of Assumption 2 is needed to obtain the strong convergence of the gradient method; see Theorem 3.10. It is satisfied if $E_t := \{u \in \mathcal{H} : \Phi(u) \leqslant t\}$ is compact for every $t \in \mathbb{R}$ and F' is continuous. Indeed, if u^n converges weakly to u and $\Theta(u^n)$ is monotonically decreasing, then $\{\Phi(u^n)\}_{n\in\mathbb{N}}$ is bounded and thus $\{u^n\} \subset E_t$ for some $t > 0$. Since E_t is compact, there is a subsequence $\{u^{n_j}\}$ such that $u^{n_j} \to u$. By continuity of F', we have $F'(u^{n_j}) \to F'(u)$.

3.1. The quadratic approximation methods in Hilbert spaces. As discussed in the previous section, the main idea of this gradient method is to replace the minimization problem (3.1) by a sequence of minimization problems, $\min_{v\in\mathcal{H}} \Theta_{s^n}(v, u^n)$, in which $\Theta_{s^n}(\cdot, u^n)$ are strictly convex and the minimization problems are easy to solve. Furthermore, the sequence of minimizers $u^{n+1} = \operatorname{argmin}_{v\in\mathcal{H}} \Theta_{s^n}(v, u^n)$ should converge to a minimizer of problem (3.1). For a fixed value of $s > 0$, we define the following quadratic approximation of $\Theta(v) = F(v) + \Phi(v)$ at a given point u,

$$\Theta_s(v, u) := F(u) + \langle F'(u), v - u\rangle + \frac{s}{2}\|v - u\|^2 + \Phi(v).$$

ETNA
Kent State University
http://etna.math.kent.edu

This functional admits a unique minimizer. The operator, which maps $u \in \mathcal{H}$ to the minimizer of $\Theta_s(\cdot, u)$ is denoted by $J_s : \mathcal{H} \to \mathcal{H}$. By completing the square we obtain a second characterization

$$\begin{aligned} J_s(u) &:= \operatorname*{argmin}_{v \in H} \{\Theta_s(v, u)\} \\ &= \operatorname*{argmin}_{v \in \mathcal{H}} \left\{ \frac{1}{2} \left\| v - \left(u - \frac{1}{s} F'(u)\right) \right\|^2 + \frac{1}{s}\Phi(v) \right\} = P_{\frac{1}{s}\Phi}(u - \frac{1}{s}F'(u)). \end{aligned} \tag{3.2}$$

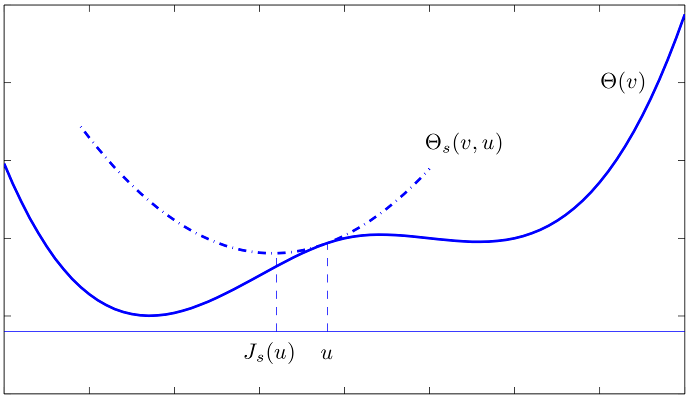

FIG. 3.1. *Sketch of the functionals* $\Theta(v)$, $\Theta_s(v, u)$ *and of the operator* $J_s(u)$.

The sequence of minimizers of these approximations is given by $u^{n+1} = J_s(u^n)$. Figure 3.1 provides a sketch of the functionals $\Theta(v)$, $\Theta_s(v, u)$ as well as $J_s(u)$. An explicit expression for the minimizer of Θ_s in the case of $\Phi = \alpha\Phi_p$ can be obtained by the soft shrinkage operator $\mathbb{S}_{\tau,p}$. The following lemma has been obtained in a similar setting in [6, 8, 21].

LEMMA 3.2. *Let F be Fréchet differentiable and let $\Phi = \alpha\Phi_p$ with Φ_p given in* (1.5).

1) The unique solution of (3.2) is given by

$$J_s(u) = \mathbb{S}_{\frac{\alpha\omega}{s},p}(u - \frac{1}{s}F'(u)).$$

2) If $u^* \in \mathcal{H}$ is a minimizer of Θ defined in (3.1), then the necessary condition for u^* is

$$u^* = \mathbb{S}_{\beta\alpha w,p}\big(u^* - \beta F'(u^*)\big) \textit{ for any fixed } \beta > 0.$$

Additionally, if F is convex, then this necessary condition is also sufficient.

We use this characterization of $J_s(u)$, which leads to the following gradient-type iteration for problem (3.1) with $\Phi = \alpha\Phi_p$

$$u^{n+1} = J_{s^n}(u^n) = \mathbb{S}_{\frac{\alpha\omega}{s^n},p}(u^n - \frac{1}{s^n}F'(u^n)). \tag{3.3}$$

The choice of the approximate step sizes $\frac{1}{s^n}$ affects the convergence properties of the iteration. This will be discussed in Section 4.

REMARK 3.3. We want to emphasize once more that this iteration coincides with several other gradient descent approaches for minimizing Θ. However, the proofs of convergence

ETNA
Kent State University
http://etna.math.kent.edu

use somewhat different assumptions and the quadratic approximation approach allows us to introduce different step size controls in the next section.

Next, we consider necessary conditions for s^n and then examine some convergence properties of this method.

3.2. Some convergence properties. In this section we follow the outline of [5, 35], where equivalent results but in finite-dimensional spaces were proved. The analytic techniques used for the proofs are similar to those of [5, 12].

For the analysis of the gradient method, we need the following result. This is based on the assumption that Θ_s is an approximation to Θ with stronger local convexity at u; see Figure 3.1.

LEMMA 3.4. *Assume that F is Fréchet differentiable with Lipschitz continuous derivative F'. Let $u \in \mathcal{H}$ and $s > 0$ be such that*

$$\Theta(J_s(u)) \leqslant \Theta_s(J_s(u), u). \tag{3.4}$$

Then for any $v \in \mathcal{H}$,

$$\Theta(v) - \Theta(J_s(u)) \geqslant \frac{s}{2}\|J_s(u) - u\|^2 + s\langle u - v, J_s(u) - u\rangle - \frac{L}{2}\|v - u\|^2,$$

where L is the Lipschitz constant of F'.

Proof. From (3.4), we have

$$\Theta(v) - \Theta(J_s(u)) \geqslant \Theta(v) - \Theta_s(J_s(u), u).$$

On the other hand, since $z = J_s(u)$ is the minimizer of $\Theta_s(., u)$, there exists a $\gamma \in \partial\Phi(z)$ such that

$$F'(u) + s(z - u) + \gamma = 0.$$

Now since F' is Lipschitz (see Remark 3.1) and Φ is convex, we have

$$\begin{aligned} F(v) &\geqslant F(u) + \langle F'(u), v - u\rangle - \frac{L}{2}\|v - u\|^2, \\ \Phi(v) &\geqslant \Phi(z) + \langle \gamma, v - z\rangle. \end{aligned} \tag{3.5}$$

Summing the above inequalities yields

$$\Theta(v) \geqslant F(u) + \langle F'(u), v - u\rangle + \Phi(z) + \langle \gamma, v - z\rangle - \frac{L}{2}\|v - u\|^2.$$

Furthermore, by definition of $z = J_s(u)$, one has

$$\Theta_s(z, u) = F(u) + \langle F'(u), z - u\rangle + \frac{s}{2}\|z - u\|^2 + \Phi(z).$$

From the previous inequality and equality, using $\gamma = -F'(u) - s(z - u)$, it follows that

$$\begin{aligned} \Theta(v) - \Theta(z) &\geqslant -\frac{s}{2}\|z - u\|^2 + \langle F'(u) + \gamma, v - z\rangle - \frac{L}{2}\|v - u\|^2 \\ &= -\frac{s}{2}\|z - u\|^2 + s\langle u - z, v - z\rangle - \frac{L}{2}\|v - u\|^2 \\ &= \frac{s}{2}\|z - u\|^2 + s\langle z - u, u - v\rangle - \frac{L}{2}\|v - u\|^2. \qquad \square \end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

REMARK 3.5.
1. By Remark 3.1, it is easy to show that (3.4) is satisfied if $s \geqslant L$.
2. Additionally, if F is convex, then $F(v) \geqslant F(u) + \langle F'(u), v-u\rangle$. Thus, following the proof above and inserting this stronger inequality into (3.5), we obtain

$$\Theta(v) - \Theta(J_s(u)) \geqslant \frac{s}{2}\|J_s(u)-u\|^2 + s\langle J_s(u)-u, u-v\rangle.$$

This inequality is exactly the one in [5, Lemma 2.3].

We are now in a position to investigate some convergence properties of the gradient method for the problem (3.1), i.e., the convergence properties of the sequence defined by (3.3).

LEMMA 3.6. *Let F satisfy Conditions 2, 3 of Assumption 2. Assume that the sequence $\{u^n\}$ is defined by (3.3), where the sequence of step sizes $\{s^n\}$ satisfies $s^n \in [\underline{s}, \overline{s}]$ with $(0 < \underline{s} \leq L \leq \overline{s})$ and*

$$\Theta(u^{n+1}) \leqslant \Theta_{s^n}(u^{n+1}, u^n).$$

Then the sequence $\Theta(u^n)$ is monotonically decreasing, $\lim_{n\to\infty}\|u^{n+1}-u^n\| = 0$, and the sequence $\{u^n\}$ is bounded.

Proof. The proof follows the idea of Beck and Teboulle [5]. By the hypothesis, we have

$$\Theta(u^{n+1}) \leqslant \Theta_{s^n}(u^{n+1}, u^n) \leqslant \Theta_{s^n}(u^n, u^n) = \Theta(u^n).$$

Thus, the sequence $\Theta(u^n)$ is monotonically decreasing as long as the hypothesis holds.

For each $k = 0, 1, \ldots, n$, applying Lemma 3.4 with $v = u = u^k$ and $s = s^k$, we obtain

$$\frac{2}{s^k}\big(\Theta(u^k) - \Theta(u^{k+1})\big) \geqslant \|u^k - u^{k+1}\|^2,$$
$$\frac{2}{\underline{s}}\big(\Theta(u^k) - \Theta(u^{k+1})\big) \geqslant \|u^k - u^{k+1}\|^2.$$

Summing the last inequality over $k = 0, \ldots, n$ gives

$$\frac{2}{\underline{s}}\big(\Theta(u^0) - \Theta(u^{n+1})\big) \geqslant \sum_{k=0}^{n}\|u^k - u^{k+1}\|^2, \quad \forall n.$$

This implies that the series $\sum_{k=0}^{\infty}\|u^k - u^{k+1}\|^2$ converges. As a consequence, we have

$$\lim_{n\to\infty}\|u^{n+1}-u^n\| = 0.$$

The boundedness of $\{u^n\}$ is a consequence of the decrease of $\{\Theta(u^n)\}$, the weak coercivity of Θ, i.e., $\Theta(u) \to \infty$ as $\|u\| \to \infty$, and Condition 2 of Assumption 2. ☐

The previous lemma implies that the sequence $\{u^n\}$ is bounded. Hence, it must have a weak accumulation point. We now aim at proving that each weak accumulation point is a stationary point of Θ, i.e., it satisfies the necessary condition for a minimizer of Θ. To this end, we only consider the case $\Phi = \alpha\Phi_p$.

First, we need the following technical lemma.

LEMMA 3.7. *Assume that*

$$u^n = \mathbb{S}_{\beta^n\alpha w,p}\big(v^n - \beta^n F'(v^n)\big).$$

If both u^n and v^n converge weakly to u^*, $F'(v^n)$ converges weakly to $F'(u^*)$, and $\beta_n > 0$, with $\lim_{n\to\infty}\beta^n = \beta^* > 0$, then

$$u^* = \mathbb{S}_{\beta^*\alpha w,p}\big(u^* - \beta^* F'(u^*)\big).$$

Proof. We first prove the lemma for $p > 1$. Using the notation $u_k = \langle u, \varphi_k \rangle$, we have that u_k^n and v_k^n converge to u_k^* and $F'(u^n)_k$ converges to $F'(u^*)_k$ for all $k \in \Lambda$ when $n \to \infty$. By assumption it holds that

$$u^n = \mathbb{S}_{\beta^n \alpha w, p}\big(v^n - \beta^n F'(v^n)\big),$$

which is equivalent to

$$u_k^n = S_{\beta^n \alpha w_k, p}\big(v_k^n - \beta^n F'(v^n)_k\big), \quad \forall k \in \Lambda.$$

By (2.1) and (2.2), these equations are equivalent to

$$u_k^n + p\beta^n \alpha \omega_k \operatorname{sgn}(u_k^n)|u_k^n|^{p-1} = v_k^n - \beta^n F'(v^n)_k, \quad \forall k \in \Lambda.$$

Taking $n \to \infty$, we get

$$u_k^* + p\beta^* \alpha \omega_k \operatorname{sgn}(u_k^*)|u_k^*|^{p-1} = u_k^* - \beta^* F'(u^*)_k, \quad \forall k \in \Lambda.$$

Therefore we have

$$u^* = \mathbb{S}_{\beta^* \alpha w, p}\big(u^* - \beta^* F'(u^*)\big).$$

We now prove the lemma for $p = 1$. By the hypothesis we have that

$$u^n = \mathbb{S}_{\beta^n \alpha w, 1}\big(v^n - \beta^n F'(v^n)\big),$$

which is equivalent to

$$u_k^n = \operatorname{sgn}(v_k^n - \beta^n F'(v^n)_k) \max\big(|v_k^n - \beta^n F'(v^n)_k| - \beta^n \alpha w_k, 0\big), \quad \forall k \in \Lambda. \tag{3.6}$$

We denote

$$\begin{aligned}
\Gamma_1 &:= \{k \in \Lambda : |u_k^* - \beta^* F'(u^*)_k| > \beta^* \alpha w_k\},\\
\Gamma_2 &:= \{k \in \Lambda : |u_k^* - \beta^* F'(u^*)_k| < \beta^* \alpha w_k\},\\
\Gamma_3 &:= \{k \in \Lambda : |u_k^* - \beta^* F'(u^*)_k| = \beta^* \alpha w_k\}.
\end{aligned}$$

We treat each of these three cases separately. Since $v_k^n - \beta^n F'(v^n)_k \to u_k^* - \beta^* F'(u^*)_k$ and $|v_k^n - \beta^n F'(v^n)_k| - \beta^n \alpha w_k \to |u_k^* - \beta^* F'(u^*)_k| - \beta^* \alpha w_k$ as $n \to \infty$ (with k being fixed), we obtain the following:

- If $k \in \Gamma_1$, then $v_k^n - \beta^n F'(v^n)_k$ and $u_k^* - \beta^* F'(u^*)_k$ have the same sign and $|v_k^n - \beta^n F'(v^n)_k| - \beta^n \alpha w_k > 0$ when n is large enough, and thus the limit of two sides of (3.6) exists and

  $$u_k^* = \operatorname{sgn}(u_k^* - \beta^* F'(u^*)_k) \max\big(|u_k^* - \beta^* F'(u^*)_k| - \beta^* \alpha w_k, 0\big), \quad \forall k \in \Gamma_1,$$

  or

  $$u_k^* = S_{\beta^* \alpha w, 1}\big(u_k^* - \beta^* F'(u^*)_k\big), \quad \forall k \in \Gamma_1.$$

- If $k \in \Gamma_2$, then $|v_k^n - \beta^n F'(v^n)_k| - \beta^n \alpha w_k < 0$ when n is large enough. Thus, (3.6) becomes $u_k^n = 0$. It follows that $u_k^* = 0$ and then

  $$u_k^* = S_{\beta^* \alpha w, 1}\big(u_k^* - \beta^* F'(u^*)_k\big), \quad \forall k \in \Gamma_2.$$

- If $k \in \Gamma_3$, then $v_k^n - \beta^n F'(v^n)_k$ and $u_k^* - \beta^* F'(u^*)_k$ have the same sign and are nonzero when n is large enough. Thus, $\frac{u_k^n}{\operatorname{sgn}(v_k^n - \beta^n F'(v^n)_k)} \to \frac{u_k^*}{\operatorname{sgn}(u_k^* - \beta^* F'(u^*)_k)}$ as $n \to \infty$. From (3.6), we deduce that $\max\left(|v_k^n - \beta^n F'(v^n)_k| - \beta^n \alpha w_k, 0\right)$ also converges and its limit is equal to zero. This implies that $u_k^* = 0$ and thus

$$u_k^* = S_{\beta^* \alpha w,1}\left(u_k^* - \beta^* F'(u^*)_k\right), \quad \forall k \in \Gamma_3.$$

Summarizing the above results, we have that

$$u_k^* = S_{\beta^* \alpha w,1}\left(u_k^* - \beta^* F'(u^*)_k\right), \quad \forall k \in \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 = \Lambda,$$

and

$$u^* = \mathbb{S}_{\beta^* \alpha w,p}\left(u^* - \beta^* F'(u^*)\right). \qquad \square$$

LEMMA 3.8. *Let F satisfy Assumption 2, $\Phi = \alpha\Phi_p$, and $\{u^n\}$ be defined in Lemma 3.6. If u^* is a weak accumulation point of $\{u^n\}$, then u^* is a stationary point of Θ.*

Proof. Let $\{u^{n_j}\}_{j\in\mathbb{N}}$ be a subsequence converging weakly to u^*. By $s^n \in [\underline{s}, \overline{s}]$ and Assumption 2, there exists a subsequence of this subsequence (again denoted by $\{u^{n_j}\}$) such that $\text{w-}\lim_{j\to\infty} u^{n_j} = u^*$, $F'(u^{n_j}) \to F'(u^*)$, and $\lim_{j\to\infty} s^{n_j} = s^* \in [\underline{s}, \overline{s}]$. Due to Lemma 3.6, $\{u^{n_j+1}\}$ also converges weakly to u^*. By (3.3), we have

$$u^{n_j+1} = \mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\left(u^{n_j} - \frac{1}{s^{n_j}} F'(u^{n_j})\right).$$

By Lemma 3.7, we obtain

$$u^* = \mathbb{S}_{\frac{\alpha\omega}{s^*},p}\left(u^* - \frac{1}{s^*} F'(u^*)\right).$$

By Lemma 3.2, u^* is a stationary point of Θ. $\square$

Next, we shall prove that the sequence $\{u^n\}_{n\in\mathbb{N}}$ has a strongly convergent subsequence. To this end, we need the following generalization of the result in [12, Lemma 3.18].

LEMMA 3.9. *Let $\{h^n\} \subset \mathcal{H}$ be uniformly bounded and $\{d^n\} \subset \mathcal{H}$ converge weakly to zero. If $s^n \in [\underline{s}, \overline{s}]$ and $\lim_{n\to\infty} \|\mathbb{S}_{\frac{\alpha w}{s^n},p}(h^n + d^n) - \mathbb{S}_{\frac{\alpha w}{s^n},p}(h^n) - d^n\| = 0$, then $\|d^n\| \to 0$ for $n \to \infty$.*

Proof. This lemma can be proven similar to [12, Lemma 3.18]. $\square$

THEOREM 3.10. *Let F satisfy Assumption 2, $\Phi = \alpha\Phi_p$, and let $\{u^n\}$ be defined as in Lemma 3.6. Then the sequence $\{u^n\}$ has a subsequence that converges strongly to a stationary point u^* of Θ.*

Proof. Let $\{u^{n_j}\}_{j\in\mathbb{N}}$ be the subsequence of $\{u^n\}$ defined in the proof of Lemma 3.8. Hence, u^* is a stationary point of Θ, and by Lemma 3.2 we have

$$u^* = \mathbb{S}_{\alpha\omega\beta,p}\left(u^* - \beta F'(u^*)\right)$$

for any fixed $\beta > 0$. We denote $d^{n_j} = u^{n_j} - u^*$ and $h^{n_j} = u^* - \frac{1}{s^{n_j}} F'(u^*)$. Due to Lemma 3.6, we have that $\lim_{j\to\infty} \|d^{n_j+1} - d^{n_j}\| = 0$. Using the previous equation for u^*

ETNA
Kent State University
http://etna.math.kent.edu

with $\beta = \frac{1}{s^{n_j}}$, we get

$$\begin{aligned} d^{n_j} - d^{n_j+1} &= d^{n_j} + u^* - \mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\big(u^{n_j} - \frac{1}{s^{n_j}}F'(u^{n_j})\big) \\ &= d^{n_j} + \mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\big(u^* - \frac{1}{s^{n_j}}F'(u^*)\big) - \mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\big(u^{n_j} - \frac{1}{s^{n_j}}F'(u^{n_j})\big) \\ &= d^{n_j} + \mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\big(h^{n_j}\big) \\ &\qquad - \mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\big(u^* - \frac{1}{s^{n_j}}F'(u^{n_j}) + d^{n_j}\big) \qquad (3.7) \\ &\qquad + \mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\big(u^* - \frac{1}{s^{n_j}}F'(u^*) + d^{n_j}\big) \qquad (3.8) \\ &\qquad - \mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\big(u^* - \frac{1}{s^{n_j}}F'(u^*) + d^{n_j}\big). \end{aligned}$$

We consider now the sum of (3.7) and (3.8). By Assumption 2, the nonexpansiveness of $\mathbb{S}$ (see, for example [12]) and $s^{n_j} \to s^*$, we have

$$\begin{aligned} &\|\mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\big(u^* - \frac{1}{s^{n_j}}F'(u^{n_j}) + d^{n_j}\big) - \mathbb{S}_{\frac{\alpha\omega}{s^{n_j}},p}\big(u^* - \frac{1}{s^{n_j}}F'(u^*) + d^{n_j}\big)\| \\ &\qquad\qquad \leqslant \frac{1}{s^{n_j}}\|F'(u^{n_j}) - F'(u^*)\| \to 0 \quad (j \to \infty). \end{aligned}$$

Consequently, combining $\|d^{n_j} - d^{n_j+1}\| \to 0$ as $j \to \infty$ and the last inequality, we observe that

$$\lim_{j\to\infty} \|\mathbb{S}_{\frac{\alpha w}{s^{n_j}},p}(h^{n_j} + d^{n_j}) - \mathbb{S}_{\frac{\alpha w}{s^{n_j}},p}(h^{n_j}) - d^{n_j}\| = 0.$$

Applying Lemma 3.9 where the sequences $\{h^n, d^n\}$ are replaced by $\{h^{n_j}, d^{n_j}\}$, we obtain the desired result. ∎

REMARK 3.11.

- A similar result as in Theorem 3.10 has been obtained in [6, 8] for constant step-sizes ($1/s^n = s$) under different assumptions on F and Φ; see [8, Theorem 1].
- For finite-dimensional spaces $\mathcal{H}$, the above results have been obtained implicitly in [35, Theorem 5] under the strong convexity condition for Θ. In that case, even a linear convergence rate of $\{u^n\}$ can be proved.
- A linear convergence rate of $\{u^n\}$ has also been obtained in [7] under the following conditions: $\Theta = F + \Phi$ is coercive, F is convex, and the sequence $\{u^n\}$ satisfies $\|u^n - u^*\| \leqslant cr^n$, where u^* is a minimizer of Θ and $r^n := \Theta(u^n) - \Theta(u^*)$.

In our setting, we do not impose the condition $\|u^n - u^*\| \leqslant cr^n$ for proving convergence rates for $\{u^n\}$ in this paper. Instead, we are aiming at weaker results concerning the decay rate of the functional values $\Theta(u^k)$.

THEOREM 3.12. *Let F be convex and satisfy the Conditions 1–3 of Assumption 2, and let $\{u^n\}$ be defined as in Lemma 3.6. Then for any $n \geqslant 1$*

$$\Theta(u^n) - \Theta(u^*) \leqslant \frac{\overline{s}\|u^0 - u^*\|^2}{2n},$$

where u^* is a minimizer of Θ.

Proof. Since F is convex, we obtain the same inequality as in [5, Lemma 2.3] by Remark 3.5. Thus, the proof is obtained as in [5, Lemma 3.1]. ∎

ETNA
Kent State University
http://etna.math.kent.edu

4. A step size selection criterion. As analyzed in the previous section, the quadratic approximation method converges when the parameters s^n satisfy the conditions stated in Lemma 3.6. We note that Remark 3.1 implies that $s \geqslant L$ yields

$$|F(v) - F(u) - \langle F'(u), v - u\rangle| \leqslant \frac{s}{2}\|v - u\|^2.$$

Hence, with $s \geqslant L$ we obtain

$$\Theta(v) = F(v) + \Phi(v) \leqslant F(u) + \langle F'(u), v - u\rangle + \frac{s}{2}\|v - u\|^2 + \Phi(v) = \Theta_s(v, u),$$

and thus the conditions in Lemma 3.6 are always satisfied if $s^n \geqslant L$ for all n.

It is well known that the choice of step sizes s^n affects the convergence of the gradient method; see, for example, [6]. Some strategies for choosing these parameters in the context of quadratic approximations in finite-dimensional spaces were proposed in [5, 35]. However, we follow a different approach. Let us have a closer look at the iteration (3.3). It is easy to see that — neglecting the soft shrinkage operator $\mathbb{S}$ — the parameters $\frac{1}{s^n}$ are the step sizes of the classical gradient method for the minimization problem $\min_{u\in\mathcal{H}} F(u)$. Therefore, we suggest to first compute an intermediate step size t^n by

$$t^n := \operatorname*{argmin}_{t>0} F(u^n - tF'(u^n)). \tag{4.1}$$

Imposing a lower and upper bound on the step size s^n then yields a first guess for the step size

$$\frac{1}{s^n} = P_{[\overline{s}^{-1},\underline{s}^{-1}]}(t^n) := \max(\min(t^n, \underline{s}^{-1}), \overline{s}^{-1}).$$

We then check whether the condition in Lemma 3.6, i.e., $\Theta(u^{n+1}) \leqslant \Theta_{s^n}(u^{n+1}, u^n)$, is satisfied. We retain s^n if the condition is satisfied, otherwise we repeatedly reduce $1/s^n$ by a factor $q < 1$. Note that the problem (4.1) does not need to be solved exactly. We only need an efficient strategy for approximating this minimizer. For this purpose, we use the Barzilai-Borwein rule proposed in [4]

$$\frac{1}{s^n} = P_{[\overline{s}^{-1},\underline{s}^{-1}]}\Big(\frac{\langle u^n - u^{n-1}, F'(u^n) - F'(u^{n-1})\rangle}{\langle F'(u^n) - F'(u^{n-1}), F'(u^n) - F'(u^{n-1})\rangle}\Big). \tag{4.2}$$

By this strategy, we summarize the quadratic approximation method with step size control in the following algorithm:

Algorithm 1

Initiation:	Initial guess u^0 such that $\Theta(u^0) < \infty$, $s^0 \in [\underline{s}, \overline{s}]$ $(0 < \underline{s} \leq L/q \leq \overline{s})$, and $q < 1$.
Iteration:	**for** $n = 0, 1, 2, \ldots$ **1.** $u^{n+1} = J_{s^n}(u^n)$. **2.** If $\Theta(u^{n+1}) > \Theta_{s^n}(u^{n+1}, u^n)$ and $s^n \in [\underline{s}, \overline{s}]$ then $\frac{1}{s^n} = \frac{1}{s^n}q$; go to **Step 1.** **3.** $\frac{1}{s^{n+1}}$ given by (4.2). **end**
Output:	the output of the algorithm is $u = u_{lim}$.

REMARK 4.1.

1. Together with Remark 3.1, the assumption in the initialization, $\overline{s} \geqslant L/q$, guarantees that $\Theta(u^{n+1}) \leq \Theta_{s^n}(u^{n+1}, u^n)$ is always satisfied after a finite number of updates $\frac{1}{s^n} = \frac{1}{s^n}q$. Hence, we do not need the condition $s^n \in [\underline{s}, \overline{s}]$ in the algorithm. It is included in case that L is not known. This remark is also relevant for the next accelerated versions.
2. If $F : \mathbb{R}^d \to \mathbb{R}$ is convex then Algorithm 1 for (3.1) is similar to the gradient method in [35] and to ISTA in [5]. The only difference is the criterion for choosing s^n.
3. For a fixed step size $s^n = s$, the proposed algorithm is also identical to the generalized conditional gradient method with $\lambda = 1$; see Remark 3.3.

5. Some accelerated versions. In this section, we aim at presenting two accelerated schemes of the quadratic approximation method, which, however, require stronger assumptions. Hence, in this section we assume that F in (3.1) is convex. This assumption has also been analyzed in [5, 35] for finite-dimensional spaces. There the authors proposed two accelerated versions and proved a convergence rate for the values of the objective functional of order $O(\frac{1}{n^2})$. This convergence rate is known to be optimal for algorithms that are based on first order schemes, i.e., for algorithms using only the values of the objective functional Θ and its gradient [2, 5, 35, 40]. Similarly, we present accelerated versions for the problem (3.1) in a general Hilbert space setting.

The first accelerated algorithm of the gradient method for the problem (3.1) is motivated by [5] and is presented in Algorithm 2.

Algorithm 2

Initiation:	Initial guess $y^0 \in \mathrm{dom}(\Phi)$, $s^0 \in [\underline{s}, \overline{s}]$ $(0 < \underline{s} \leq L/q \leq \overline{s})$, and $t_0 = 1, q < 1$.
Iteration:	**for** $n = 0, 1, 2, \ldots$ **1.** $u^n = J_{s^n}(y^n)$ **2.** If $\Theta(u^n) > \Theta_{s^n}(u^n, y^n)$ and $s^n \in [\underline{s}, \overline{s}]$ then $\frac{1}{s^n} = \frac{1}{s^n}.q$; go to **Step 1.** **3.** $t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}$ **4.** $y_{n+1} = u^n + \left(\frac{t_n - 1}{t_{n+1}}\right)(u^n - u^{n-1})$ **5.** $s^{n+1} = s^n$ **end**
Output:	the output of the algorithm is $u = u_{lim}$.

The convergence rate for the values of the objective functional Θ for this algorithm is given in the following theorem.

THEOREM 5.1. *Let F be convex and satisfy the Conditions 1–3 of Assumption 2. Let $\{u^n\}$ be generated by Algorithm 2 and u^* be a minimizer of problem (3.1). Then for any $n \geqslant 1$*

$$\Theta(u^n) - \Theta(u^*) \leqslant \frac{C\|u^0 - u^*\|^2}{(n+1)^2} \quad (C = 2L/q).$$

Proof. By Remark 3.5, we obtain the same inequality as in [5, Lemma 2.3] and we proceed as in the proof of [5, Theorem 4.3]. □

The second accelerated algorithm presented in Algorithm 3 is motivated by [35]. Note that in [35], the author proposed this algorithm for a general functional Φ in finite-dimensional spaces. Here we apply it for a specific functional $\Phi = \alpha\Phi_p$, but we extend it to the Hilbert space setting. In the context of problem (3.1) with $\Phi = \alpha\Phi_p$, the solution v^n in Step 6 of Algorithm 3 is given explicitly in the following lemma.

LEMMA 5.2. *Let $\psi_n(u)$ be as in Step 6 of Algorithm 3. Then, $v^n = \operatorname{argmin}_{u\in H} \psi_n(u)$ is given by*

$$v^n = \mathbb{S}_{\alpha A_n,p}\big(u^0 - \sum_{k=1}^{n} a_k F'(u^k)\big)\ (n > 0).$$

Proof. The initial quadratic approximation is defined as $\psi_0(u) = \frac{1}{2}\|u - u^0\|^2$. The following iterates satisfy

$$\psi_n(u) = \frac{1}{2}\|u - u^0\|^2 + \sum_{k=1}^{n} a_k\Phi(u) + \sum_{k=1}^{n} a_k\big(F(u^k) + \langle F'(u^k), u - u^k\rangle\big).$$

Now, the proof is similar to that of Lemma 3.2. □

REMARK 5.3. From the formula of v^n, it seems that Algorithm 3 is a method with "infinite memory". In fact, it is not since we can define a variable z^n with the initial value $z^0 := u^0$ and Step 6 of Algorithm 3 is replaced by

$$z^{n+1} := z^n - a_{n+1}F'(u^{n+1}),\ v^{n+1} = \mathbb{S}_{\alpha A_{n+1},p}\big(z^{n+1}\big)\ .$$

Algorithm 3

Initialization:	Initial guess $u^0 \in \operatorname{dom}(\Phi)$, $A_0 = 0$, $v^0 = u^0$, $s^0 \in [\underline{s}, \overline{s}]$ $(0 < \underline{s} \le L/q \le \overline{s})$, and $\psi_0(u) = \frac{1}{2}\|u - u^0\|^2$.
Iteration:	**for** $n = 0, 1, 2, \ldots$ **1.** $a_{n+1} = \frac{1+\sqrt{1+2A_n s^n}}{s^n}$ **2.** $y^n = \frac{A_n u^n + a_{n+1}v^n}{A_n + a_{n+1}}$ **3.** $u^{n+1} = J_{s^n}(y^n)$ **4.** If $\|F'(u^{n+1}) - F'(y^n)\|^2 > s^n\langle F'(y^n) - F'(u^{n+1}), y^n - u^{n+1}\rangle$ and $s^n \in [\underline{s}, \overline{s}]$ then $\frac{1}{s^n} = \frac{1}{s^n}.q$; go to **Step 1** **5.** $\frac{1}{s^{n+1}} = P_{[\overline{s}^{-1}, \underline{s}^{-1}]}\frac{\langle u^{n+1} - y^n, F'(u^{n+1}) - F'(y^n)\rangle}{\langle F'(u^{n+1}) - F'(y^n), F'(u^{n+1}) - F'(y^n)\rangle}$ $A_{n+1} = A_n + a_{n+1}$ **6.** $v^{n+1} = \operatorname{argmin}_{u\in\mathcal{H}} \psi_{n+1}(u)$ with $\psi_{n+1}(u) = \psi_n(u)$ $+a_{n+1}(F(u^{n+1}) + \langle F'(u^{n+1}), u - u^{n+1}\rangle + \Phi(u))$ **end**
Output:	the output of the algorithm is $u = u_{lim}$.

Finally, the convergence rate for the values of the objective functional Θ in Algorithm 3 is obtained similarly as in [35, Theorem 6].

THEOREM 5.4. *Let F be convex and satisfy the Conditions 1–3 of Assumption 2. Let $\{u^n\}$ be generated by Algorithm 3, and u^* be a minimizer of problem (3.1). Then for any $n \geqslant 1$*

$$\Theta(u^n) - \Theta(u^*) \leqslant \frac{C\|u^0 - u^*\|^2}{n^2} \quad (C = L/q).$$

Proof. Note that if s^n satisfies the condition in Step 4 of Algorithm 3, then it also satisfies the condition (**) of the accelerated method in [35]; see [35, Lemma 4]. Thus, the theorem's proof is done similarly as that in [35, Theorem 6]. □

ETNA
Kent State University
http://etna.math.kent.edu

6. Numerical examples. In this section, we implement the algorithms described above for a parameter identification problem for an elliptic partial differential equation. We will use the notation which is customary in this field, i.e., the quantity searched for is now denoted by a and u denotes the solution of the elliptic equation, which is also the data for this problem. To be precise, we aim at estimating the coefficient a in the elliptic boundary problem

$$\begin{aligned} -\operatorname{div}(a\nabla u) &= y \text{ in } \Omega \subset \mathbb{R}^2, \\ u &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{6.1}$$

with a fixed $y \in L^2(\Omega)$. Let a^* denote the parameter to be recovered and let u^* denote the solution of (6.1) with parameter a^* and right-hand side y. As data we assume that u^δ is given, where u^δ is the solution of the elliptic equation with parameter a^* but perturbed right-hand side y^δ with $\|y - y^\delta\|_{L_2} \leq \delta$. Hence, the available data satisfy $\|u^* - u^\delta\|_{H^1(\Omega)} \leq C\delta$ with a certain positive constant. Our task is to determine an approximation of a^* from u^δ.

A number of papers, such as [1, 9, 10, 18, 23, 26, 29, 30, 31, 39, 41, 42], have examined this problem or variations of it; see also [3].

We let

$$\mathcal{A} = \{a \in L^\infty(\Omega) : 0 < \underline{a} \leqslant a \leqslant \overline{a},\ \operatorname{supp}(a - a^0) \subset\subset \Omega\}$$

and define $K : \mathcal{A} \subset L^\infty(\Omega) \to H_0^1(\Omega), a \mapsto u$, the solution operator of (6.1) with fixed y. The parameter identification problem regularized by sparsity constraints leads us to the following constrained minimization problem

$$\min_{a\in\mathcal{A}} \Theta(a) = \int_\Omega a|\nabla K(a) - \nabla u^\delta|^2 dx + \alpha\Phi_p(a - a^0).$$

We set $\Theta(a) = +\infty$ if $a \notin \mathcal{A} \cap \operatorname{dom}(\Phi)$. Then this problem is equivalent to

$$\min_{a\in L^2(\Omega)} \Theta(a) = \int_\Omega a|\nabla K(a) - \nabla u^\delta|^2 dx + \alpha\Phi_p(a - a^0).$$

It is known that $F(a) = \int_\Omega a|\nabla K(a) - \nabla u^\delta|^2 dx$ is convex and Lipschitz differentiable with respect to the L^∞-norm [23], but the Lipschitz differentiability of it with respect to the L^2-norm is, to the best of our knowledge, unknown. However, we will see that the algorithms work well for this problem. This fact also concerns Remark 4.1.

$$F'(a)h = -\int_\Omega h(|\nabla K(a)|^2 - |\nabla u^\delta|^2)dx.$$

For illustrating our algorithms, we assume that Ω is the unit disk and

$$a^*(x_1, x_2) = \begin{cases} 4 & (x_1, x_2) \in B_{0.4}(0, 0.3) \\ 1 & \text{otherwise} \end{cases}, \qquad y(x_1, x_2) = 4a^*,$$

where $B_r(x_1, x_2)$ is the disk with center at (x_1, x_2) and radius r. Here, we take $a^0 = 1$ and thus $a^* \in \mathcal{A}$ since $\operatorname{supp}(a^* - a^0) \subset\subset \Omega$.

We discretized the problems (forward and inverse) by the finite element method [42] and set

$$\Phi_p(\vartheta) = \sum_{k=1}^N \left|\langle \vartheta, \varphi_k\rangle_{L^2(\Omega)}\right|^p,$$

ETNA
Kent State University
http://etna.math.kent.edu

where $\{\varphi_k\}_{k=1,\dots,N}$ is the basis consisting of piecewise linear finite elements in the discretized space. Since we assume that $\operatorname{supp}(a^* - a^0) \subset\subset \Omega$ is small, many coefficients of the unknown parameter $\langle a^* - a^0, \varphi_k \rangle_{L^2(\Omega)}$ are equal to zero, i.e., $a^* - a^0$ has a sparse expansion in the basis of the finite element method.

To obtain u^* and u^δ, we solve (6.1) by the finite element method on a mesh with 1272 triangles. The solution of (6.1) as well as the parameter a are represented by piecewise linear finite elements. The algorithms described in the previous section will compute a sequence a^n for approximating a^*. In order to maintain the ellipticity of the operator, we add as usual an additional truncation step in the numerical procedure, which, however, is not covered by our theoretical investigation, i.e., we cut off values of a^n which are below $a^0 = 1$ in each iteration.

In the remainder of this section we describe the following experiments: at first, we compare the effect of our choice of parameters in the gradient method with other choices proposed in [5, 35], i.e., we compare

- Algorithm 1: method of quadratic approximations with step sizes chosen according to Algorithm 1 and $q = 0.5$, $[\underline{s}, \overline{s}] := [10^{-2}, 10^2]$, $\alpha := 5 \cdot 10^{-5}$.
- Algorithm 1N: gradient method of [35] with $\gamma_u = \gamma_d = 2$.
- Algorithm 1B: gradient method of [5] with $\eta = 2$, i.e., this is ISTA with backtracking.

Secondly, we compare the gradient method with its accelerated versions Algorithm 2 and Algorithm 3. For Algorithm 1, Algorithm 2, and Algorithm 3 we set $q = 0.5$, $\alpha := 5 \cdot 10^{-5}$, and $[\underline{s}, \overline{s}] := [10^{-2}, 10^2]$. We measure the convergence of the computed minimizers to the true parameter a^* by considering the mean square error sequence

$$MSE(a^n) = \int_\Omega (a^n - a^*)^2 dx.$$

6.1. Numerical experiments with $\delta = 0$. We first discuss numerical results without noise, i.e., $u^\delta = u^*$. Figure 6.1 displays the resulting step sizes $\frac{1}{s^n}$ in Algorithm 1, Algorithm 1B, and Algorithm 1N for the first 300 iterations. Large but controlled step sizes are preferable for fast convergence. Towards the end of the iteration, the step sizes in Algorithm 1 are typically larger than those of the others. The step sizes in Algorithm 1B are the smallest. Furthermore, we observe that the initial guesses for the step sizes in Algorithm 1 and Algorithm 1B always satisfy the conditions in Lemma 3.6, but Algorithm 1N needs some iterations to establish them. Therefore, Algorithm 1N takes more time than Algorithm 1 and Algorithm 1B; see Figure 6.2. Figure 6.1 shows that $\{MSE(a^n)\}$ in Algorithm 1B decreases most slowly. In the first iterations, $\{MSE(a^n)\}$ in Algorithm 1N decreases faster than that in Algorithm 1, but after that it decreases more slowly.

The decrease of the objective functionals is illustrated in Figure 6.2. In this example, Algorithm 1 exhibits the best convergence rate.

Figure 6.3 displays a^* and a^n with $n = 300$ for all three algorithms. It shows that the algorithms recover a^* very well and Algorithm 1 gives the best approximation of a^* with a fixed number of iterations.

From this analysis, we conclude that the step sizes computed by (4.2) are preferable; they are typically larger than those chosen by the other algorithms and they always satisfy the conditions in Lemma 3.6.

Now we compare Algorithm 1 with its accelerated versions (Algorithm 2 and Algorithm 3). Figure 6.4 displays the values of the step sizes $\frac{1}{s^n}$ in Algorithm 1, Algorithm 2, and Algorithm 3. We observe that the initial guesses for the step sizes in Algorithm 1 and Algorithm 2 always satisfy the conditions in Step 2, but the initial guess in Algorithm 3 often needs

ETNA
Kent State University
http://etna.math.kent.edu

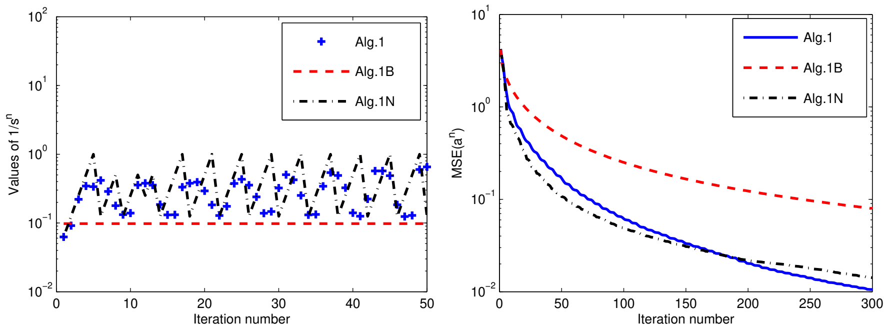

FIG. 6.1. *The values of* $\frac{1}{s^n}$ *and* $MSE(a^n)$ *in Algorithm 1, Algorithm 1B, and Algorithm 1N.*

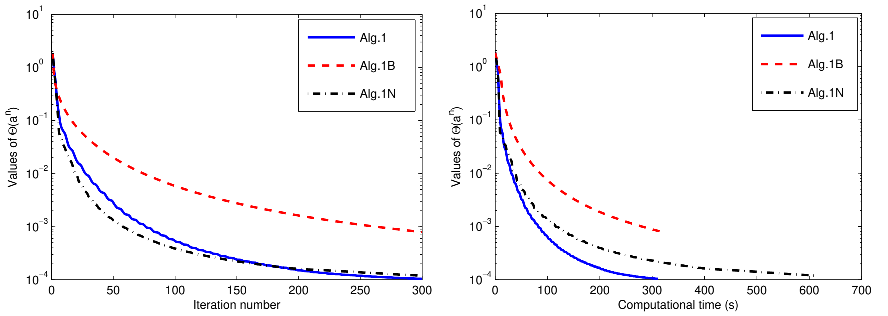

FIG. 6.2. *The decrease rate of* $\Theta(a^n)$ *in Algorithm 1, Algorithm 1B, and Algorithm 1N.*

one refinement iteration in order to satisfy the condition in Step 4 of Algorithm 3. Moreover, we observe that the convergence order is $O(1/n^2)$ for Algorithm 2 and Algorithm 3 as predicted by the analytic results of the previous section. Hence, these algorithms converge faster than Algorithm 1. This is confirmed in Figure 6.5.

The convergence rate of the objective functional with respect to the number of iterations and the computational time is illustrated in Figure 6.5. During the first iterations, the convergence rate of Algorithm 3 is faster than that of Algorithm 2. The convergence rate is the slowest for Algorithm 1. This agrees with the theory; see Theorem 3.12, Theorem 5.1, and Theorem 5.4. However, this is slightly misleading since each iteration of the accelerated algorithms needs more time than the original algorithm. Overall, the convergence of the functional values with respect to the computational time is equivalent for all versions.

Figure 6.6 illustrates a^* and a^n with $n = 300$ computed by all three algorithms. The accelerated algorithms reconstruct the parameter a^* better than Algorithm 1. The reconstructions of a^* in Algorithm 2 and Algorithm 3 are almost exact.

6.2. Numerical experiments with noisy data. This section deals with noisy data. To obtain $u^\delta \in H_0^1(\Omega)$, we first choose $y^\delta = y + 5\frac{R}{\|R\|_{L^2(\Omega)}}$, where R is computed with the MATLAB routine $randn(size(y))$ with setting $randn('state', 0)$. u^δ is then obtained by solving (6.1) with y replaced by y^δ. We obtain

$$\|u^\delta - u^*\|_{H^1(\Omega)} = 0.0928 \approx 0.1, \quad \frac{\|u^\delta - u^*\|_{H^1(\Omega)}}{\|u^*\|_{H^1(\Omega)}} = 0.0044.$$

Figure 6.7 displays the step sizes $\frac{1}{s^n}$ of Algorithm 1, Algorithm 1B, and Algorithm 1N. Similar to the case of exact data, Algorithm 1 tends to choose larger step sizes and Al-

ETNA
Kent State University
http://etna.math.kent.edu

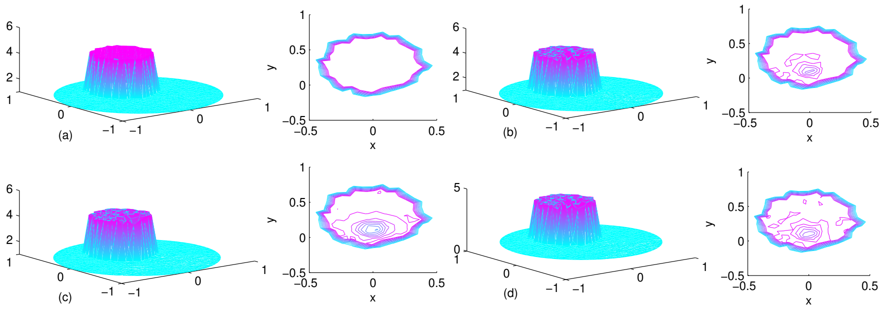

FIG. 6.3. *3D-plots and contour plots: a) exact a^*; b)–d) a^n with $n = 300$ in Algorithm 1, Algorithm 1B and Algorithm 1N, respectively.*

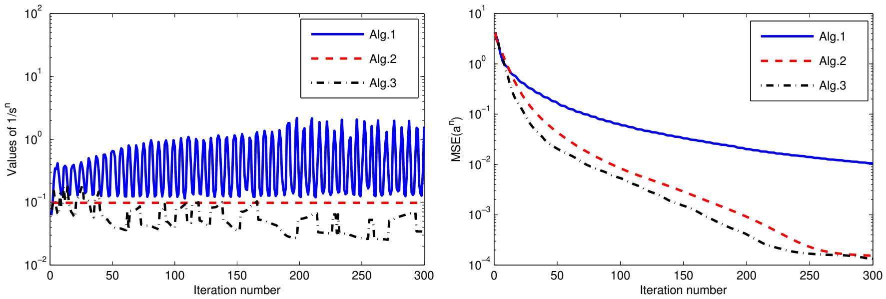

FIG. 6.4. *The values of $\frac{1}{s^n}$ and $MSE(a^n)$ in Algorithm 1, Algorithm 2, and Algorithm 3.*

gorithm 1B chooses the smallest step sizes. In this case, the interval $[\underline{s}, \overline{s}]$ is needed for Algorithm 1 at some iterations otherwise the step sizes computed by the Barzilai-Borwein rule without using the projection $P_{[\overline{s}^{-1}, \underline{s}^{-1}]}$ are out of the interval $[\underline{s}, \overline{s}]$, e.g., at some iterations in the blue ellipses in the figure. The figure also shows that $\{MSE(a^n)\}$ in Algorithm 1B decreases most slowly. During the first iterations, $\{MSE(a^n)\}$ in both Algorithm 1 and Algorithm 1N decrease fast, but the error increases again after a few iterations. This might have several reasons. We suspect that the regularization parameters were chosen too small ($\alpha = 5 \cdot 10^{-5}$), i.e., the resulting ill-conditioned system shows the typical semi-convergent behavior of iteration methods for inverse problems. If we change the value of α, then the shape and values of the sequences $MSE(a^n)$ in the algorithms are changed too, but the semi-convergent behavior will be observed after a certain number of iterations. As a remedy, a suitable stopping criterion could be incorporated. Such a criterion could be that the algorithms is stopped when $|\Theta(a^n) - \Theta(a^{n-1})| \le \epsilon$ for some ϵ small enough. Alternatively the discrepancy principle [17] could be used, or one of the ϵ-free stopping criteria described, e.g., in [22, 25, 27, 28] could be adapted to the present nonlinear situation. However, we do not consider these problems in this paper and such a criterion is not used here.

Figure 6.8 shows that the decay rate of $\Theta(a^n)$ in Algorithm 1 is the fastest with respect to the iteration counter and time. Similar to the noise free case, the computational time of Algorithm 1N is higher than that of Algorithm 1, and Algorithm 1B spent the least time.

In both two cases, exact data and noisy data, the initial guesses of the step sizes computed by (4.2) are efficient in practice. They are adaptive and large enough; only occasionally further iterations are needed for reducing the step sizes in order to satisfy the respective criteria.

Figure 6.9 displays a^* and a^n, where n is taken with respect to the minimum values

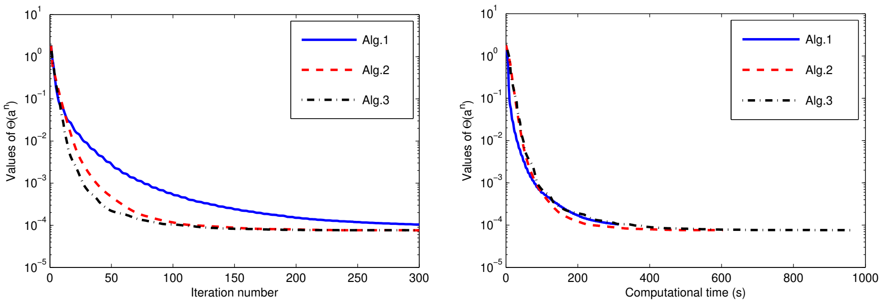

FIG. 6.5. *The decrease rate of* $\Theta(a^n)$ *in Algorithm 1, Algorithm 2, and Algorithm 3.*

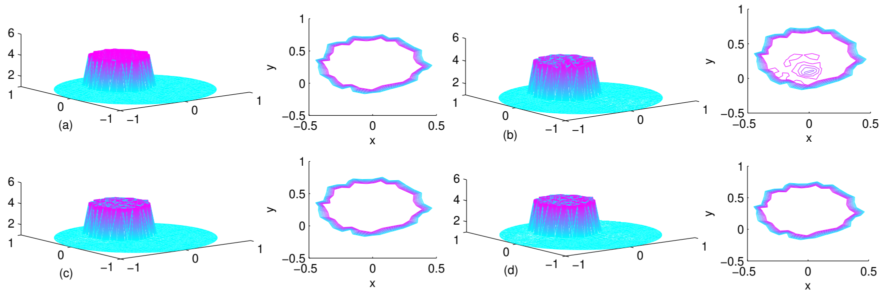

FIG. 6.6. *3D-plots and contour plots: a) exact* a^**; b)–d)* a^n *with* $n = 300$ *in Algorithm 1, Algorithm 2, and Algorithm 3, respectively.*

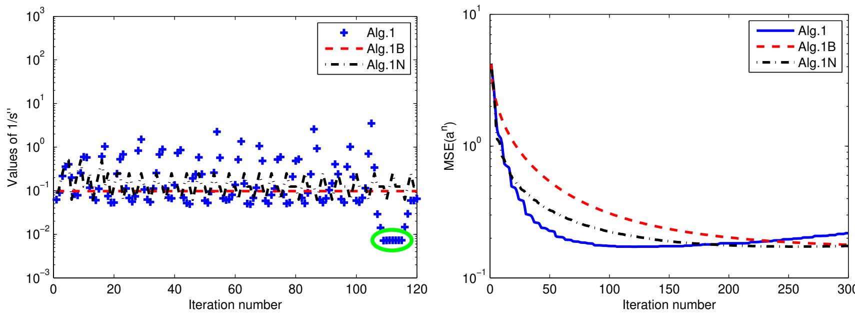

FIG. 6.7. *The values of* $\frac{1}{s^n}$ *and* $MSE(a^n)$ *in Algorithm 1, Algorithm 1B, and Algorithm 1N.*

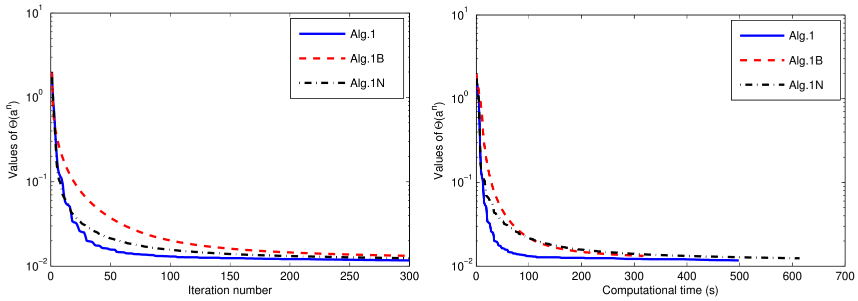

FIG. 6.8. *The decrease rate of* $\Theta(a^n)$ *in Algorithm 1, Algorithm 1B, and Algorithm 1N.*

ETNA
Kent State University
http://etna.math.kent.edu

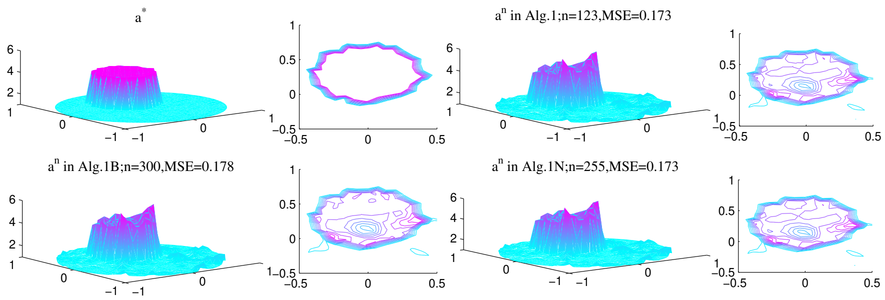

FIG. 6.9. *3D-plots and contour plots of a^* and a^n in Algorithm 1, Algorithm 1B, and Algorithm 1N.*

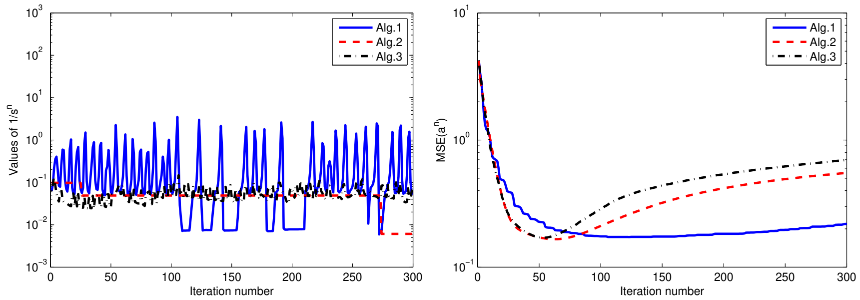

FIG. 6.10. *The values of $\frac{1}{s^n}$ and $MSE(a^n)$ in Algorithm 1, Algorithm 2 and Algorithm 3.*

of $MSE(a^n)$ in Algorithm 1, Algorithm 1B, and Algorithm 1N, respectively. It shows that the algorithms still recover a^* quite well and have the same accuracy.

Now we consider Algorithm 1 and its accelerated versions (Algorithm 2 and Algorithm 3). Figure 6.10 shows that the minimum values of $MSE(a^n)$ in Algorithm 2 and Algorithm 3 are smaller than that of Algorithm 1. Therefore, with a suitable stopping criterion, the accelerated algorithms can obtain good approximations of a^*.

In Figure 6.11, the convergence rate of $\Theta(a^n)$ in Algorithm 1 is the slowest. They seem to be the same in Algorithm 2 and Algorithm 3, and the two accelerated algorithms still take more time than the original algorithm.

Figure 6.12 illustrates a^* and a^n, where n is taken with respect to the minimum values of $MSE(a^n)$ in Algorithm 1, Algorithm 2, and Algorithm 3, respectively. Here, $MSE(a^n)$ in Algorithm 2 and Algorithm 3 are smaller than that in Algorithm 1.

7. Conclusion. We have proposed an algorithm based on quadratic approximations as well as two accelerated versions for the minimization problem

$$\min_{u\in\mathcal{H}} \Theta(u) := F(u) + \Phi(u),$$

where $\mathcal{H}$ is a Hilbert space, $F : \mathcal{H} \to \mathbb{R}$ is a smooth but not necessarily convex mapping and $\Phi : \mathcal{H} \to \mathbb{R}$ is defined by $\Phi(u) = \alpha \sum_{k\in\Lambda} \omega_k |\langle u, \varphi_k\rangle|^p$, where $p \in [1, 2]$, $\omega_k \geqslant \omega_{\min} > 0, \forall k$, and $\{\varphi_k\}$ is an orthonormal basis of $\mathcal{H}$.

Under Assumption 2, Algorithm 1 is proved to converge. We have also analyzed different strategies for improving the step size selection. In addition, if F is convex then Algorithm 2 and Algorithm 3 are proved to converge. The convergence rate of the objective functional Θ in Algorithm 1 is $O(1/n)$, two accelerated algorithms (Algorithm 2 and Algorithm 3) are of order $O(1/n^2)$. This rate is known to be optimal for general gradient methods. The numerical examples demonstrate the efficiency of the algorithms.

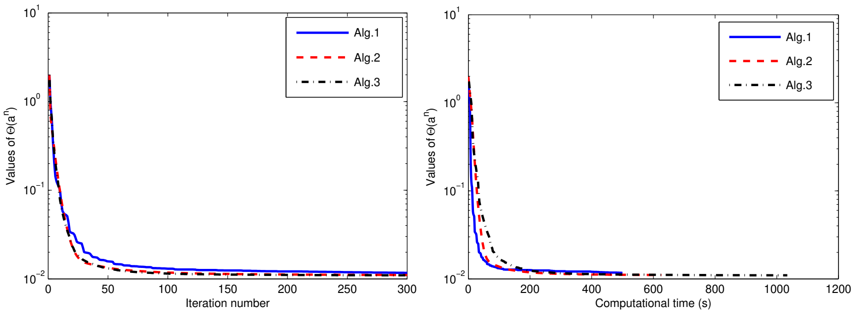

FIG. 6.11. *The decrease rate of $\Theta(a^n)$ in Algorithm 1, Algorithm 2 and Algorithm 3.*

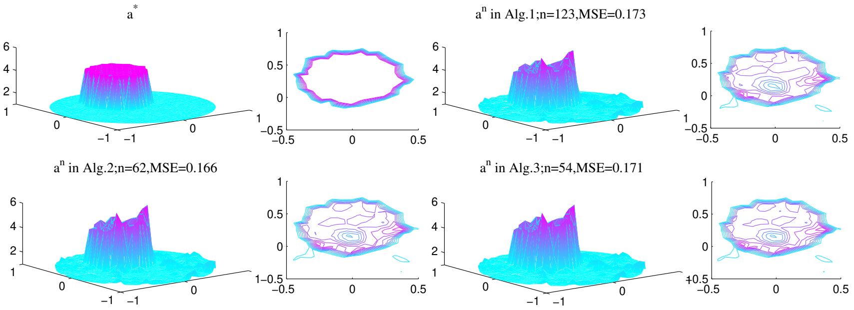

FIG. 6.12. *3D-plots and contour plots of exact a^* and a^n in Algorithm 1, Algorithm 2 and Algorithm 3.*

Acknowledgments. The authors would like to thank Prof. Dr. Dinh Nho Hao, who introduced and suggested to investigate Nesterov's algorithm. The third author was supported by a scholarship of the graduate program "Scientific Computing in Engineering (SCiE)" for supporting the Ph.D scholarship. The second author gratefully acknowledges the financial support by Deutsche Forschungsgemeinschaft (DFG) for Subproject C2 within the SFB 747 'Mikrokaltumformen'.

REFERENCES

[1] R. ACAR AND C. R. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.

[2] J. AUJOL, *Some first-order algorithms for total variation based image restoration*, J. Math. Imaging Vision, 34 (2009), pp. 307–327.

[3] H. T. BANKS AND K. KUNISCH, *Estimation Techniques for Distributed Parameter Systems*, Birkhäuser, Boston, 1989.

[4] J. BARZILAI AND J. M. BORWEIN, *Two-point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.

[5] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[6] T. BONESKY, K. BREDIES, D. A. LORENZ, AND P. MAASS, *A generalized conditional gradient method for nonlinear operator equations with sparsity constraints*, Inverse Problems, 23 (2007), pp. 2041–2058.

[7] K. BREDIES AND D. A. LORENZ, *Linear convergence of iterative soft-thresholding*, J. Fourier Anal. Appl., 14 (2008), pp. 813–837.

[8] K. BREDIES, D. A. LORENZ, AND P. MAASS, *A generalized conditional gradient method and its connection to an iterative shrinkage method*, Comput. Optim. Appl., 42 (2009), pp. 173–193.

[9] T. F. CHAN AND X. TAI, *Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients*, J. Comput. Phys., 193 (2004), pp. 40–66.

[10] Z. CHEN AND J. ZOU, *An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems*, SIAM J. Control Optim., 37 (1999), pp. 892–910.

[11] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[12] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math, 57 (2004), pp. 1413–1457.

[13] V. F. DEMYANOV AND A. M. RUBINOV, *Approximate Methods in Optimization Problems*, American Elsevier, New York, 1970.

[14] J. C. DUNN, *Rates of convergence for conditional gradient algorithms near singular and nonsingular extremals*, SIAM J. Control Optim., 17 (1979), pp. 187–211.

[15] ———, *Convergence rates for conditional gradient sequences generated by implicit step length rules*, SIAM J. Control Optim., 18 (1980), pp. 473–487.

[16] ———, *Global and asymptotic convergence rate estimates for a class of projected gradient processes*, SIAM J. Control Optim., 19 (1981), pp. 368–400.

[17] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[18] R. FALK, *Error estimates for the numerical identification of a variable coefficient*, Math. Comp., 40 (1983), pp. 537–546.

[19] M. FRANK AND P. WOLFE, *An algorithm for quadratic programming*, Naval Res. Logist. Quart., 3 (1956), pp. 95–110.

[20] M. GRASMAIR, M. HALTMEIER, AND O. SCHERZER, *Sparse regularization with l^q penalty term*, Inverse Problems, 24 (2008), 055020 (13 pages).

[21] R. GRIESSE AND D. A. LORENZ, *A semismooth Newton method for Tikhonov functionals with sparsity constraints*, Inverse Problems, 24 (2008), 035007 (19 pages).

[22] U. HÄMARIK, R. PALM, AND T. RAUS, *Comparison of parameter choices in regularization algorithms in case of different information about noise level*, Calcolo, 48 (2011), pp. 47–59.

[23] D. N. HAO AND T. N. T. QUYEN, *Convergence rates for Tikhonov regularization of coefficient identification problems in Laplace-type equation*, Inverse Problems, 26 (2010), 125014 (23 pages).

[24] B. HOFMANN, B. KALTENBACHER, C. PÖSCHL, AND SCHERZER, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems, 23 (2007), pp. 987–1010.

[25] B. JIN AND D. A. LORENZ, *Heuristic parameter-choice rules for convex variational regularization based on error estimates*, SIAM J. Numer. Anal., 48 (2010), pp. 1208–1229.

[26] B. JIN AND P. MAASS, *An analysis of electrical impedance tomography with applications to Tikhonov regularization*, ESAIM Control Optim. Calc. Var., in press, 2012.

[27] S. KINDERMANN, *Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems*, Electron. Trans. Numer. Anal., 38 (2011), pp. 233–257. http://etna.math.kent.edu/vol.38.2011/pp233-257.dir

[28] S. KINDERMANN AND A. NEUBAUER, *On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization*, Inverse Probl. Imaging, 2 (2008), pp. 291–299.

[29] I. KNOWLES, *Parameter identification for elliptic problems*, J. Comput. Appl. Math., 131 (2001), pp. 175–194.

[30] R. V. KOHN AND B. LOWE, *A variational method for parameter identification*, RAIRO Modél. Math. Anal. Numér., 22 (1988), pp. 119–158.

[31] C. KRAVARIS AND J. H. SEINFELD, *Identification of parameters in distributed parameter systems by regularization*, SIAM J. Control Optim., 23 (1985), pp. 217–241.

[32] D. A. LORENZ, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, J. Inverse Ill-Posed Probl., 16 (2008), pp. 463–478.

[33] D. A. LORENZ AND A. RÖSCH, *Error estimates for joint Tikhonov and Lavrentiev regularization of constrained control problems*, Appl. Anal., 89 (2010), pp. 1679–1691.

[34] I. LORIS, *On the performance of algorithms for the minimization of ℓ_1-penalized functionals*, Inverse Problems, 25 (2009), 035008 (16 pages).

[35] Y. NESTEROV, *Gradient methods for minimizing composite objective function*, CORE Discussion Papers 2007/76, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), Sept. 2007.

[36] E. POLAK, *An historical survey of computational methods in optimal control*, SIAM Rev., 15 (1973), pp. 553–584.

[37] R. RAMLAU AND E. RESMERITA, *Convergence rates for regularization with sparsity constraints*, Electon. Trans. Numer. Anal., 37 (2010), pp. 87–104. http://etna.mcs.kent.edu/vol.37.2010/pp87-104.dir

[38] R. RAMLAU AND G. TESCHKE, *A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints*, Numer. Math., 104 (2006), pp. 177–203.

[39] G. R. RICHTER, *Numerical identification of a spatially varying diffusion coefficient*, Math. Comp., 36 (1981), pp. 375–386.

[40] P. WEISS, L. FÉRAUD, AND G. AUBERT, *Efficient schemes for total variation minimization under constraints in image processing*, SIAM J. Sci. Comput., 31 (2009), pp. 2047–2080.
[41] L. W. WHITE AND J. ZHOU, *Continuity and uniqueness of regularized output least squares optimal estimators*, J. Math. Anal. Appl., 196 (1995), pp. 53–83.
[42] J. ZOU, *Numerical methods for elliptic inverse problems*, Int. J. Comput. Math., 70 (1998), pp. 211–232.