Iterative Regularization of Parameter Identification Problems by SQP Methods

Martin Burger* Wolfram Mühlhuber*

SFB F 013 Numerical and Symbolic Scientific Computing
Johannes Kepler Universität, Freistädterstr. 313, A-4040 Linz, Austria

Abstract

The aim of this paper is to design and to analyze sequential quadratic programming (SQP) methods as iterative regularization methods for ill-posed parameter identification problems. We discuss two variants of the original SQP-algorithm, in which an additional stabilizer ensures the strict convexity and well-posedness of the quadratic programming problems that have to be solved in each step of the iteration procedure. We show that the sequential quadratic programming problems are equivalent to stable saddle-point problems, which can be analyzed by standard methods. In addition, the investigation of these saddle-point problems offers new possibilities for the numerical treatment of the identification problem compared to standard numerical methods for inverse problems.

One of the resulting iteration algorithms, called *Levenberg-Marquardt SQP method*, is analyzed with respect to convergence and regularizing properties under an appropriate choice of the stopping index depending on the noise level. Finally, we show that the conditions needed for convergence are fulfilled for several important types of applications and we test the convergence behavior in numerical examples.

Keywords: Parameter Identification, Sequential Quadratic Programming, Iterative Regularization, Ill-Posed Problems, Saddle Point Problems, Partial Differential Equations.
AMS Subject Classification (2000): 35R30, 90C55, 65J20

1 Introduction

Since distributed parameters have to be determined from indirect measurements in many applications that are modeled by partial differential equations, *parameter identification* has become an important part of mathematical modeling (cf. e.g. [2, 11, 16, 22, 23] and the references therein). The majority of these identification problems is *ill-posed*, i.e., the parameter does not depend on the data, which cannot be measured exactly in practice, in a stable way. Therefore regularization methods have to be used in order to obtain stable approximations of the solution in presence of data noise. We refer to [13, 29] for an overview of regularization methods for inverse ill-posed problems.

Classical approaches to the regularization of parameter identification problems are direct methods such as *Tikhonov regularization* (cf. e.g. [9, 14, 15]), which replace the least-squares problem by a close stable problem. Recently, also the application of *iterative regularization*

*This work has been supported by the Austrian National Science Foundation FWF under project grants SFB F 13/08 and F 13/09.

methods methods has been investigated (cf. e.g. [20, 21, 27, 28, 37]), where the iteration followed the feasible path defined by the underlying state equation. We will review this approach in Section 1.2. Our aim is to design new iterative regularization methods based on the idea of sequential quadratic programming, which means to minimize a second-order approximation of the Lagrange functional subject to the linearized state equation in each iteration step, resulting in a sequence of quadratic programming problems. An overview of SQP methods will be given in Section 1.3.

An important difference of SQP-type methods to the feasible path approach is that the equation (1.2) is interpreted as a constraint in the product space $X \times Q$ and approximated by a linearized version in each iteration step. So far, SQP-type method have been used in this context only as an optimization method for the Tikhonov-functional (cf. [19, 24, 32]), i.e., the regularization procedure is independent of the SQP-iteration. Our approach is based on the paradigm of Levenberg-Marquardt methods, i.e., we will add a stabilizer to the arising quadratic optimization problem, which penalizes large iteration steps. The regularizing effect will then come from an appropriate early termination of the iteration procedure, with a stopping index chosen in dependence of the noise level δ.

The paper is organized as follows: In Section 2 we will present two new classes of iterative regularization methods for parameter identification problems and discuss some of their basic properties. In Section 3 we will carry out the convergence analysis for a rather general class of identification problems, before applying our results to some well-studied identification problems in Section 4. Numerical experiments with typical model problems are presented in Section 5. Finally, we conclude and discuss possible extensions and variants in Section 6.

1.1 Notations and Definitions

The basic setup of the identification problems to be discussed in this paper is as follows: given an observation

$$z := E\hat{u}, \tag{1.1}$$

where $E : X \to Z$ is a bounded linear operator, we want to identify the parameter $q \in Q_{ad} \subset Q$ in the underlying equation

$$e(u, q) = f, \tag{1.2}$$

where $e : X \times Q \to Y$ is a continuous nonlinear operator. In this setup we assume that Q and Z are Hilbert spaces, that Q_{ad} is a closed subset of Q with nonempty interior and that X and Y are appropriate Banach spaces. In addition, we assume that the operator e is homogeneous, i.e,

$$e(0, 0) = 0, \tag{1.3}$$

which is no restriction of generality, since for an arbitrary operator e we can transform (1.2) into an equivalent equation with homogeneous operator via

$$\tilde{e}(u, q) := e(u, q) - e(0, 0), \qquad \tilde{f} = f - e(0, 0).$$

In practice one has to deal with data z^δ that are corrupted by noise instead of the exact data; we assume that the observation error is bounded by

$$\|z - z^\delta\|_Z \le \delta, \tag{1.4}$$

where $z = E\hat{u}$ such that there exists a $\hat{q} \in Q_{ad}$ with

$$e(\hat{u}, \hat{q}) = f. \tag{1.5}$$

The general notation for state (u), parameter (q) and Lagrangian variable (λ) in this paper will be as follows: unless further noticed, (u, q, λ) and (φ, σ, μ) denote arbitrary elements of the spaces X, Q and Y^*, the pair $(\hat{u}, \hat{q})$ denotes an exact solution of the parameter identification problem (as introduced above), and $(u_k, q_k, \lambda_k) \in X \times Q \times Y^*$ denotes the actual iterate.

Note that under typical conditions, parameter identification problems in equations of the form (1.2) are ill-posed, i.e., an arbitrarily small error in the data z can lead to an arbitrarily large deviation in the reconstructed parameter q. In presence of noise, a solution of the equation $Eu = z^\delta$ need not exist, and therefore one has to consider the corresponding normal equation respectively the least-squares problem

$$\frac{1}{2}\|Eu - z\| \to \min_{(u,q) \in X \times Q_{ad}} \tag{1.6}$$

subject to (1.2). Because of the ill-posedness of the problem, a direct application of standard SQP-type methods to this least-squares problem is not possible, since the quadratic problems arising in each of the iteration steps are most likely ill-posed. Therefore we will present and analyze a new SQP-type approach in this paper, which leads to stable quadratic subproblems due to an additional penalty term in the parameter space.

1.2 Iterative Regularization on Feasible Paths

Iterative regularization methods are usually formulated for operator equations of the form

$$F(q) = z^\delta, \tag{1.7}$$

where $F : Q \to Y$ is a continuous nonlinear operator. All common iterative regularization methods use Frèchet-derivatives of F; they can be divided into two different classes, namely explicit methods such as steepest descent methods (cf. [37]) or the Landweber iteration (cf. e.g. [21, 42, 43])

$$q_{k+1} = q_k - \omega F'(q_k)^*(F(q_k) - z^\delta), \tag{1.8}$$

with appropriately chosen damping parameter $\omega > 0$, and implicit methods such as Newton-type (cf. e.g. [12, 20, 27]) or Quasi-Newton methods (cf. e.g. [28, 39])

$$B_k(q_{k+1} - q_k) = -F'(q_k)^*(F(q_k) - z^\delta), \tag{1.9}$$

where B_k is a stable approximation of the Gauss-Newton matrix $F'(q_k)^*F'(q_k)$.

The regularizing effect of an iterative regularization method comes from the early termination of the iteration procedure, where the stopping index is chosen in dependence of the noise level δ and the noisy data z^δ. We refer to the survey paper by Engl and Scherzer [17] for an overview of iterative regularization methods.

The application of iterative regularization methods to parameter identification problems is usually based on a feasible-path approach, i.e., the state equation is eliminated and the resulting parameter-to-output map is the concatenation of the parameter-to-state map $q \mapsto u$ and the observation operator E. A numerical disadvantage of this method is the possibly large number of state equations that have to be solved during the iteration procedure. E.g., for the Landweber iteration, each iteration enforces the evaluation of the operator F and the adjoint of its derivative $F'(q_k)^*$. This means that one has to solve the state equation and a related equation for the adjoint, which is basically of the same effort (cf. e.g. [21]). A Newton-type method usually yields a reasonable approximation of the solution in less iterations than an explicit method, but the effort in each step of the iteration is higher, since many evaluations of $F'(q_k)$ are needed, each one of similar effort as the solution of the state equation. For some parameter identification problems, numerical investigations even show that the overall numerical effort for standard Newton-type methods is not lower than the one for the 'slower' Landweber-iteration (cf. e.g. [6, 7]).

A method of particular interest in the following is the *Levenberg-Marquardt method* (cf. [20]), which is of the form (1.9) with

$$B_k = F'(q_k)^* F'(q_k) + \alpha_k I, \qquad \alpha_k \in \mathbb{R}^+. \tag{1.10}$$

Hanke [20] showed that the Levenberg-Marquardt method is locally convergent if the nonlinearity condition

$$\|F(\tilde{q}) - F(q) - F'(q)(\tilde{q} - q)\| \leq c\|\tilde{q} - q\| \; \|F(\tilde{q}) - F(q)\|, \tag{1.11}$$

with some constant $c \in \mathbb{R}^+$, is fulfilled for all q, $\tilde{q}$ in a neighborhood of a solution. In the presence of noise, local convergence is obtained if the stopping index is chosen according to the generalized discrepancy principle, i.e.,

$$\|F(q_{k_*}) - z^\delta\|_Z \leq \tau\delta < \|F(q_k) - z^\delta\|_Z, \qquad \forall \; k < k_*, \tag{1.12}$$

with appropriate $\tau > 1$.

1.3 Sequential Quadratic Programming

The classical *sequential quadratic programming* method (also called *Lagrange-Newton* method) is applied to general optimization problems of the form

$$J(x) \to \min_{x \in X} \tag{1.13}$$

subject to equality and inequality constraints of the form

$$e(x) = 0, \qquad \text{and} \qquad c(x) \leq 0, \tag{1.14}$$

where the functional $J : X \to \mathbb{R}$ and the operators $e : X \to Y_1$, $c : X \to Y_2$ are twice continuously Frèchet-differentiable and $\leq$ represents an order on the Banach space Y_2.

An iteration step of the SQP-method is given by $x_{k+1} = x_k + y_k$, where y_k is a minimizer of the quadratic optimization problem

$$J(x_k) + J'(x_k)y + \frac{1}{2}\mathcal{L}''(x_k; \lambda_k, \mu_k)(y, y) \to \min_{y \in X}, \tag{1.15}$$

subject to

$$\begin{aligned} e(x_k) + e'(x_k)y &= 0, && (1.16)\\ c(x_k) + c'(x_k)y &\leq 0, && (1.17) \end{aligned}$$

where $\mathcal{L}$ represents the Lagrangian of the optimization problem (1.13), (1.14) given by

$$\mathcal{L}(x, \lambda, \mu) = J(x) + \langle e(x), \lambda \rangle + \langle c(x), \mu \rangle. \tag{1.18}$$

In general, the linearized constraints (1.16) and (1.17) might define an empty set of admissible points, but we will exclude this case in the following motivated by the specific structure of the underlying state equations under investigation (see Section 2).

An iteration step of the sequential quadratic programming method is well-defined if the quadratic objective functional in (1.15) is strictly convex on the set defined by the linear constraints (1.16) and (1.17), which is not obvious for problems with little insight into the structure of the objective functional and the constraints. However, for certain classes of problems such as in the optimal control of elliptic or parabolic differential equations, SQP-methods have been extensively studied and applied with particular numerical success (cf. e.g. [3, 25, 35, 41]). Frequently used are also two variants of the SQP-method, namely *augmented Lagrangian SQP-methods* (cf. e.g. [25, 24, 30]), where a penalty term originating from the constraint (1.16) is added to the objective functional, and *reduced SQP-methods* (cf. e.g. [32, 33, 41]), where (1.16) is eliminated a-priorily.

The convergence analysis of sequential quadratic programming methods is based on the quadratic well-posedness of the optimization problem (cf. [1, 25, 30, 33] for further details), which is not satisfied for an ill-posed parameter identification problem. Therefore we will investigate stabilized versions of the SQP-method, which guarantee the well-posedness of the quadratic programming problems and converge also for problems without second-order regularity. Our SQP-approach for parameter identification problems will be introduced in detail in Section 2.

1.4 Linear Saddle-Point Problems

Using first-order optimality conditions, which are not only necessary but also sufficient for regular quadratic problems, an iteration step in the sequential quadratic programming method (disregarding inequality constraints for the moment) can be rewritten as the solution of a linear saddle-point problem. The solution and numerical approximation of linear saddle-point problems arising from Lagrangian multipliers have been well-studied over the last decades after the seminal paper by Brezzi [4]. In the following let U and Λ be two Hilbert spaces, let $g \in U^*$, $f \in \Lambda^*$ and let $a : U \times U \to \mathbb{R}$ and $b : U \times \Lambda \to \mathbb{R}$ be continuous bilinear forms. Then a symmetric linear saddle-point problem in variational formulation consists of searching for a solution $(u, \lambda) \in U \times \Lambda$ of

$$a(u, v) + b(v, \lambda) = \langle g, v \rangle, \qquad \forall\, v \in U, \tag{1.19}$$

$$b(u, \mu) = \langle f, \mu \rangle, \qquad \forall\, \mu \in \Lambda. \tag{1.20}$$

The well-posedness of (1.19), (1.20) can be studied under additional conditions on a and b, namely the so-called *kernel-ellipticity* of a,

$$\exists\, \alpha_a \in \mathbb{R}^+ :\ a(v, v) \geq \alpha_a \|v\|_U^2, \qquad \forall\, v \in K^b := \{\ v \in U \mid b(v, \mu) = 0, \forall\, \mu \in \Lambda\ \}, \tag{1.21}$$

and the *LBB-condition* upon b,

$$\exists\, \alpha_b \in \mathbb{R}^+ :\ \inf_{\mu \in \Lambda} \sup_{v \in U} \frac{b(v, \mu)}{\|v\|_U \|\mu\|_\Lambda} \geq \alpha_b. \tag{1.22}$$

Under these assumptions, the following classical result can be shown (cf. [4, 5]):

Theorem 1.1. *Let a, b as above, such that* (1.21) *and* (1.22) *are satisfied. Then the linear saddle-point problem* (1.19), (1.20) *has a unique solution* $(u, \lambda) \in U \times \Lambda$, *which depends continuously on the data* $(f, g) \in \Lambda^* \times U^*$.

2 The SQP-Approach to Parameter Identification

Following the exposition in Section 1.3, the obvious SQP-approach to parameter identification problems is to apply the classical method of sequential quadratic programming to the least-squares problem

$$\frac{1}{2}\|Eu - z\|^2 \to \min_{(u,q)\in X\times Q}, \tag{2.1}$$

subject to the equality constraint (1.2) and to $q \in Q_{ad}$. Since our main focus in this paper is the treatment of the equality constraint, we will omit the additional constraint $q \in Q_{ad}$ in the following.

Due to the ill-posedness of the original problem, a minimizer of the quadratic programming problem (1.15)-(1.17) needs not exist and if one exists, it might not depend on the data in a stable way. Thus, it seems a good choice to add stabilizing terms analogous to Newton-type methods to the objective functionals arising in each step of the iteration. This results in the following iteration method:

Method 1 (Iteratively Regularized Sequential Quadratic Programming Method). Let $(u_0, q_0, \lambda_0) \in X \times Q \times Y^*$ be a given initial value and let $(\beta_k)_{k\in\mathbb{N}}$ be a bounded sequence of positive real numbers. The method of *iteratively regularized sequential quadratic programming* (IRSQP) consists of the iteration procedure

$$(u_{k+1}, q_{k+1}, \lambda_{k+1}) = (\overline{u}_k, \overline{q}_k, \overline{\lambda}_k), \tag{2.2}$$

where $(\overline{u}_k, \overline{q}_k)$ is the minimizer of the quadratic programming problem

$$\frac{1}{2}\|Eu - z\|_Z^2 + \frac{\beta_k}{2}\|q - q_k\|_Q^2 + \langle \lambda_k, e''(u_k, q_k)(u - u_k, q - q_k)^2\rangle \to \min_{(u,q)\in X\times Q}. \tag{2.3}$$

subject to the linear constraint

$$e(u_k, q_k) + e'(u_k, q_k)(u - u_k, q - q_k) = f, \tag{2.4}$$

and $\overline{\lambda}_k$ is the corresponding Lagrange-multiplier.

Note that the IRSQP-method involves second order derivatives of the operator e, which are usually ignored by iterative methods for least-squares problems. Therefore, we define a variant of Method 1, which takes into account the special structure of the objective functional.

Method 2 (Levenberg-Marquardt Sequential Quadratic Programming Method). Let $(u_0, q_0) \in X \times Q$ be a given initial value and let $(\beta_k)_{k\in\mathbb{N}}$ be a bounded sequence of positive real numbers. The *Levenberg-Marquardt sequential quadratic programming* method (LMSQP) consists of the iteration procedure

$$(u_{k+1}, q_{k+1}) = (\overline{u}_k, \overline{q}_k), \tag{2.5}$$

where $(\overline{u}_k, \overline{q}_k) \in X \times Q$ is the minimizer of the quadratic programming problem

$$\frac{1}{2}\|Eu - z^\delta\|_Z^2 + \frac{\beta_k}{2}\|q - q_k\|_Q^2 \to \min_{(u,q)\in X\times Q}, \tag{2.6}$$

subject to the linear constraint

$$e(u_k, q_k) + e'(u_k, q_k)(u - u_k, q - q_k) = f. \tag{2.7}$$

An important question is the well-definedness of the iteration procedures for the IRSQP and the LMSQP method, i.e., the existence and uniqueness of the minimizers of (2.3), (2.4) and (2.6), (2.7), respectively. Another important property for an iterative regularization method is the stable dependence of the iterates on the previous iterates and on the data. We will investigate these questions in the following sections.

2.1 Well-Posedness of the Quadratic Programming Problems

In the following we will verify the well-posedness of the quadratic programming problems (2.3), (2.4) and (2.6), (2.7) under reasonable assumptions on the equation operator e.

In typical applications, the equation (1.2), respectively its linearization, admits a unique solution with respect to the state, i.e.,

$$e_u(u,q)^{-1} : Y \to X \text{ exists and is a continuous linear operator for all } (u,q) \in X \times Q. \tag{2.8}$$

Proposition 2.1. *Let e be continuously Frèchet-differentiable, let* (2.8) *hold and let $\beta_k > 0$. Then the quadratic programming problem* (2.6), (2.7) *has a unique solution $(\overline{u}_k, \overline{q}_k) \in X \times Q$, which is also the only local minimum.*

Proof. We reformulate (2.6), (2.7) in terms of $v = u - u_k$ and $s = q - q_k$. The set of constraints is then given by

$$M = \{\ (v,s) \mid e'(u_k,q_k)(v,s) = f - e(u_k,q_k)\ \},$$

which is a closed, convex and nonempty set, because $e'(u_k, q_k)$ is a continuous linear operator and

$$(v_0, s_0) := \big(e_u(u_k,q_k)^{-1}(f - e(u_k,q_k)), 0\big) \in M,$$

which is due to (2.8). Since the objective functional

$$J_k(v,s) := \frac{1}{2}\|Ev + Eu_k - z\|_Z^2 + \frac{\beta_k}{2}\|s\|_Q^2$$

is convex on M, the main theorem of convex optimization (cf. [44, Thm.47.C]) implies the existence of a solution as well as the convexity and closedness of the solution set.

Now let $\epsilon \in (0,1)$ and $(v_i, s_i) \in M$ for $i = 1,2$, then a straight-forward estimate yields

$$J_k(\epsilon v_1 + (1-\epsilon)v_2, \epsilon s_1 + (1-\epsilon)s_2) \le \epsilon J_k(v_1,s_1) + (1-\epsilon)J_k(v_2,s_2) - \epsilon(1-\epsilon)\|s_1 - s_2\|^2.$$

Since (2.8) implies

$$(v_1 - v_2) = -e_u(u_k,q_k)^{-1}e_q(u_k,q_k)(s_1 - s_2), \qquad \forall\ (v_i,s_i) \in M, i = 1,2,$$

we have $s_1 = s_2$ only if $(v_1, s_1) = (v_2, s_2)$. Thus,

$$J_k(\epsilon v_1 + (1-\epsilon)v_2, \epsilon s_1 + (1-\epsilon)s_2) < \epsilon J_k(v_1, s_1) + (1-\epsilon)J_k(v_2, s_2)$$

for all $\epsilon \in (0,1)$ and all distinct points $(v_i, s_i) \in M$, $i = 1, 2$, i.e., the functional J_k is strictly convex. Hence, from the main theorem of convex optimization we may conclude that there exists a unique global minimum and no further local minima. □

Proposition 2.2. *Let e be twice continuously Frèchet-differentiable, let* (2.8) *hold and let*

$$\beta_k > 2 \ \max\left\{1, \|e_u(u_k, q_k)^{-1} e_q(u_k, q_k)\|^2\right\} \ \|\lambda_k\|_{Y^*} \ \|e''(u_k, q_k)\|. \tag{2.9}$$

Then the quadratic programming problem (2.3), (2.4) *has a unique solution $(\overline{u}_k, \overline{q}_k) \in X \times Q$, which is also the only local minimum.*

Proof. As in the proof of Proposition 2.1, we write the quadratic programming problem (2.3), (2.4) in terms of (v, s). The arising objective functional

$$J_k(v,s) := \frac{1}{2}\|Ev + Eu_k - z\|_Z^2 + \frac{\beta_k}{2}\|s\|_Q^2 + \frac{1}{2}\langle \lambda_k, e''(u_k, q_k)(v,s)^2\rangle$$

is twice continuously Frèchet-differentiable with

$$\begin{aligned} J_k''(v,s)(\varphi,\sigma)^2 &= \|E\varphi\|_Z^2 + \beta_k\|\sigma\|_Q^2 + \langle \lambda_k, e''(u_k, q_k)(\varphi,\sigma)^2\rangle \\ &\geq \beta_k\|\sigma\|_Q^2 - \|\lambda_k\|_{Z^*} \ \|e''(u_k, q_k)\| \ (\|\varphi\|_X^2 + \|\sigma\|_Q^2) \end{aligned}$$

for all $(\varphi, \sigma) \in X \times Q$ and all $(v, s) \in M$, where M is defined as in the proof of Proposition 2.1. Now let $(\varphi + v, s + \sigma) \in M$, i.e.,

$$e'(u_k, q_k)(\varphi, \sigma) = 0,$$

then (2.8) implies that

$$\|\varphi\|_X \leq \|e_u(u_k, q_k)^{-1} e_q(u_k, q_k)\| \ \|\sigma\|_Q.$$

Hence, we obtain that

$$\begin{aligned} &J_k''(v,s)(\varphi,\sigma)^2 \geq \\ &\quad \left(\frac{\beta_k}{2} \min\left\{1, \|e_u(u_k, q_k)^{-1} e_q(u_k, q_k)\|^{-2}\right\} - \|\lambda_k\|_{Z^*} \ \|e''(u_k, q_k)\|\right) (\|\varphi\|_X^2 + \|\sigma\|_Q^2), \end{aligned}$$

for all $(v, s) \in M$, $(v + \varphi, s + \sigma) \in M$, which implies with (2.9) the strict convexity of J_k on M (cf. [44, p.48]) and the assertions follow as in the proof of Proposition (2.1). □

Note that condition (2.9) is a restriction for the choice of β_k in the IRSQP-method, but it does not exclude the possibility of β_k tending to zero, since we may expect that $\lambda_k \to 0$ if the method converges to a solution of the identification problem.

So far we have not discussed the Lagrangian of the problem and the arising first-order optimality conditions, which are not only necessary but also sufficient under the assumptions of Propositions 2.1 and 2.2, respectively, since the objective functionals are strictly convex. We will treat this problem in the following section.

2.2 The Karush-Kuhn-Tucker System

Based on the standard theory of convex optimization, we can formulate the Lagrangian of the problems (2.3), (2.4) and (2.6), (2.7) as

$$\begin{aligned}\mathcal{L}_k(u,q;\lambda) &= \frac{1}{2}\|Eu-z\|_Z^2 + \frac{\beta_k}{2}\|q-q_k\|_Q^2 + \frac{\eta}{2}\langle \lambda_k, e''(u_k,q_k)(u-u_k,q-q_k)^2\rangle \\ &\quad + \langle \lambda, e'(u_k,q_k)(u-u_k,q-q_k) + e(u_k,q_k) - f\rangle, \end{aligned} \tag{2.10}$$

where $\eta = 1$ in the case of the IRSQP-method and $\eta = 0$ for the LMSQP-method. The solutions $(\overline{u}_k, \overline{q}_k, \overline{\lambda}_k)$ of the quadratic programming problems are saddle points of the Lagrangian $\mathcal{L}_k$ (cf. [44, p. 392ff]), i.e.,

$$\mathcal{L}_k(\overline{u}_k,\overline{q}_k,\lambda) \le \mathcal{L}_k(\overline{u}_k,\overline{q}_k,\overline{\lambda}_k) \le \mathcal{L}_k(u,q,\overline{\lambda}_k), \qquad \forall\ (u,q,\lambda) \in X \times Q \times Y^*, \tag{2.11}$$

and satisfy the optimality condition

$$0 = \mathcal{L}_k'(\overline{u}_k,\overline{q}_k,\overline{\lambda}_k), \tag{2.12}$$

where $\mathcal{L}_k'$ denotes the Frèchet-derivative of $\mathcal{L}_k$ in $X \times Q \times Y^*$.

In order to rewrite (2.12) as a linear system for (u,q,λ), the so-called *Karush-Kuhn-Tucker system*, we define the following operators

$$K_k : X \to Y, \qquad K_k u = e_u(u_k,q_k)u, \quad \forall\ u \in X \tag{2.13}$$

$$L_k : Q \to Y, \qquad L_k q = e_q(u_k,q_k)q, \quad \forall\ q \in X \tag{2.14}$$

$$M_k : X \to X^*, \qquad \langle M_k u, v\rangle = \langle e_{uu}(u_k,q_k)(u,v), \lambda_k\rangle, \quad \forall\ (u,v) \in X^2 \tag{2.15}$$

$$N_k : Q \to Q^*, \qquad \langle N_k q, s\rangle = \langle e_{qq}(u_k,q_k)(q,s), \lambda_k\rangle, \quad \forall\ (q,s) \in Q^2 \tag{2.16}$$

$$P_k : X \to Q^*, \qquad \langle P_k u, q\rangle = \langle e_{qu}(u_k,q_k)(q,u), \lambda_k\rangle, \quad \forall\ (u,q) \in X \times Q. \tag{2.17}$$

Using these operators and the notation I_Q for the identity on Q, we may conclude that $(u_{k+1}-u_k, q_{k+1}-q_k, \lambda_{k+1})$ solves the linear system

$$\begin{pmatrix} E^*E + \eta M_k & \eta P_k^* & K_k^* \\ \eta P_k & \beta_k I_Q + \eta N_k & L_k^* \\ K_k & L_k & 0 \end{pmatrix} \begin{pmatrix} u \\ q \\ \lambda \end{pmatrix} = \begin{pmatrix} E^*(z^\delta - Eu_k) \\ 0 \\ f - e(u_k,q_k) \end{pmatrix}. \tag{2.18}$$

Note that assumption (2.8) implies that K_k is a regular operator, while L_k is not necessarily invertible.

Finally, we analyze the Karush-Kuhn-Tucker-system (2.18) in the framework of linear saddle-point problems as defined in Section 1.4. For this sake we define the symmetric bilinear form a_k on $(X \times Q)^2$ by

$$a_k(u,q;\varphi,\sigma) := \langle Eu, E\varphi\rangle_Z + \beta_k\langle q,\sigma\rangle_Q + \eta\big(\langle \varphi, M_k u\rangle + \langle \sigma, N_k q + P_k u\rangle + \langle q, P_k\varphi\rangle\big) \tag{2.19}$$

and the bilinear form $b_k : (X \times Q) \times Y^* \to \mathbb{R}$ by

$$b_k(u,q;\lambda) := \langle K_k u, \lambda\rangle + \langle L_k q, \lambda\rangle. \tag{2.20}$$

With the right-hand sides

$$f_k \ := \ f - e(u_k,q_k) \qquad \in Y, \tag{2.21}$$

$$g_k \ := \ (E^*(z^\delta - Eu_k), 0) \qquad \in X^* \times Q, \tag{2.22}$$

we can now rewrite the system (2.18) in the standard form

$$a_k(u,q;\varphi,\sigma) + b_k(\varphi,\sigma;\lambda) = \langle g_k,(\varphi,\sigma)\rangle, \qquad \forall\ (\varphi,\sigma)\in X\times Q, \tag{2.23}$$
$$b_k(u,q;\mu) = \langle f_k,\mu\rangle, \qquad \forall\ \mu\in Y^*. \tag{2.24}$$

Based on the theory introduced in Section 1.4 we can derive a statement on the well-posedness of the linear saddle-point problem (2.23), (2.24):

Theorem 2.3. *Suppose that the assumptions of Proposition 2.1 are satisfied if* $\eta = 0$ *in* (2.19) *and that the assumptions of Proposition 2.2 are satisfied if* $\eta = 1$ *in* (2.19), *respectively. Then the indefinite system* (2.23), (2.24), *with the bilinear forms* a_k *and* b_k *defined via* (2.19), (2.20), *has a unique solution* $(u,q,\lambda)\in X\times Q\times Y^*$, *which depends continuously on the right-hand sides* f_k *and* g_k.

Proof. We first show the kernel-ellipticity (1.21) of a_k. Suppose $(u,q)\in\mathcal{N}_{b_k}$, then $u = -K_k^{-1}L_kq$ and thus, with $\eta=0$ or $\eta=1$ and (2.9), we may deduce that

$$\begin{aligned} a_k(u,q;u,q) &\geq \beta_k\|q\|_Q^2 - \eta\|e''(u_k,q_k)\|\|\lambda_k\|(\|u\|^2+\|q\|^2) \\ &\geq \left(\frac{\beta_k}{2}\min\left\{1,\|K_k^{-1}L_k\|^{-2}\right\} - \eta\|e''(u_k,q_k)\|\|\lambda_k\|\right)(\|u\|^2+\|q\|^2) \\ &\geq \epsilon(\|u\|^2+\|q\|^2) \end{aligned}$$

for some $\epsilon>0$. The LBB-condition (1.22) for b_k follows from

$$\inf_{\lambda\in Y^*}\sup_{(u,q)\in X\times Q}\frac{b_k(u,q;\lambda)}{\|(u,q)\|\|\lambda\|} \geq \inf_{\lambda\in Y^*}\frac{b_k(K_k^{-1}\lambda,0;\lambda)}{\|K_k^{-1}\lambda\|\|\lambda\|} = \inf_{\lambda\in Y^*}\frac{\|\lambda\|^2}{\|K_k^{-1}\lambda\|\|\lambda\|} \geq \frac{1}{\|K_k^{-1}\|}.$$

Since the continuity of a_k and b_k follows from the continuity of the Frèchet-derivatives, Theorem 1.1 implies the assertion. ☐

2.3 The Iteration in the Parameter Space

In the following we consider the behavior of the LMSQP-iteration in the parameter space, i.e., after elimination of the state u_{k+1} and the Lagrange parameter λ_{k+1}, which is possible because of the regularity of K_k (see (2.8)). For a better distinction, we denote the updates of the LMSQP-method by superscript SQP and those of the classical Levenberg-Marquardt method following the feasible path by superscript FP.

The updates u^{SQP} and λ^{SQP} in the LMSQP-method can be computed consecutively from q^{SQP} via

$$u^{SQP} = -K_k^{-1}\left(L_kq^{SQP} - f + e(u_k,q_k)\right) \tag{2.25}$$
$$\lambda^{SQP} = -(K_k^*)^{-1}E^*\left(Eu^{SQP} - z + Eu_k\right) \tag{2.26}$$

Thus, with the notation $G_k := -EK_k^{-1}L_k$, we may rewrite the optimality condition for the update q^{SQP} as

$$(\beta_kI_Q + G_k^*G_k)\ q^{SQP} = G_k^*(z - Eu_k - EK_k^{-1}(f - e(u_k,q_k))). \tag{2.27}$$

In order to compare the LMSQP-method with the classical Levenberg-Marquardt method on feasible paths, we assume that $(u_0, q_0) \in X \times Q$ solves (1.2). Then we have

$$F'(q_0)s = -EK_0^{-1}L_0 s = G_0 s, \tag{2.28}$$

and an iteration step q^{FP} for the feasible-path LM-method solves

$$(\beta_0 I_Q + G_0^* G_0)\ q^{FP} = G_0^*(z - Eu_0), \tag{2.29}$$

which coincides with (2.27) in our particular case, i.e., the iterates q_1 computed with the LMSQP-method is the same as with the LM-method on the feasible path. The difference of our SQP approach to the classical method following the feasible path occurs in the second step of the iteration, since u_1 is not on the feasible path anymore. If e' is Lipschitz-continuous, we only have

$$\|u_1^{SQP} - u_1^{FP}\|_X = \mathcal{O}(\|u_1^{FP} - u_0\|_X^2 + \|q_1^{SQP} - q_0\|_Q^2). \tag{2.30}$$

In general, this causes the operator G_k to be different from $F'(q_k)$ during the iteration and in addition, the right-hand side differs from the one for a feasible path approach, since $f - e(u_k, q_k)$ need not vanish.

3 Convergence Analysis

In the following we will investigate the convergence behavior and the regularizing properties of the LMSQP-Method, i.e., we discuss the behavior of the iterates as $k \to \infty$ as well as the choice of an appropriate stopping index $k_* = k_*(\delta, z^\delta)$ in presence of data noise. We prove some preliminary properties of the iterative method under rather general conditions, while the final convergence analysis will be carried out under further restrictions on the structure of the state equation, which, however, still includes most important applications. We think that an abstract convergence analysis similar to the one for the Levenberg-Marquardt method by Hanke [20] could be carried out in principle, but to our impression the variety of technical conditions, which are needed for such an analysis, would shadow the basic principles. The problem-adapted convergence analysis for a specific structure corresponds very well to the objective of this paper, namely to construct a problem-adapted method for parameter identification problems.

If we denote by e_k and f_k the error terms

$$(e_k, f_k) := (u_k - \hat{u}, q_k - \hat{q}), \tag{3.1}$$

we can rewrite the Karush-Kuhn-Tucker system (2.18) with $\eta = 0$ as

$$\begin{pmatrix} E^*E & 0 & K_k^* \\ 0 & \beta_k I_Q & L_k^* \\ K_k & L_k & 0 \end{pmatrix} \begin{pmatrix} e_{k+1} \\ f_{k+1} \\ \lambda_{k+1} \end{pmatrix} = \begin{pmatrix} E^*(z^\delta - z) \\ \beta_k(q_k - \hat{q}) \\ r_k \end{pmatrix}, \tag{3.2}$$

where the r_k denotes the remainder

$$r_k := e(\hat{u}, \hat{q}) - e(u_k, q_k) + e'(u_k, q_k)(e_k, f_k). \tag{3.3}$$

These notations will be used without further notice in the analysis carried out in the subsequent sections.

3.1 Preliminary Results

As for all other iterative regularization methods (cf. [17]), we will need a condition on the nonlinearity, which allows to prove convergence results; the condition we use will be summarized in the following:

Assumption 1. Let (2.8) be satisfied for all $(u, q) \in X \times B_{2\rho}(q_0)$ and let $(\hat{u}, \hat{q}) \in B_\zeta(u_0) \times B_\rho(q_0)$ be a solution of the parameter identification problem (1.1), (1.2). Defining the remainder $r(u, q)$ by

$$r(u,q) \quad := \quad e(\hat{u},\hat{q}) - e(u,q) - e'(u,q)(\hat{u}-u,\hat{q}-q), \tag{3.4}$$

we assume that there exists a constant $\gamma_1 < 1$ such that

$$\|Ee_u(u,q)^{-1} r(u,q)\|_Z \leq \gamma_1 \|Eu - z\|_Z, \qquad \forall\ (u,q) \in X \times B_{2\rho}(q_0). \tag{3.5}$$

The first fundamental property of an iterative regularization method, namely the well-posedness of each iteration step has been verified in Section 2, and therefore it only remains to construct a strategy for the choice of a stopping criterion in presence of noise, i.e., for $\delta > 0$. For the choice of the stopping index k_* depending on the noise level δ and the noisy data z^δ, we adapt the generalized discrepancy principle (1.12), which can now be formulated as

$$\|Eu_{k_*} - z^\delta\|_Z \leq \tau\delta < \|Eu_k - z^\delta\|_Z, \qquad \forall\ k < k_*. \tag{3.6}$$

For an appropriate choice of τ, this allows us to prove the following monotonicity property of the iterates:

Lemma 3.1. *Let Assumption 1 be fulfilled, let the noise be bounded by* (1.4)*, and assume that*

$$\beta_0^{-1}(\|Ee_0\|_Z - \delta - \epsilon_h)^2 + \|f_0\|_Q^2 \leq \rho^2. \tag{3.7}$$

In addition, β_k is chosen such that $\beta_k \leq \beta_{k-1}$ for all $k \in \mathbb{N}$ and that

$$\overline{\gamma}_1 := \gamma_1 \sup_{k\in\mathbb{N}} \sqrt{\frac{\beta_{k-1}}{\beta_k}} < 1, \tag{3.8}$$

and the stopping index k_* is chosen according to the generalized discrepancy principle (3.6) *with*

$$\tau > 1 + \frac{\gamma_1 + \overline{\gamma}_1}{\gamma_1(1-\overline{\gamma}_1)}, \tag{3.9}$$

then $q_k \in B_{2\rho}(q_0)$ and the estimates

$$(\|Ee_{k+1}\|_Z - \delta)^2 + \beta_k\|f_{k+1}\|_Q^2 + \beta_k\|q_{k+1} - q_k\|_Q^2 \leq (\gamma_1\|Ee_k\|_Z + \delta)^2 + \beta_k\|f_k\|_Q^2 \tag{3.10}$$

and

$$\beta_k^{-1}(\|Ee_{k+1}\|_Z - \delta)^2 + \|f_{k+1}\|_Q^2 \leq \beta_{k-1}^{-1}(\|Ee_k\|_Z - \delta)^2 + \|f_k\|_Q^2 \tag{3.11}$$

hold for all $k < k_*$.

Proof. Assume that $q_k \in B_{2\rho}(q_0)$. Then, with (3.2) and

$$\lambda_{k+1} = -(K_k^*)^{-1} E^*(Eu_{k+1} - z^\delta),$$

we deduce the identity

$$\begin{aligned} 2\|Ee_{k+1}\|_Z^2 &+ \beta_k \|f_{k+1}\|_Q^2 + \beta_k \|q_{k+1} - q_k\|_Q^2 \\ &= 2\|Ee_{k+1}\|_Z^2 + \beta_k \|f_k\|_Q^2 + 2\beta_k \langle f_{k+1}, q_{k+1} - q_k \rangle \\ &= 2\langle z^\delta - z, Ee_{k+1}\rangle_Z + \beta_k \|f_k\|_Q^2 + 2\langle Eu_{k+1} - z^\delta, EK_k^{-1} r(u_k, q_k)\rangle_Z. \end{aligned}$$

Using (1.4) and (3.5) we obtain the estimate

$$\begin{aligned} \|Ee_{k+1}\|_Z^2 - 2\delta \|Ee_{k+1}\|_Z + \beta_k \|f_{k+1}\|_Q^2 + \beta_k \|q_{k+1} - q_k\|_Q^2 \\ \leq 2\gamma_1 \delta \|Ee_k\|_Z + \gamma_1^2 \|Ee_k\|_Z^2 + \beta_k \|f_k\|_Q^2, \end{aligned}$$

which implies (3.10) by adding δ^2 on both sides. (3.11) follows from dividing (3.10) by β_k and the fact that

$$\sqrt{\frac{\beta_{k-1}}{\beta_k}}\,(\gamma_1 \|Ee_k\|_Z + \delta) \leq \overline{\gamma_1} \|Ee_k\|_Z + \frac{\overline{\gamma_1}}{\gamma_1}\delta \leq \|Ee_k\|_Z - \delta.$$

for $k < k_*$ and τ satisfying (3.9).

Finally, an inductive proof using (3.11) and (3.7) shows that $q_k \in B_{2\rho}(q_0)$ and the above estimates hold for all $k < k_*$. □

In Lemma 3.1, we have deduced a property of the iterates for all $k < k_*$. A reasonable choice of the stopping index is expected to yield a finite index k_* if the noise level is positive. For the generalized discrepancy principle we shall show in the following Lemma that this property holds:

Lemma 3.2. *Under the assumptions of Lemma 3.1, the discrepancy principle yields a finite stopping index $k_*(\delta, z^\delta)$ if $\delta > 0$ and τ is chosen according to* (3.9).

Proof. Noticing that $\overline{\gamma_1} \geq \gamma_1$ under the assumptions of Lemma 3.1, we may conclude from (3.10) that there exists a positive real number ϵ such that

$$\epsilon\delta^2 + \beta_k^{-1}(\overline{\gamma}_1 \|Ee_{k+1}\|_Z + \delta)^2 + \|f_{k+1}\|_Q^2 \leq \beta_{k-1}^{-1}(\overline{\gamma}_1 \|Ee_k\|_Z + \delta)^2 + \|f_k\|_Q^2,$$

if (3.9) holds. Hence, we may conclude that

$$k\epsilon\delta^2 \leq \beta_0^{-1}(\overline{\gamma}_1 \|Ee_0\|_Z + \delta)^2 + \|f_0\|_Q^2,$$

for all $k < k_*$, which implies the finiteness of k_*. □

3.2 Convergence for Exact Data

Now we turn our attention to the convergence analysis in the case of exact data, i.e., for $\delta = 0$. In this case we will show that the pair (q_k, u_k) converges to a solution of the parameter identification problem as $k \to \infty$.

We start with summing up the assumptions on e needed for our analysis:

Assumption 2. In addition to Assumption 1, assume that e is of the form

$$e(u,q) = A(u) + N(u,q), \qquad \forall\ (u,q) \in X \times Q, \tag{3.12}$$

with continuously Frèchet-differentiable (nonlinear) operators $A : X \to Y$ and $N : X \times Q \to Y$, such that

$$N(u,.) \in \mathcal{L}(Q,Y), \qquad \forall\ u \in X. \tag{3.13}$$

Moreover, we assume that A and N satisfy the nonlinearity conditions

$$\|Ee_u(u,q)^{-1}A'(v)w\|_Y \leq \gamma_2\|Ew\|_Y, \quad \forall\ (u,v,w,q) \in B_{2\zeta}(u_0)^2 \times X \times B_{2\rho}(q_0), \tag{3.14}$$

and

$$\|Ee_u(u,q)^{-1}N_u(v,s)w\|_Y \leq \gamma_3\|Ew\|_Y, \quad \forall\ (u,v,w,q,s) \in B_{2\zeta}(u_0)^2 \times X \times B_{2\rho}(q_0)^2, \tag{3.15}$$

for some positive constants γ_2 and γ_3.

Lemma 3.3. *Let Assumption 2 be satisfied. Then the estimates*

$$\begin{aligned} \|Ee_u(u,q)^{-1}(A(v) - A(w))\|_Y &\leq \gamma_2\|E(v-w)\|_Y, & (3.16)\\ \|Ee_u(u,q)^{-1}(N(v,s) - N(w,s))\|_Y &\leq \gamma_3\|E(v-w)\|_Y, & (3.17) \end{aligned}$$

hold for all $(u,v,w,q,s) \in B_{2\zeta}(u_0)^3 \times B_{2\rho}(q_0)^2$.

Proof. Since A is continuously Frèchet-differentiable, we obtain with (3.14) and the monotonicity of integration that

$$\begin{aligned} \|Ee_u(u,q)^{-1}(A(v) - A(w))\|_Y &\leq \int_0^1 \|Ee_u(u,q)^{-1}(A'(w + t(v-w))(v-w)\|_Y\ dt\\ &\leq \int_0^1 \gamma_2\|E(v-w)\|_Y\ dt = \gamma_2\|E(v-w)\|_Y. \end{aligned}$$

The second estimate (3.17) can be deduced from (3.15) in an analogous way. □

Theorem 3.4 (Convergence for Exact Data). *Let Assumption 2 and* (3.7) *be fulfilled. Furthermore, let* $\delta = 0$ *and let* β_k *be chosen such that* $\beta_k \leq \beta_0$ *for all* $k \in \mathbb{N}$ *and that* (3.8) *is satisfied. Then the sequence of iterates* $(u_k, q_k) \in X \times Q$ *obtained with the LMSQP-method converges to a solution* $(\overline{u}, \overline{q})$ *of* (1.2) *with* $E\overline{u} = z$, *if* ρ *and* ζ *are sufficiently small.*

Proof. For $\delta = 0$, (3.10) implies that

$$R_{k+1} + \|f_{k+1}\|_Q^2 + \|q_{k+1} - q_k\|_Q^2 \leq \overline{\gamma}R_k + \|f_k\|_Q^2, \qquad \forall\ k \in \mathbb{N}_0,$$

with $R_k := \beta_{k-1}^{-1}\|Ee_k\|_Z^2$ for $k \geq 0$ and $\beta_1 = \beta_0$. By induction one can show that $q_k \in B_{2\rho}(q_0)$ for all $k \in \mathbb{N}$ and that

$$\sum_{j=0}^{\infty} \left(\frac{1}{\beta_{j-1}}\|Ee_j\|_Z^2 + \|q_{j+1} - q_j\|_Q^2 \right) < \infty. \tag{3.18}$$

As a direct consequence we obtain the convergence of $Eu_k \to z$ and $q_{k+1} - q_k \to 0$ as $k \to \infty$.

The next step is to show the strong convergence of $q_k \to \overline{q}$, which is equivalent to proving that f_k is a Cauchy-sequence. For this sake let $m > k$ be arbitrary and let $m \geq \ell \geq k$ be such that

$$\|Eu_\ell - z\|_Z + \|Eu_{\ell-1} - z\|_Z \leq \|Eu_j - z\|_Z + \|Eu_{j-1} - z\|_Z, \qquad \forall\, j \in \{k, \ldots, m\}. \quad (3.19)$$

Then we have

$$\|f_m - f_k\| \leq \|f_m - f_\ell\| + \|f_\ell - f_k\|$$

and

$$\begin{aligned} \|f_m - f_\ell\|^2 &= 2\langle f_\ell - f_m, f_\ell\rangle + \|f_m\|^2 - \|f_\ell\|^2 \\ \|f_\ell - f_k\| &= 2\langle f_\ell - f_k, f_\ell\rangle + \|f_k\|^2 - \|f_\ell\|^2. \end{aligned}$$

Since

$$\|f_k\|^2 - \|f_j\|^2 \leq R_j - R_k - \|q_k - q_{k-1}\|^2 \to 0$$

for $k > j$, we conclude that $\|f_m\|^2 - \|f_\ell\|^2$ and $\|f_k\|^2 - \|f_\ell\|^2$ converge to zero as $k, m \to \infty$. Thus, it suffices to show that the scalar products $\langle f_\ell - f_m, f_\ell\rangle$ and $\langle f_\ell - f_k, f_\ell\rangle$ converge to zero. For this sake we use (3.2) to obtain

$$\begin{aligned} |\langle f_\ell - f_k, f_\ell\rangle| &= |\sum_{j=k}^{\ell-1}\langle q_{j+1} - q_j, f_\ell\rangle| = |\sum_{j=k}^{\ell-1}\beta_j^{-1}\langle L_j f_\ell, \lambda_{j+1}\rangle| \\ &\leq \sum_{j=k}^{\ell-1}\beta_j^{-1}\left(|\langle L_{\ell-1}q_\ell - L\hat{q}, \lambda_{j+1}\rangle| + |\langle (L - L_j)\hat{q}, \lambda_{j+1}\rangle| + |\langle (L_{\ell-1} - L_j)q_\ell, \lambda_{j+1}\rangle|\right) \\ &= (I) + (II) + (III) \end{aligned}$$

The three parts of the sum on the right-hand side can be estimated using (3.2), (3.19) and Assumption 2, which leads to

$$\begin{aligned} &\sum_{j=k}^{\ell-1}\beta_j^{-1}|\langle L_{\ell-1}q_\ell - L\hat{q}, \lambda_{j+1}\rangle| \\ &= \sum_{j=k}^{\ell-1}\beta_j^{-1}|\langle EK_j^{-1}(A(\hat{u}) - A(u_{\ell-1}) + A'(u_{\ell-1})(e_{\ell-1} - e_\ell) - N_u(u_{\ell-1}, q_{\ell-1})(u_\ell - u_{\ell-1})), Ee_{j+1}\rangle| \\ &\leq \sum_{j=k}^{\ell-1}\beta_j^{-1}\left((\gamma_2 + \gamma_3)\|Ee_\ell\|_Z + (2\gamma_2 + \gamma_3)\|Ee_{\ell-1}\|_Z\right)\|Ee_{j+1}\|_Z \\ &\leq (4\gamma_2 + 2\gamma_3)\sum_{j=k}^{\ell}\beta_j^{-1}\|Ee_j\|_Z^2 \end{aligned}$$

and

$$\begin{aligned} \sum_{j=k}^{\ell-1}\beta_j^{-1}|\langle (L - L_j)\hat{q}, \lambda_{j+1}\rangle| &= \sum_{j=k}^{\ell-1}\beta_j^{-1}|\langle Ee_u(u_j, q_j)^{-1}(N(\hat{u}, \hat{q}) - N(u_j, \hat{q})), Ee_{j+1}\rangle| \\ &\leq \gamma_3\sum_{j=k}^{\ell}\beta_j^{-1}\|Ee_j\|_Z^2 \end{aligned}$$

$$\begin{aligned}\sum_{j=k}^{\ell-1}\beta_j^{-1}|\langle(L_{\ell-1}-L_j)q_\ell,\lambda_{j+1}\rangle| &= \sum_{j=k}^{\ell-1}\beta_j^{-1}|\langle Ee_u(u_j,q_j)^{-1}(N(u_{\ell-1},q_\ell)-N(u_j,q_\ell)),Ee_{j+1}\rangle| \\ &\le \gamma_3\sum_{j=k}^{\ell-1}\beta_j^{-1}(\|Ee_{\ell-1}\|_Z+\|Ee_j\|_Z)\|Ee_{j+1}\|_Z \\ &\le 3\gamma_3\sum_{j=k}^{\ell}\beta_j^{-1}\|Ee_j\|_Z^2.\end{aligned}$$

Because of the boundedness of $\sum_{j=0}^{\infty}\beta_j^{-1}\|Ee_j\|_Z^2$, all three terms converge to zero as $k\to\infty$ and thus, $\langle f_\ell - f_k, f_\ell\rangle \to 0$. In an analogous way we can show that $\langle f_\ell - f_m, f_\ell\rangle \to 0$ and hence, (e_k) is a Cauchy-sequence.

Let $\overline{q}$ denote the limit of q_k and let $\overline{u}$ be the corresponding solution of (1.2) with $q=\overline{q}$, which lies in $B_{2\zeta}(u_0)$ if ρ and ζ are sufficiently small. Then we may conclude that

$$u_{k+1}-\overline{u} = e_u(u_k,q_k)^{-1}\left(e(\overline{u},\overline{q})-e(u_k,q_k)-e'(u_k,q_k)(\overline{u}-u_k,\overline{q}-q_k)\right),$$

and with the regularity of e_u and the continuous Frèchet-differentiability of A and N we obtain that

$$\|u_{k+1}-\overline{u}\|_X = o(\|u_k-\overline{u}\|_X),$$

and hence, $u_k\to\overline{u}$ for ζ and ρ sufficiently small. Finally, $E\overline{u}=z$ follows from $Eu_k\to Ez$. □

We want to mention that the proof of Theorem 3.4 implies that the residual decays at least with

$$\|Eu_k - z\|_Z = o(\sqrt{\beta_k}), \tag{3.20}$$

while the convergence of (u_k,q_k) may be arbitrarily slow, which is a well-known effect for ill-posed problems.

3.3 Convergence for Noisy Data

Now we turn our attention to the case of noisy data, i.e., $\delta>0$. In order to compare the perturbed iterations for different noisy data, we will use the notation (u_k^δ,q_k^δ) for the iterations obtained with specific noisy data z^δ in the following.

Theorem 3.5 (Convergence for Noisy Data). *Let Assumption 2 and* (3.7) *be fulfilled with ζ, ρ sufficiently small, and let the noise be bounded by* (1.4). *Moreover, let β_k be chosen such that $\beta_k\le\beta_0$ for all $k\in\mathbb{N}$ and that* (3.8) *is satisfied. If the perturbed iteration is stopped with $k_*=k_*(\delta,z^\delta)$ according to the generalized discrepancy principle* (3.6) *with τ satisfying* (3.9)*, then*

$$(q^\delta_{k_*(\delta,z^\delta)},u^\delta_{k_*(\delta,z^\delta)})\to(\overline{q},\overline{u}),\qquad \text{in } X\times Q,\quad \text{as } \delta\to 0, \tag{3.21}$$

where $(\overline{u},\overline{q})$ is a solution of (1.2) *with $E\overline{u}=z$.*

Proof. Now let $\delta_n\to 0$ and denote by k_n the stopping index $k_*(\delta_n,z^{\delta_n})$ according to (3.6). Assuming that k is a finite accumulation point of the sequence $\{k_n\}$ we can show that this implies $q^{\delta_n}_{k_n-1}\to\hat{q}$ in the same way as in the proof of Theorem 2.4 in [21]. If $k_n\to\infty$, we may assume without loss of generality that k_n increases monotonically with n and for $k_n > m$, (3.11) implies

$$\begin{aligned} \|q_{k_n}^{\delta_n} - \overline{q}\|_Q^2 &\leq \|q_m^{\delta_n} - \overline{q}\|_Q^2 + \beta_{m-1}^{-1}(\|Eu_m^{\delta_n} - z\|_Z - \delta_n)^2 \\ &\leq 2\|q_m^{\delta_n} - q_m\|_Q^2 + 2\|q_m - \hat{q}\|_Q^2 + 2\beta_{m-1}^{-1}(\|Eu_m^{\delta_n} - Eu_m\|_Z^2 + \|Eu_m - z\|_Z^2). \end{aligned}$$

For each $\epsilon > 0$ we can now choose m sufficiently large such that

$$4\|q_m - \hat{q}\|_Q^2 + 4\beta_{m-1}^{-1}\|Eu_m - z\|_Z^2 \leq \epsilon^2$$

and the continuous dependence of the iterates on δ implies that (with m now fixed)

$$4\|q_m^{\delta_n} - q_m\|_Q^2 + 4\beta_{m-1}^{-1}\|Eu_m^{\delta_n} - Eu_m\|_Z^2 \leq \epsilon^2$$

for δ_n sufficiently small. Hence, we obtain that

$$\limsup_{n\to\infty} \|q_{k_n}^{\delta_n} - \overline{q}\|_Q \leq \epsilon,$$

and since ϵ is arbitrary, we may conclude the convergence of $q_{k_n}^{\delta_n}$ to $\overline{q}$ as $n \to \infty$. The convergence $u_{k_n}^{\delta_n} \to \overline{u}$ in X can be shown similarly as in the proof of Theorem 3.4. □

4 Applications

In this section we present four typical classes of identification problems, to which the SQP-type methods presented above can be applied. In particular, our aim is to show that all assumptions needed for the convergence of the LMSQP method are fulfilled, which implies that the convergence results deduced in the previous section are applicable.

4.1 Inverse Source Problems

In mathematical terms, an inverse source problem can be interpreted as the identification of a right-hand side, a boundary condition or an initial condition in a differential equation (cf. [22]). We will slightly generalize this notion to a problem of the form

$$e(u,q) = A(u) + Cq, \qquad \forall\ (u,q) \in X \times Q, \tag{4.1}$$

where the nonlinear operator $A : X \to Y$ is continuously Frèchet-differentiable and $C \in \mathcal{L}(Q,Y)$. In this case we have $L_k = C$ for all $k \in \mathbb{N}$, the nonlinearity condition (3.5) reads

$$\|EA'(u)^{-1}(A(\hat{u}) - A(u) - A'(u)(\hat{u} - u))\|_Y \leq \gamma_1\|Eu - z\|_Y, \qquad \forall\ u \in B_\zeta(\hat{u}), \tag{4.2}$$

and conditions (3.14), (3.15) reduce to

$$\|EA'(u)^{-1}A'(v)w)\|_Y \leq \gamma_2\|Ew\|_Y, \qquad \forall\ (u,v,w) \in B_\zeta(\hat{u})^2 \times X, \tag{4.3}$$

for some constant $\gamma_2 \in \mathbb{R}^+$.

Obviously, all nonlinearity conditions are satisfied if A is a regular linear operator (cf. [18] for an overview), but also for typical nonlinearities in partial differential equations, e.g. for a typical nonlinear heat conduction model of the form

$$e(u,q) = -\text{div}\ (a(u)\nabla u) - q, \tag{4.4}$$

with $q \in Q = L^2(\Omega)$, $u \in X = H_0^1(\Omega)$.

4.2 Identification of Conductivities

The identification of a conductivity (or piezoelectric head in groundwater filtration) from distributed measurements denotes the problem of reconstructing $q \in H^d(\Omega)$ from an observation of the state $z^\delta \in L^2(\Omega)$, where the exact observation z solves the elliptic boundary value problem.

$$-\text{ div }(q\nabla u) = f \qquad \text{in } \Omega, \tag{4.5}$$

$$u = g \qquad \text{on } \partial\Omega. \tag{4.6}$$

The solution of this problem has been investigated by several authors (cf. e.g. [2, 9, 24, 20, 21, 27, 36, 43]) using a variety of regularization methods. Variants of this problem are the reconstruction of q in the corresponding Neumann or Robin problem, or the transient case of identifying $q \in H^d(\Omega)$ in the parabolic initial-boundary problem

$$\frac{\partial u}{\partial t} - \text{ div }(q\nabla u) = f \qquad \text{in } \Omega \times (0,T), \tag{4.7}$$

$$u = g \qquad \text{on } \partial\Omega \times (0,T), \tag{4.8}$$

$$u = h \qquad \text{in } \Omega \times \{0\}. \tag{4.9}$$

Another interesting case is the identification of a nonlinearity of the form $q = q(u)$ in (4.5) or (4.7), which has been investigated recently (cf. [10, 15, 34]).

For the sake of simplicity we restrict ourselves to (4.5), (4.6) with $g = 0$ and $\partial\Omega$ sufficiently smooth, as discussed e.g. in [36]. This problem can be written in the abstract form (1.2) with $X = H_0^1(\Omega)$, $Y = X^* = H^{-1}(\Omega)$, $Q = H^d(\Omega)$ and

$$e(u,q) = -\text{ div }(q\nabla u). \tag{4.10}$$

The set of admissible parameters is given by

$$Q_{ad} = \{\ q \in H^d(\Omega) \mid q \geq \epsilon \text{ a.e. in } \Omega\ \},$$

for given constant $\epsilon > 0$, and the operator E can be defined as the embedding operator from $H_0^1(\Omega)$ into the output space $Z = L^2(\Omega)$.

The regularity of the derivative $e_u(u,q) = -\text{div }(q\nabla *)$ for each $q \in Q_{ad}$ is guaranteed by the standard theory of weak solutions for elliptic differential equations, noticing that the embedding $H^d(\Omega) \hookrightarrow L^\infty(\Omega)$ is continuous for $d = 1,2,3$. In order to verify the nonlinearity condition (3.5), let $v(\varphi) = (e_u(u,q)^*)^{-1}E^*\varphi$ for $\varphi \in L^2(\Omega)$, i.e., $v \in H_0^1(\Omega)$ is the solution of

$$-\text{ div }(q\nabla v) = \varphi.$$

Then we may compute

$$\begin{aligned} \|Ee_u(u,q)^{-1}r(u,q)\|_Y &= \sup_{\varphi\in Z} \frac{\langle (e_u(u,q)^*)^{-1}E^*\varphi, r(u,q)\rangle_Y}{\|\varphi\|_Z} \\ &= \sup_{\varphi\in Z} \frac{\langle -\text{div }((\hat{q}-q)\nabla v(\varphi)), \hat{u}-u\rangle_Z}{\|\varphi\|_Z}. \end{aligned}$$

The standard elliptic regularity theory shows that there exists a positive constant c_1 such that $\|v(\varphi)\|_{H^2(\Omega)} \leq c_2\|\varphi\|_{L^2\Omega}$ and embedding results yield

$$\begin{aligned}\|Ee_u(u,q)^{-1}r(u,q)\|_Y &\leq c_2\frac{\|\hat{q}-q\|_Q\|\hat{u}-u\|_Y\|v(\varphi)\|_{H^2(\Omega)}}{\|\varphi\|_Z}\\ &\leq c_1c_2\rho\|\hat{u}-u\|_Y,\end{aligned}$$

i.e., (3.5) holds for $\rho c_1 c_2 < 1$. The nonlinearity conditions (3.14) and (3.15) can be verified in an analogous way. Hence, if $\hat{q}$ is in the interior of Q_{ad}, we obtain local convergence of the LMSQP-method. If $\hat{q}$ not in the interior of the admissible set, local convergence of the LMSQP-method can be enforced by incorporating the non-negativity condition $q_{k+1} \geq \epsilon$ into the quadratic programming problems to be solved in each step of the LMSQP-method.

4.3 Identification of Potentials and Convections

A typical model problem for elliptic parameter identification problems is the identification of the potential q in

$$-\Delta u + qu = f \qquad \text{in } \Omega, \tag{4.11}$$

$$u = g \qquad \text{on } \partial\Omega. \tag{4.12}$$

As for the conductivity, different parameter models such as $q = q(x)$ or $q = q(u)$ have been investigated in this context as well as the corresponding parabolic problem (cf. e.g. [2, 13, 21, 27, 38]). The problem setup is analogous to the identification of the conductivity, with only modifying the equation operator e according to (4.11). The nonlinearity conditions needed for the convergence statements can be shown similarly to Section 4.2, but in this case no elliptic regularity is needed due to the fact that the nonlinearity is related to a lower order term in the state equation.

A related problem is the identification of the convection in an elliptic equation (or the corresponding parabolic problem) of the form

$$-\Delta u + q.\nabla u = f, \qquad \text{in } \Omega, \tag{4.13}$$

with appropriate boundary conditions on $\partial\Omega$. Here we have X, Y, Z, E as above and

$$Q = \{\ q \in H^1(\Omega)^d \mid \text{div}\ q = 0\ \},$$

which guarantees the well-posedness of the state equation and the identifiability of the parameter (cf. [26]).

4.4 Inverse Robin Problems

An inverse Robin problem is concerned with the identification of the transfer coefficient in a third-type boundary condition of the form

$$\frac{\partial u}{\partial \nu} + qu = g, \qquad \text{on } \partial\Omega, \tag{4.14}$$

where $\frac{\partial u}{\partial \nu}$ denotes the derivative of u in outward normal direction and g is a prescribed flux. The underlying state equation is either of elliptic or parabolic type. A typical measurement in such problems is the Dirichlet value of the solution u on $\partial\Omega$, i.e., the output operator E is the trace map into $Z = L^2(\partial\Omega)$. The spaces X, Y can be chosen as in the examples above, and the set of admissible parameters is given by

$$Q_{ad} = \{\ q \in H^{d-1}(\partial\Omega) \mid q \geq 0 \text{ a.e. on } \partial\Omega\ \} \tag{4.15}$$

or, if $q = q(u)$ (cf. [40]), by

$$Q_{ad} = \{\ q \in H^1([z_1, z_2])q \geq 0 \text{ a.e. in } [z_1, z_2] \mid ,\ \} \tag{4.16}$$

for an appropriate interval $[z_1, z_2]$.

For the discussion of the nonlinearity conditions we restrict our attention to the equation

$$-\Delta u = f \qquad \text{in } \Omega. \tag{4.17}$$

In this case, $v = e_u(u,q)^{-1}r(u,q)$ is the solution of the boundary value problem

$$-\Delta v = 0 \qquad \text{in } \Omega, \tag{4.18}$$

$$\frac{\partial v}{\partial \nu} + qv = (\hat{q} - q)(\hat{u} - u) \qquad \text{on } \partial\Omega. \tag{4.19}$$

The standard theory of elliptic boundary value problems shows that

$$\|Ev\|_{L^2(\partial\Omega)} \leq \|E\|\ \|v\|_{H^1(\Omega)} \leq \|E\|\ \|(\hat{q} - q)(\hat{u} - u)\|_{H^{-\frac{1}{2}}(\partial\Omega)},$$

and an embedding result yields (3.5). Similarly, one can show that the nonlinearity conditions (3.14) and (3.15) are fulfilled, and thus, the LMSQP-method is locally convergent for this problem.

5 Numerical Experiments

In order to test our theoretical results, we carry out some numerical experiments with a model problem for parameter identification, namely with the identification of a potential in an elliptic boundary-value problem. For details on the numerical approximation and realization of the LMSQP-method we refer to the forthcoming paper [8]. Example 5.1 was implemented in the software system MATLAB, while the implementation of Example 5.2 is based on the finite-element package FEPP [31], which has been developed at the Department for Computational Mathematics and Optimization of the University Linz.

Example 5.1. Our first example is the identification of the potential q in (4.11), (4.12) from a state observation $u \in L^2(\Omega)$, with $\Omega = (0,1)$, $g = 0$ and

$$f(x) = \frac{1}{2} + \sin x, \qquad x \in \Omega.$$

The exact potential is given by

$$q(x) = x(1-x),$$

which is an element of $Q = H^1(\Omega)$. The data are generated by solving the state equation on a fine grid and subsequent interpolation to a coarser grid; the noise is an additive high-frequency perturbation. We used uniform grids with $n = 1601$ nodes for the discretization

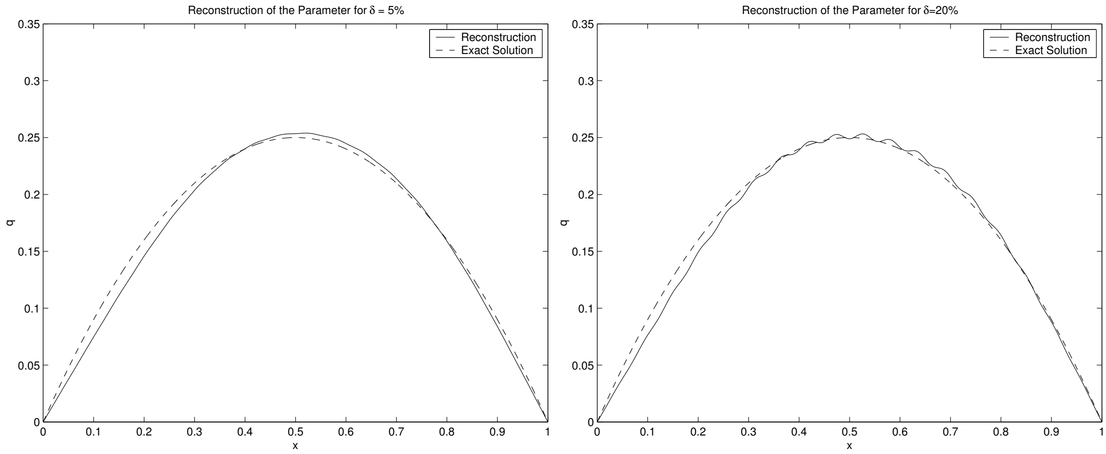

Figure 1: Reconstruction (solid) and exact solution (dashed) for noise level $\delta = 5\%$ (left) and $\delta = 20\%$ (right).

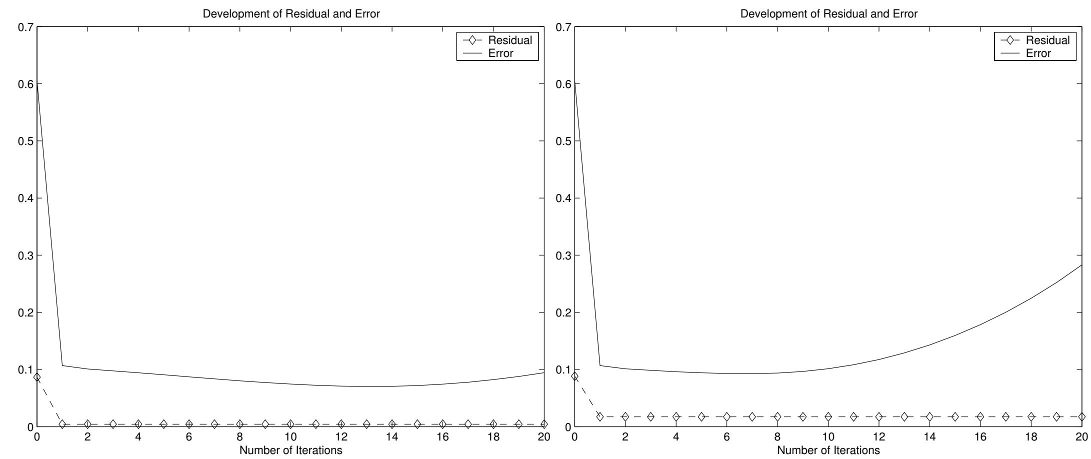

Figure 2: Development of the error $\|q_k - \hat{q}\|_{H^1(\Omega)}$ (solid) and the residual $\|u_k - z^\delta\|_{L^2(\Omega)}$ during the iteration for noise $\delta = 5\%$ (left) and $\delta = 20\%$ (right).

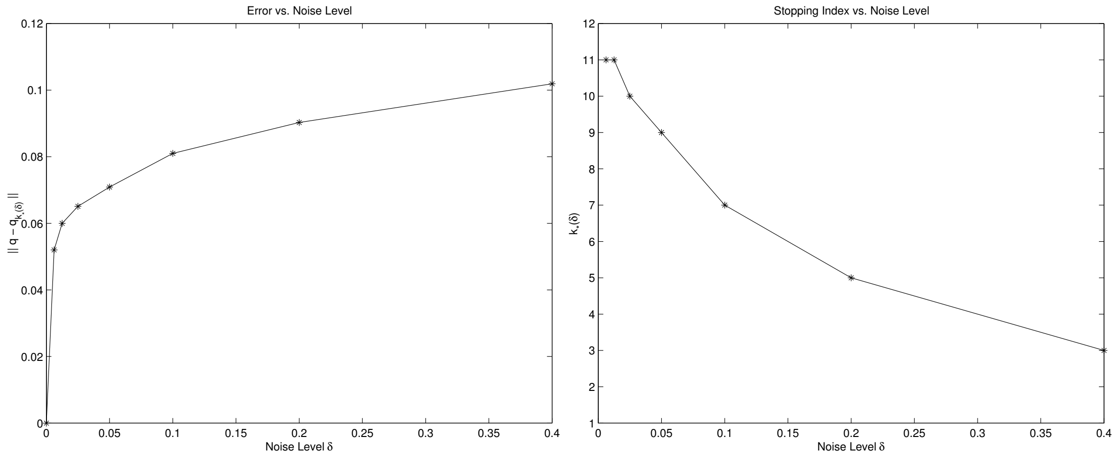

Figure 3: Error $\|q_{k_*(\delta)} - \hat{q}\|_{H^1(\Omega)}$ (left) and stopping index $k_*(\delta)$ plotted vs. the noise level δ.

of the state u and the Lagrange-parameter λ and $m = 401$ nodes for the parameter q. The parameters β_k are chosen according to $\beta_{k+1} = 0.9\beta_k$, with $\beta_0 = 10^{-6}$.

Figure 1 shows the results obtained with the LMSQP method for noise levels $\delta = 5\%$ and $\delta = 20\%$, where the discrepancy principle has been used as a stopping rule. Surprisingly, the approximation is still reasonable even for a large noise level like $\delta = 20\%$, but the reconstruction is not as smooth as for $\delta = 5\%$. The corresponding evolutions of the error $\|q_k - \hat{q}\|_{H^1(\Omega)}$ and the residual $\|u_k - z^\delta\|_{L^2(\Omega)}$ are plotted in Figure 2; one observes that in both cases the error decreases up to some iteration index and then starts to increase again, which numerically demonstrates the necessity of an appropriate stopping rule. We want to mention that the stopping index obtained from the discrepancy principle was always close to the iteration index, where the error is minimal. A comparison of the results for the two different noise levels also confirms the intuition that the instability is more pronounced if the noise level is larger, which results in an earlier and steeper ascent of the error during the iteration for larger δ.

The behavior as $\delta \to 0$ is illustrated in Figure 3; one observes that the error between the reconstruction $q_{k_*(\delta)}$ and the exact solution $\hat{q}$ is decreasing to zero as $\delta \to 0$, but the speed of convergence is slower than linear convergence with respect to δ, which is a typical effect for an ill-posed problem. The plot of the stopping index $k_*(\delta)$ obtained with the discrepancy principle numerically confirms that k_* increases as the noise level tends to zero. However, the absolute value of k_* is still low even for relatively small noise levels like 1%, which demonstrates the rather high convergence speed of the method. We want to mention that numerical tests with a Levenberg-Marquardt method on the feasible path and the IRSQP-method (both with the same choice of the parameters β_k and τ) produced exactly the same stopping indices and the same errors between the reconstruction and the exact solution up to the fourth digit, which numerically confirms our idea that the convergence behavior of the LMSQP-method (and the IRSQP-method) is similar to the one of a Levenberg-Marquardt method following the feasible path.

Finally, the results of a numerical test with exact data are shown in Figure 4. The left plot shows the development of the error $\|q_k - \hat{q}\|_{H^1(\Omega)}$ and the residual $\|u_k - z\|_{L^2(\Omega)}$ during the

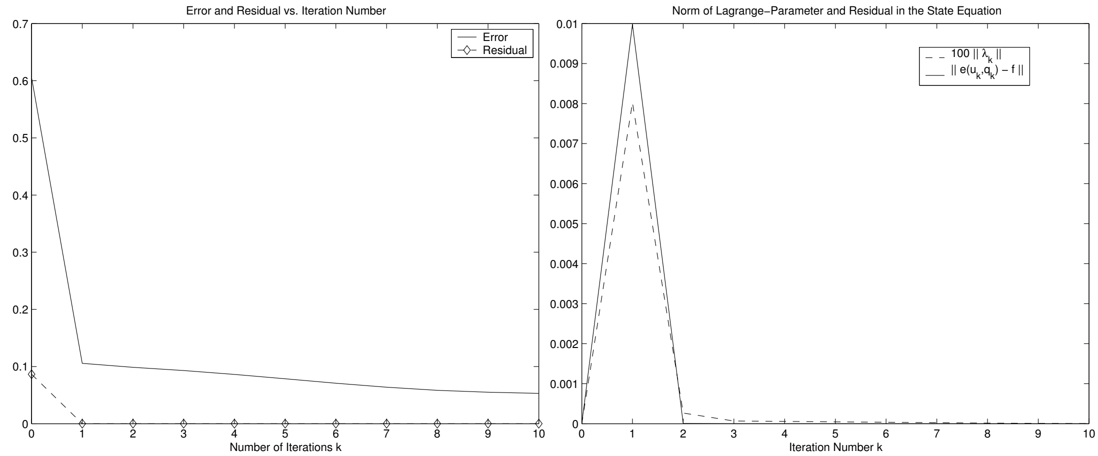

Figure 4: Error and residual (left) as well as the scaled norm of λ_k and the residual in the state equation (right) plotted vs. the iteration number k.

iteration. One observes that both are monotonically decreasing to zero, but the convergence speed in the parameter is rather slow, which has to be expected due to the ill-posedness of the problem. The right picture shows a plot of the scaled norm of the Lagrange parameter $(100\|\lambda_k\|)$ and of the residual in the state equation $(\|e(u_k, q_k) - f\|)$ vs. the iteration number. Both start at the value 0 at $k = 0$, since the initial Lagrange parameter is chosen as $\lambda_0 = 0$ and u_0 solves the state equation with parameter q_0. A significant deviation from zero occurs only in the first iteration step, whereas both norms decay to zero very fast in the consecutive steps, i.e., the deviation from the real equation constraint is very small.

Example 5.2. Our second numerical example is the identification of $q \in L^2(\Omega)$ in (4.11) and (4.12) from a state observation $u \in L^2(\Omega)$. The domain Ω is a ball in $\mathbb{R}^2$ with missing first quadrant, i.e., in radial coordinates

$$\Omega = \{\ (r\cos\theta, r\sin\theta) \mid r \in [0,1), \theta \in (\pi/2, 2\pi)\ \}. \tag{5.1}$$

The exact parameter to be reconstructed is $\hat{q} \equiv 1$, the right-hand side in (4.11) is given by

$$f = \frac{3\pi}{4}\big(3\pi\cos(\frac{3\pi}{2}r) + \frac{2}{r}\sin(\frac{3\pi}{2}r)\big) + \cos(\frac{3\pi}{2}r) + 3 \qquad \text{with } r = \sqrt{x^2+y^2}.$$

The corresponding solution of the state equation is $\hat{u} = \cos(\frac{3\pi}{2}r) + 3$. The data are generated using the exact solution $\hat{u}$ perturbed by uniformly distributed random noise. For the discretization we used triangular finite elements with piecewise quadratic shape functions for the state u and the Lagrange parameter λ and piecewise constant shape functions for the parameter q. The triangulation consisted of $n = 3065$ nodes and $m = 1472$ elements. We want to mention that this identification problem is quite challenging not only due to the complicated geometry, but also due to the fact that q is only identifiable in the interior and not along the boundary.

Figure 5 shows the parameter q_{k_*} reconstructed with the LMSQP-method for noise level $\delta = 1\%$ and the difference between the measured data and the observation corresponding to the reconstructed state u_{k_*}. It can be seen that along the boundary, where the parameter

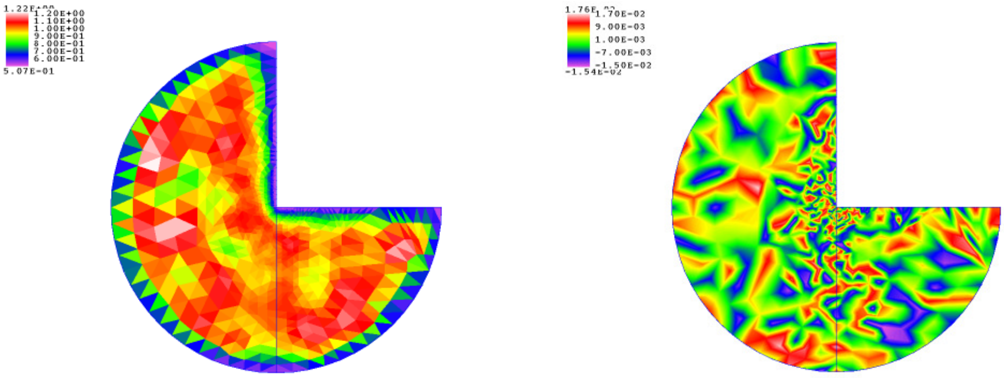

Figure 5: Reconstruction of the parameter and difference between reconstructed state and the noisy data for noise level $\delta = 1\%$.

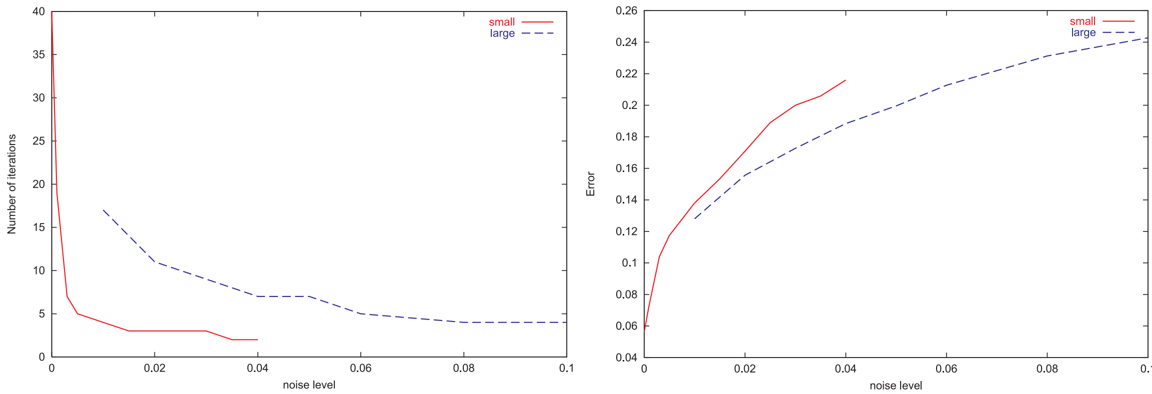

Figure 6: Stopping index $k_*(\delta)$ and error $\|q_{k_*(\delta)} - \hat{q}\|_{L^2(\Omega)}$ plotted vs. the noise level δ for two different regularization parameters.

can not be identified, the reconstructed values differ significantly from the exact parameter. However, the difference between exact and reconstructed parameter is very small in the interior. Therefore, we may consider the results as another indicator for the good quality of the reconstructions obtained with the LMSQP method.

The stopping indices obtained with the generalized discrepancy principle for various noise levels and two different choices of penalty parameters β_k are shown in Figure 6. It can be seen that for larger noise levels the instability is more pronounced and therefore larger penalty parameters are of advantage. Furthermore, from the behavior for $\delta \to 0$ one observes that the reconstruction $q_{k_*(\delta)}$ converges to the exact solution $\hat{q}$ as δ decreases to 0 but the convergence is rather slow.

Figure 7 shows the convergence of $\|u_k - \hat{u}\|_{L^2(\Omega)}$ and $\|q_k - \hat{q}\|_{L^2(\Omega)}$ for exact data. Both pictures show a rather fast decrease during the first iterations followed by a slower decrease during the latter ones. This effect is quite typical for such problems and caused by its ill-posedness.

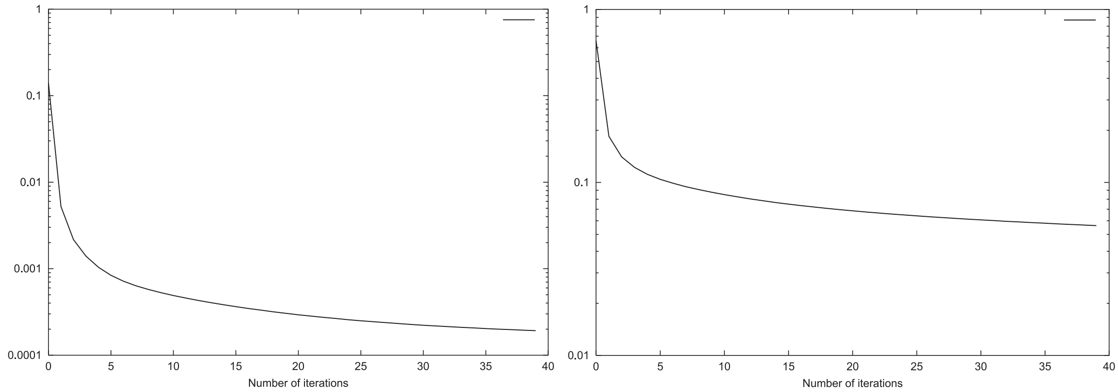

Figure 7: Development of $\|u_k - \hat{u}\|_{L^2(\Omega)}$ and $\|q_k - \hat{q}\|_{L^2(\Omega)}$ for exact data.

6 Conclusions and Outlook

We have presented two iterative SQP-type methods that can be applied to ill-posed parameter identification problems. It has been shown that the subproblems to be solved in each iteration step are well-posed and equivalent to an indefinite linear system that satisfies the standard regularity conditions. Moreover, we have analyzed the LMSQP-method with respect to convergence and verified its regularizing properties. The numerical comparison with standard feasible-path methods shows that the convergence behavior is similar, in particular the number of iterations needed is very small.

So far, we have not discussed in detail the numerical treatment of the SQP-type methods, which leaves more freedom than classical methods following the feasible path. In particular, the numerical approximation and solution of the Karush-Kuhn-Tucker system can be realized in different ways, e.g., by a reduced SQP-approach or by a simultaneous approach in the product space $X \times Q \times Y^*$. We shall show in a subsequent paper (cf. [8]) that the first is very similar to the feasible path approach with respect to its numerical treatment, while the second provides new possibilities for the design of efficient solution methods (cf. also [19]). This is an important advantage of SQP-methods compared to feasible-path methods and allows the numerical solution with fine discretization in reasonable time, which is not the case for classical approaches. We refer to [8] for further details on this subject.

References

[1] W.Alt, *The Lagrange-Newton method for infinite-dimensional optimization problems*, Numer. Funct. Anal. and Optimiz. **11** (1990), 201-224.

[2] H.T.Banks, K.Kunisch, *Estimation Techniques for Distributed Parameter Systems*, (Birkhäuser, Basel, Boston, Berlin,1989).

[3] A.Barclay, P.E.Gill, J.B.Rosen, *SQP methods and their application to numerical optimal control* in W.H.Schmidt, K.Heier, L.Bittner, R.Bulirsch, eds., *Variational Calculus, Optimal Control and Applications. Proceedings of the 12th international conference in honour of L.Bittner and R.Kloetzler* (Birkhäuser, Basel, 1998), 207-222.

[4] F.Brezzi, *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers*, R.A.I.R.O., R-2, **8** (1974), 129-151.

[5] F.Brezzi, M.Fortin, *Mixed and Hybrid Finite Element Methods*, Springer Series in Computational Mathematics **15** (Springer, New York, 1991).

[6] M.Burger, *Iterative regularization of a parameter identification problem occurring in polymer crystallization*, SIAM J. Numer. Anal. (2001), to appear.

[7] M.Burger, *Direct and Inverse Problems in Polymer Crystallization Processes* (PhD-Thesis, University Linz, 2000).

[8] M.Burger, W.Mühlhuber, *Numerical approximation of an SQP-type method for parameter identification problems*, Preprint (2001).

[9] G.Chavent, K.Kunisch, *On weakly nonlinear inverse problems*, SIAM J. Appl. Math. **56** (1996), 542-572.

[10] G.Chavent, P.Lemmonier, *Identification de la non-linearité d'une equation parabolique quasilineaire*, Appl. Math. and Optim. **1** (1974), 121-162.

[11] D.Colton, R.Ewing, W.Rundell (eds.), *Inverse Problems in Partial Differential Equations* (SIAM, Philadelphia, 1990).

[12] P.Deuflhard, H.W.Engl, O.Scherzer, *A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions*, Inverse Problems **14** (1998) 1081-1106.

[13] H.W.Engl, M.Hanke and A.Neubauer, *Regularization of Inverse Problems* (Kluwer, Dordrecht, 1996).

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

[15] H.W.Engl, P.Kügler, *Identification of a temperature dependent heat conductivity by Tikhonov regularization*, J. of Inverse and Ill-Posed Problems (2001), to appear.

[16] H.W.Engl, W.Rundell, eds., *Inverse Problems in Diffusion Processes. Proceedings of the GAMM-SIAM Symposium* (SIAM Proceedings, Philadelphia, 1995), 217-232.

[17] H.W.Engl, O.Scherzer, *Convergence rate results for iterative methods for solving nonlinear ill-posed problems*, to appear in: D.Colton, H.W.Engl, J.McLaughlin, A.Louis, W.Rundell, eds., *Solution Methods for Inverse Problems* (Springer, Vienna, New York, 2000).

[18] H.Engl, O.Scherzer, M.Yamamoto, *Uniqueness and stable determination of forcing terms in linear partial differential equations with overspecified boundary data*, Inverse Problems **10** (1994), 1253-1276.

[19] E.Haber, U.Ascher, D.Oldenburg, *On optimization techniques for solving nonlinear inverse problems*, Inverse Problems **16** (2000), 1263-1280.

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

[21] M.Hanke, A.Neubauer, O.Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math. **72** (1995), 21-37.

[22] V.Isakov, *Inverse Source Problems* (AMS, Providence, RI, 1990).

[23] V.Isakov, *Inverse Problems for Partial Differential Equations* (Springer, New York, 1998).

[24] K.Ito, K.Kunisch, *The augmented Lagrangian method for parameter estimation in elliptic systems*, SIAM J. Cont. Optim. **28** (1990), 113-136.

[25] K.Ito, K.Kunisch, *Augmented Lagrangian-SQP-methods in Hilbert spaces and application to control in the coefficients problems*, SIAM J. Optim. **6** (1996), 96-125.

[26] K.Ito, K.Kunisch, *Estimation of the convection in elliptic equations*, Inverse Problems **13** (1997), 995-1013.

[27] B.Kaltenbacher, *Some Newton-type methods for the regularization of nonlinear ill-posed problems*, Inverse Problems **13** (1997), 729-753.

[28] B.Kaltenbacher, *On Broyden's method for the regularization of nonlinear ill-posed problems*, Numer. Funct. Anal. and Optimiz. **19** (1998), 807-833.

[29] A.Kirsch, *An introduction to the mathematical theory of inverse problems*, Springer Series in Applied Mathematical Sciences **120** (Springer, Berlin , 1996).

[30] D.Kleis, E.W.Sachs, *Convergence rate of the augmented Lagrangian SQP method*, J. Optim. Theory Appl. **95** (1997), 49-74.

[31] M.Kuhn, U.Langer, J.Schöberl, *Scientific computing tools for 3D magnetic field problems*, in: J.R.Whiteman, ed., *The Mathematics of Finite Elements and Applications X* (Elsevier, Amsterdam, 2000), 239-258.

[32] K.Kunisch, E.W.Sachs, *Reduced SQP methods for parameter identification problems*, SIAM J. Numer. Anal. **29** (1992), 1793-1820.

[33] F.S.Kupfer, *An infinite-dimensional convergence theory for reduced SQP methods in Hilbert space*, SIAM J. Optim. **6** (1996), 126-163.

[34] P.Kügler, *Identification of a temperature-dependent heat conductivity by Tikhonov regularization* (Diploma Thesis, University of Linz, 2000).

[35] F.Leibfritz, E.W.Sachs, *Inexact SQP interior point methods and large scale optimal control problems*, SIAM J. Cont. Optim. **38** (1999), 272-293.

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

[37] A.Neubauer, O.Scherzer, *A convergence rate result for a steepest descent method and a minimal error method for the solution of nonlinear ill-posed problems*, Z. Anal. Anwend. **14** (1995), 369-377.

[38] D.E.Reeve, M.Spivack, *Determination of a source term in the linear diffusion equation*, Inverse Problems **10** (1994), 1335-1344.

[39] A.Rieder, *On the regularization of nonlinear ill-posed problems via inexact Newton iterations*, Inverse Problems **15** (1999), 309-327.

[40] A.Rösch, *Identification of nonlinear heat transfer laws by optimal control*, Numer. Funct. Anal. Optimization **15** (1994), 417-434.

[41] E.W.Sachs, *Control applications of reduced SQP methods*, in R.Bulirsch, D.Kraft, eds., *Computational Optimal Control. Proceedings of the 9th IFAC Workshop on Control Applications of Optimization* (Birkhäuser, Basel, 1994), 89-104.

[42] O.Scherzer, *Convergence criteria of iterative methods based on Landweber iteration for solving nonlinear problems*, J.Math.Anal.Appl. **194** (1995), 911-933.

[43] O.Scherzer, *A modified Landweber iteration for solving parameter estimation problems*, Appl. Math. Optim. **38**, 45-68 (1998).

[44] E.Zeidler, *Nonlinear Functional Analysis and its Applications, III: Variational Methods and Optimization* (Springer, New York, 1985).