SQP methods for parameter identification

Martin Burger*Wolfram Mühlhuber†

October 6, 2003

Abstract

The aim of this paper is to discuss sequential quadratic programming (SQP) methods as iterative regularization methods for ill-posed parameter identification problems. We focus on variant 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.

A Galerkin-type discretization of the resulting iteration algorithm, called *Levenberg-Marquardt SQP (LMSQP) method*, leads to a convergent approximation whose Karush-Kuhn-Tucker (KKT) system is well-posed, but indefinite. We discuss the simulataneous solution of the discretized KKT-system by preconditioned iteration methods for indefinite problems. From a discussion of the numerical effort we conclude that these approaches may lead to a considerable speed-up with respect to standard iterative regularization methods that eliminate the underlying state equation. The numerical efficiency of the LMSQP-method is confirmed by numerical examples.

Keywords: parameter identification, sequential quadratic programming, iterative regularization, ill-posed problems, saddle point problems, Galerkin methods, indefinite systems

1 Introduction

Since distributed parameters have to be determined from indirect measurements in many applications that are modeled by PDEs, parameter identification has become an important part of mathematical modeling. Such problems appear in many applications, where mathematical models of physical, chemical, biological or economical processes are used (cf. e.g. [1, 15, 20] 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 [17, 30] 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. [14, 19, 18]), which replace the least-squares problem by a close stable problem. Recently, also the application of *iterative regularization methods* has been investigated (cf. e.g. [24, 25, 28, 29, 37]), where the iteration followed the

*Department of Mathematics, UCLA, Los Angeles, USA. e-mail:martin.burger@jku.at
†Audi AG, Ingolstadt, Germany

feasible path defined by the underlying state equation. The main effort of these methods is the evaluation of the *parameter-to-output map*, i.e., the operator mapping the parameter to the corresponding observation as it involves the solution of the underlying state equation for given parameter, which is numerically realized e.g. by standard discretizations such as finite elements.

Our aim is discuss 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 important difference of SQP-type methods to the feasible path approach is that the underlying state equation is interpreted as a constraint in the product space of state and parameter variables and approximated by a linearized version in each iteration step. The formulation in the product space opens new possibilities with respect to the discretization and numerical implementation, in particular a product space discretization leads to a system with sparse matrices to be solved in each iteration step, while classical feasible path approaches use the parameter-to-output map, whose linearization and discretization leads to smaller, but dense matrices (cf. e.g. [23]).

The paper is organized as follows: In Section 2 we introduce the Levenberg-Marquardt SQP-method for parameter identification problems and review some of its basic properties. In Section 3 we investigate the numerical approximation of the LMSQP-method for an elliptic state problem by a Galerkin-type approach and discuss the well-posedness, stability and approximation properties of the discretized Karush-Kuhn-Tucker (KKT) system. Moreover, we present and analyze a multi-level approach, which leads to a further speed-up of the method. For an analysis of the convergence properties of the presented method as well as for proofs of the stated propositions we refer to [12, 13] for details. The central part of each iteration step is the numerical solution of the discretized KKT-system. As we use an iterative solver, we describe several ways of preconditioning the occurring equation systems in Section 4. Numerical experiments with typical model problems are presented in Section 5.

2 The SQP approach for parameter identification

2.1 Definitions and Notation

The basic setup of the identification problems treated in the paper is as follows: Given an observation

$$z = E\hat{u}, \tag{1}$$

where $E : U \to Z$ is a bounded linear operator and $\hat{u}$ denotes the exact state. We want to identify the parameter $q \in Q_{ad} \subset Q$ in an underlying equation

$$e(u, q) = f, \tag{2}$$

where $e : U \times Q \to Y$ 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 U and Y are appropriate Banach spaces. In addition, we assume that the operator e is homogeneous, i.e,

$$e(0, 0) = 0, \tag{3}$$

which is no restriction of generality, since for an arbitrary operator e we can transform (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 z. We assume that the observation error is bounded by

$$\|z - z^\delta\|_Z \leq \delta, \tag{4}$$

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

$$e(\hat{u}, \hat{q}) = f. \tag{5}$$

The pair $(\hat{u}, \hat{q})$ denotes an exact solution of the parameter identification problem (as introduced above).

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 $Eu = z^\delta$ does not always exist, and therefore one has to consider the corresponding normal equation respectively the least-squares problem

$$\frac{1}{2}\|Eu - z^\delta\|^2 \to \min_{(u,q)\in U\times Q} \tag{6}$$

subject to the equality constraint (2) and to $q \in Q_{ad}$. Since our main focus is the treatment of the state equation as equality constraint, we will omit the additional constraint $q \in Q_{ad}$ in the following which leads to

$$\begin{aligned} \frac{1}{2}\|Eu - z^\delta\|^2 &\to \min_{(u,q)\in U\times Q} \\ \text{subject to} \qquad e(u,q) &= f. \end{aligned} \tag{7}$$

2.2 SQP methods in the product space

Because of the ill-posedness of problem (7), a direct application of a standard SQP scheme is not possible, since a minimizer of the quadratic subproblems arising in each iteration step needs not exist and if one exists, it might not depend on the data in a stable way. Therefore we need to modify our SQP-type approach. Adding a stabilizing term to the objective functionals arising in each iteration step (similar to the classical Levenberg-Marquardt method) seems a good choice and leads to stable quadratic subproblems. This leads to

$$\frac{1}{2}\|Eu - z^\delta\|_Z^2 + \langle e''(u_k,q_k)(u-u_k, q-q_k)^2, \lambda_k\rangle + \frac{\beta_k}{2}\|q - q_k\|_Q^2 \to \min_{(u,q)\in U\times Q}, \tag{8}$$

subject to the linearized state equation as constraint as introduced in [12]. Since we are only interested in the case of attainable data here, the Lagrangian variable in the least-squares system (7) must be small close to a solution. Therefore we can neglect the second order derivatives of e, which leads to the following iterative regularization scheme originally proposed in [12]:

Method 1 (Levenberg-Marquardt SQP Method). *Let* $(u_0, q_0) \in U \times Q$ *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 *(LMSQP) method consists of the iteration procedure*

$$(u_{k+1}, q_{k+1}) = (\overline{u}_k, \overline{q}_k), \tag{9}$$

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

$$\frac{1}{2}\|Eu - z^\delta\|_Z^2 + \frac{\beta_k}{2}\|q - q_k\|_Q^2 \to \min_{(u,q)\in U\times Q}, \tag{10}$$

subject to the linear constraint

$$e(u_k, q_k) + e'(u_k, q_k)(u - u_k, q - q_k) = f. \tag{11}$$

An important issue of applying the above stated method is the question, whether the iteration procedure is well-defined. Besides the existence and uniqueness of minimizers of the quadratic programming problems, the stable dependence of the iterates on the previous iterates and on the data is of high interest. We will investigate these questions in the following subsection.

2.3 Well-posedness of the quadratic programming problems

In the following we will verify the well-posedness of the quadratic programming problem (10), (11) under reasonable assumptions on the state equation e. Besides that, we will also analyze the KKT-System of the problem in the frame of linear saddle point problems. In a latter part of this paper this approach will also be used for solving the occurring quadratic programming problems.

In typical applications, the equation (2), respectively its linearization, admits a unique solution with respect to the state, i.e.,

$$e_u(u, q)^{-1} : Y \to U \text{ exists and is a continuous linear operator for all } (u, q) \in U \times Q. \tag{12}$$

Under this assumption we can prove the well-posedness of Method 1 which is done in the following proposition:

Proposition 1. *Let* e *be continuously Fréchet-differentiable, let* (12) *hold and let* $\beta_k > 0$. *Then the quadratic programming problem* (10), (11) *has a unique solution* $(\overline{u}_k, \overline{q}_k) \in U \times Q$, *which is also the only local minimum.*

So far we have not discussed the Lagrangian of the problem and the arising first-order optimality conditions. These are not only necessary but also sufficient under the assumptions needed for showing the well-posedness of the quadratic programming problems, since the objective functionals are strictly convex.

2.4 The Karush-Kuhn-Tucker system

Based on the standard theory of convex optimization, we can formulate the Lagrangian of the problem (10), (11) as

$$\begin{aligned}\mathcal{L}_k(u, q; \lambda) = &\frac{1}{2}\|Eu - z^\delta\|_Z^2 + \frac{\beta_k}{2}\|q - q_k\|_Q^2 \\ &+ \langle \lambda, e'(u_k, q_k)(u - u_k, q - q_k) + e(u_k, q_k) - f\rangle,\end{aligned} \tag{13}$$

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. ZEIDLER [51, p. 392ff]), i.e.,

$$\mathcal{L}_k(\overline{u}_k, \overline{q}_k, \lambda) \leq \mathcal{L}_k(\overline{u}_k, \overline{q}_k, \overline{\lambda}_k) \leq \mathcal{L}_k(u, q, \overline{\lambda}_k), \qquad \forall\ (u, q, \lambda) \in U \times Q \times Y^*, \tag{14}$$

and satisfy the optimality condition

$$0 = \mathcal{L}_k'(\overline{u}_k, \overline{q}_k, \overline{\lambda}_k), \tag{15}$$

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

In order to rewrite (15) as a linear system for (u, q, λ), the so-called *Karush-Kuhn-Tucker system*, we define the following operators:

$$K_k : U \to Y, \qquad K_k u = e_u(u_k, q_k)u, \quad \forall\ u \in U, \tag{16}$$

$$L_k : Q \to Y, \qquad L_k q = e_q(u_k, q_k)q, \quad \forall\ q \in Q \tag{17}$$

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

$$\begin{pmatrix} E^*E & 0 & K_k^* \\ 0 & \beta_k I_Q & L_k^* \\ K_k & L_k & 0 \end{pmatrix} \begin{pmatrix} u \\ q \\ \lambda \end{pmatrix} = \begin{pmatrix} E^*(z^\delta - Eu_k) \\ 0 \\ f - e(u_k, q_k) \end{pmatrix}. \tag{18}$$

Note that assumption (12) implies that K_k is a regular operator, while L_k is not necessarily invertible.

In the last part of this subsection we want to analyze the Karush-Kuhn-Tucker system (18). As it forms a symmetric, and indefinite linear system of equations, the analysis can be done in the framework of linear saddle-point problems.

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 [10]. In the following let X and Λ be two Hilbert spaces, let $g \in X^*$, $f \in \Lambda^*$ and let $a : X \times X \to \mathbb{R}$ and $b : X \times \Lambda \to \mathbb{R}$ be continuous bilinear forms. Then a symmetric linear saddle-point problem in variational formulation consists of searching for a solution $(x, \lambda) \in X \times \Lambda$ of

$$a(x, v) + b(v, \lambda) = \langle g, v \rangle, \qquad \forall\ v \in X, \tag{19}$$

$$b(x, \mu) \qquad\quad = \langle f, \mu \rangle, \qquad \forall\ \mu \in \Lambda, \tag{20}$$

where $a(\cdot, \cdot)$ is supposed to be symmetric on $X \times X$. The well-posedness of (19), (20) can be studied under additional assumptions on a and b, namely the so-called *kernel-ellipticity* of a,

$$\exists\ \alpha_a \in \mathbb{R}^+ :\ a(v, v) \geq \alpha_a \|v\|_X^2, \qquad \forall\ v \in K^b := \{v \in X \mid b(v, \mu) = 0, \forall\ \mu \in \Lambda\}, \tag{21}$$

and the *LBB-condition* upon b,

$$\exists\ \alpha_b \in \mathbb{R}^+ :\ \inf_{\mu \in \Lambda} \sup_{v \in X} \frac{b(v, \mu)}{\|v\|_X \|\mu\|_\Lambda} \geq \alpha_b. \tag{22}$$

Under these assumptions, the linear saddle-point problem (19), (20) has a unique solution $(x, \lambda) \in X \times \Lambda$, which depends continuously on the data $(f, g) \in \Lambda^* \times X^*$ (cf. BREZZI [10] or BREZZI AND FORTIN [11]).

In order to transform our problem to the standard notation introduced above we define the symmetric bilinear form a_k on $(U \times Q) \times (U \times Q)$ by

$$a_k(u, q; \varphi, \sigma) := \langle Eu, E\varphi \rangle_Z + \beta_k \langle q, \sigma \rangle_Q \tag{23}$$

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

$$b_k(u, q; \lambda) := \langle K_k u, \lambda \rangle + \langle L_k q, \lambda \rangle. \tag{24}$$

With the right-hand sides

$$\begin{aligned} f_k &= f - e(u_k, q_k) && \in Y, \\ g_k &= (E^*(z^\delta - Eu_k), 0) && \in U^* \times Q^*, \end{aligned} \tag{25}$$

we can now rewrite the system (18) in the standard form

$$a_k(u, q; \varphi, \sigma) + b_k(\varphi, \sigma; \lambda) = \langle g_k, (\varphi, \sigma) \rangle, \qquad \forall\ (\varphi, \sigma) \in U \times Q, \tag{26}$$

$$b_k(u, q; \mu) = \langle f_k, \mu \rangle, \qquad \forall\ \mu \in Y^*. \tag{27}$$

Using the abstract theory of linear saddle point problems presented above, we can derive a statement on the well-posedness of the linear saddle-point problem (26), (27):

Theorem 2. *Under the assumptions of Proposition 1 the indefinite system* (26), (27), *with the bilinear forms a_k and b_k defined via* (23), (24), *has a unique solution $(u, q, \lambda) \in U \times Q \times Y^*$, which depends continuously on the right-hand sides $f_k \in Y$ and $g_k \in U^* \times Q^*$.*

3 Discretization techniques

In the following we investigate the discretization of the LMSQP-method by a Galerkin approach. First of all, we assume that we have discretized data $z^{\delta,\eta} \in Z_\eta \subset Z$ of the form

$$z^{\delta,\eta} = R_\eta z^\delta, \tag{28}$$

where $R_\eta : Z \to Z_\eta$ is the orthogonal projector onto the finite-dimensional subspace Z_η. Note that we can give an error estimate for $z^{\delta,\eta}$ using (4) and $\|R_\eta\| = 1$, which yields

$$\delta_\eta := \|R_\eta z^\delta - z\|_Z \le \|R_\eta (z^\delta - z)\|_Z + \|R_\eta z - z\|_Z \le \delta + \inf_{y \in Z_\eta} \|y - z\|_Z. \tag{29}$$

Additionally we assume that U is a Hilbert space and that the image space of e can be identified with the dual of U, for which reason we write U^* instead of Y in the following. Finally, we assume that e is continuously Frèchet-differentiable on $U \times Q$ and that the partial derivative e_u is self-adjoint and satisfies the coercivity condition

$$\langle e_u(u, q)v, v \rangle \ge \alpha_e \|v\|_U^2, \qquad \forall (u, q, v) \in U \times Q \times U, \tag{30}$$

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

The above setup is typical for a partial differential equation of elliptic type, which is also the main type of application we have in mind. We want to mention that the infinite-dimensional analysis carried out in the previous section was not restricted to elliptic problems, but only assumed well-posedness of the state equation for given parameter. However, since the numerical approximation techniques for elliptic problems differ from the ones for parabolic or hyperbolic problems (cf. e.g. QUARTERONI AND VALLI [40] for an overview), one cannot expect a successful unified approach to corresponding parameter identification problems. For this reason we restrict ourselves to the investigation of the elliptic case.

3.1 The discretized LMSQP method and its well-posedness

Now let $U_h \subset U$, $Q_h \subset Q$ be finite-dimensional subspaces of U and Q, with the corresponding orthogonal projectors $P_h : U \to U_h$ and $\tilde{P}_h : Q \to Q_h$. Then we can discretize the LMSQP-Method as follows:

Method 2 (Galerkin LMSQP-Method). *Let U_h, Q_h and Z_η be as above and let $(u_0, q_0) \in U_h \times Q_h$ be a given initial value. Moreover, let $(\beta_k)_{k\in\mathbb{N}}$ be a bounded sequence of positive real numbers. The* Galerkin Levenberg-Marquardt sequential quadratic programming *(GLMSQP) method consists of the iteration procedure*

$$(u_{k+1}, q_{k+1}) = (\overline{u}_k, \overline{q}_k), \tag{31}$$

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

$$\frac{1}{2}\|R_\eta(Eu - z^\delta)\|_Z^2 + \frac{\beta_k}{2}\|q - q_k\|_Q^2 \to \min_{(u,q)\in U_h\times Q_h}, \tag{32}$$

subject to the linear constraint

$$\langle e(u_k, q_k) + e'(u_k, q_k)(u - u_k, q - q_k), \varphi\rangle = \langle f, \varphi\rangle, \qquad \forall\ \varphi \in U_h. \tag{33}$$

Note that the constraint (33) can be rewritten in operator form as

$$P_h^* K_k P_h(u - u_k) + P_h^* L_k \tilde{P}_h(q - q_k) = P_h^*(f - e(u_k, q_k)), \tag{34}$$

to be solved for $(u, q) \in U_h \times Q_h$, with the notation

$$K_k : U \to U^*, \qquad K_k u = e_u(u_k, q_k)u, \quad \forall\ u \in U, \tag{35}$$

$$L_k : Q \to U^*, \qquad L_k q = e_q(u_k, q_k)q, \quad \forall\ q \in Q, \tag{36}$$

and $P_h^* : U_h^* \to U^*$ is the adjoint of P_h. Under the assumption (30), we obtain that

$$\langle P_h^* K_k P_h v, v\rangle = \langle K_k P_h v, P_h v\rangle = \langle K_k v, v\rangle \geq \alpha_e \|v\|_U^2 \tag{37}$$

for all $v \in U_h$, i.e., the discrete bilinear form associated with the operator $P_h^* K_k P_h$ is coercive on U_h. This implies by the Lax-Milgram theorem, that (34) is uniquely solvable with respect to u for given $q \in Q_h$. Consequently, in an analogous way to the proof of Proposition 1 we may show the following result on the well-posedness of the quadratic programming problem that has to be solved in each step of Method 2 (GLMSQP method).

Proposition 3. *Let e be continuously Frèchet-differentiable, let* (30) *hold and let $\beta_k > 0$. Then the quadratic programming problem* (32), (33) *has a unique solution $(\overline{u}_k, \overline{q}_k) \in U_h \times Q_h$, which is also the only local minimum.*

3.2 The discretized Karush-Kuhn-Tucker system

In Subsection 2.4, the Karush-Kuhn-Tucker system for the infinite-dimensional version of the LMSQP-method has been derived and analyzed in the framework of linear saddle point problems. Now we will discuss the discretized analogue of this system, namely the first-order optimality conditions for the quadratic programming problem (32), (33).

The Lagrangian of (32), (33) is given by

$$\begin{aligned}\mathcal{L}_k(u,q;\lambda) = \frac{1}{2}\|R_\eta(Eu - z^\delta)\|_Z^2 + \frac{\beta_k}{2}\|q - q_k\|_Q^2 + \\ + \langle \lambda, e'(u_k,q_k)(u-u_k, q-q_k) + e(u_k,q_k) - f\rangle,\end{aligned} \tag{38}$$

for $(u,q,\lambda) \in U_h \times Q_h \times U_h$. Since P_h and $\tilde{P}_h$ are equal to the identity on U_h and Q_h, respectively, we can rewrite the Lagrangian as

$$\begin{aligned}\mathcal{L}_k(u,q;\lambda) = \frac{1}{2}\|R_\eta(EP_hu - z^\delta)\|_Z^2 + \frac{\beta_k}{2}\|\tilde{P}_h(q - q_k)\|_Q^2 + \\ + \langle P_h\lambda, K_kP_h(u-u_k) + L_k\tilde{P}_h(q-q_k) + e(u_k,q_k) - f\rangle,\end{aligned} \tag{39}$$

with the operators K_k and L_k defined by (35), (36). The KKT-system can now be deduced by computing the partial derivatives of the Lagrangian with respect to u, q and λ, i.e., $(u_{k+1} - u_k, q_{k+1} - q_k, \lambda_{k+1})$ solves the linear saddle-point problem

$$\begin{pmatrix} P_h^*E^*R_\eta^*R_\eta EP_h & 0 & P_h^*K_k^*P_h \\ 0 & \beta_k\tilde{P}_h^*\tilde{P}_h & \tilde{P}_h^*L_k^*P_h \\ P_h^*K_kP_h & P_h^*L_k\tilde{P}_h & 0 \end{pmatrix} \begin{pmatrix} u \\ q \\ \lambda \end{pmatrix} = \begin{pmatrix} P_h^*E^*R_\eta^*R_\eta(z^\delta - Eu_k) \\ 0 \\ P_h^*(f - e(u_k,q_k)) \end{pmatrix}. \tag{40}$$

As in Subsection 2.3, we define the symmetric bilinear form $a_k : (U \times Q) \times (U \times Q) \to \mathbb{R}$ by

$$a_k^\eta(u,q;\varphi,\sigma) := \langle R_\eta Eu, R_\eta E\varphi\rangle_Z + \beta_k\langle q,\sigma\rangle_Q \tag{41}$$

and the bilinear form $b_k : (U \times Q) \times U \to \mathbb{R}$ by

$$b_k(u,q;\lambda) := \langle K_ku,\lambda\rangle + \langle L_ku,\lambda\rangle. \tag{42}$$

Moreover, we use the right-hand sides

$$f_k := f - e(u_k,q_k) \quad \in U^*, \tag{43}$$

$$g_k^\eta := (E^*R_\eta^*R_\eta(z^\delta - Eu_k), 0) \quad \in U^* \times Q. \tag{44}$$

Then the KKT-system (40) can be interpreted as the Galerkin approximation of an indefinite variational problem, i.e., $(u,q,\lambda) \in U_h \times Q_h \times U_h$ is the solution of

$$a_k^\eta(u,q;\varphi,\sigma) + b_k(\varphi,\sigma;\lambda) = \langle g_k^\eta, (\varphi,\sigma)\rangle, \qquad \forall\ (\varphi,\sigma) \in U_h \times Q_h, \tag{45}$$

$$b_k(u,q;\mu) = \langle f_k,\mu\rangle, \qquad \forall\ \mu \in U_h. \tag{46}$$

In an analogous way to the previous section we can show that the bilinear form a_k^η satisfies the discrete kernel-ellipticity condition on $U_h \times Q_h$, i.e., there exists a constant $\alpha_a > 0$ such that

$$a_k^\eta(u,q;u,q) \geq \alpha_a\|(u,q)\|^2, \quad \forall\ (u,q) \in \mathcal{K}_b^h$$

with

$$\mathcal{K}_b^h := \{(v,s) \in U_h \times Q_h \mid b(v,s;\lambda) = 0,\ \forall\ \lambda \in U_h\},$$

and that b satisfies the discrete LBB-condition

$$\inf_{\lambda \in U_h} \sup_{(u,q) \in U_h \times Q_h} \frac{b_k(u,q;\lambda)}{\|(u,q)\|\ \|\lambda\|} \geq \alpha_b,$$

for some $\alpha_b > 0$.This implies the following well-posedness result for the discretized problem (45), (46):

Theorem 4. *Let e be continuously Frèchet-differentiable, let* (30) *hold and let $\beta_k > 0$. Then the indefinite system* (45), (46) *has a unique solution $(u, q, \lambda) \in U_h \times Q_h \times U_h$, which depends continuously on the right-hand sides f_k and g_k^η.*

Since the constants α_a and α_b are the same as in the corresponding infinite-dimensional conditions in $U \times Q$, they are in particular independent of the discrete subspaces U_h and Q_h. This allows us to deduce an approximation result for the solutions of (45), (46) to the solution $(u, q, \lambda) \in U \times Q \times U$ of the infinite-dimensional KKT-System, given in variational form in (26), (27), i.e.

$$a_k(u, q; \varphi, \sigma) + b_k(\varphi, \sigma; \lambda) = \langle g_k, (\varphi, \sigma) \rangle, \qquad \forall\ (\varphi, \sigma) \in U \times Q, \tag{47}$$

$$b_k(u, q; \mu) = \langle f_k, \mu \rangle, \qquad \forall\ \mu \in U, \tag{48}$$

with a_k given by

$$a_k(u, q; \varphi, \sigma) := \langle Eu, E\varphi \rangle_Z + \beta_k \langle q, \sigma \rangle_Q, \tag{49}$$

b_k, f_k as above and g_k defined by

$$g_k := (E^*(z^\delta - Eu_k), 0) \in U^* \times Q. \tag{50}$$

Theorem 5. *Suppose that the assumptions of Theorem 4 are satisfied and let*

$$(u_h, q_h, \lambda_h) \in U_h \times Q_h \times U_h$$

denote the unique solution of (45), (46). *Then there exists a constant $c > 0$ independent of U_h and Q_h such that*

$$\|(u - u_h, q - q_h, \lambda - \lambda_h)\| \le c \left(r_{\eta,h}^\delta + \inf_{(v,s,\mu) \in U_h \times Q_h \times U_h} \|(u - v, q - s, \lambda - \mu)\| \right), \tag{51}$$

where (u, q, λ) denotes the unique solution of (26), (27) *and*

$$r_{\eta,h}^\delta := \|(R_\eta - I) z^\delta\|_Z + \sup_{v \in U_h, \|v\| = 1} \|(R_\eta - I) E v\|_Z. \tag{52}$$

Theorem 5 provides an error estimate for the solutions of the discretized saddle-point problem (45), (46), consisting of two parts corresponding to the numerical approximation in the image space Z and in the pre-image spaces U and Q. An obvious estimate for the first term is

$$r_{\eta,h}^\delta \le \inf_{y \in Z_\eta} \|y - z^\delta\|_Z + \sup_{v \in U_h, \|v\|_U = 1} \inf_{\tilde{y} \in Z_\eta} \|\tilde{y} - Ev\|_Z,$$

which possibly does not lead to a quantitative estimate, since there is no additional information on the smoothness of the noisy data. An alternative estimate is

$$r_{\eta,h}^\delta \le \delta + \inf_{y \in Z_\eta} \|y - z\|_Z + \sup_{v \in U_h, \|v\|_U = 1} \inf_{\tilde{y} \in Z_\eta} \|\tilde{y} - Ev\|_Z.$$

The infimum of $\|y - z\|_Z$ can usually be estimated more easily, since the exact data z are smoother due to the fact that $\hat{u}$ is the solution of the state equation for some parameter $\hat{q}$. E.g., if the state equation is of elliptic type with solution $\hat{u} \in H^1(\Omega)$, $E : H^1(\Omega) \to L^2(\Omega)$

is the embedding operator, and R_η results from a standard finite element discretization on a grid with fineness η, then we have at least

$$\inf_{y \in Z_\eta} \|y - z\| = \mathcal{O}(\eta).$$

Another important observation is that the last term vanishes if the discrete spaces Z_η and U_h are equal, which can be achieved in some applications.

The second term in (51) shows that the Galerkin approximation of the KKT-system is of optimal order in $U_h \times Q_h \times U_h$; it can be estimated by standard methods for finite element discretizations; quantitative estimates can be obtained using the regularity of the iterates. This part depends of course strongly on the specific application.

3.3 Nested multi-level optimization techniques

Important tools for the efficient numerical approximation of infinite-dimensional optimization problems are *multi-level optimization methods*. In the nested multi-level setup, one starts the optimization procedure at a coarse level $U_{h_1} \times Q_{h_1}$, where the iteration procedure can be carried out efficiently. If an appropriate stopping criterion is satisfied, one interpolates the state and parameter obtained in this way to a finer level $U_{h_2} \times Q_{h_2}$ (for $h_2 < h_1$), serving now as a starting value on this level. This procedure is repeated until the finest level is reached. Usually, nested spaces are used in this approach, i.e., $U_{h_1} \subset U_{h_2}$, $Q_{h_1} \subset Q_{h_2}$ (for $h_2 < h_1$), which leads to simple interpolation operators. Since one cannot choose the discretization of the data arbitrarily in general , we consider only the case of fixed η here, but a multi-level approach in η can be realized in an analogous way, if necessary.

Nested multi-level methods outperform standard discretization techniques in many cases (cf. e.g. HEINKENSCHLOSS [26], HEISE [27], LUKÁŠ [35, 36]); usually a considerable number of iterations is needed on the coarse level only, where the numerical effort per iteration is very low. On the finest levels, the stopping criterion is often satisfied already after one iteration step and so the overall effort is less than for a direct discretization on the finest level. For the Galerkin LMSQP method, this leads to Algorithm 1.

Algorithm 1 Nested Multi-Level Galerkin LMSQP

Require: a decreasing sequence $\{h_\ell\}_{\ell=1,\dots,L}$ with nested spaces $U_{h_\ell} \subset U_{h_{\ell+1}}$, $Q_{h_\ell} \subset Q_{h_{\ell+1}}$ (e.g. $h_\ell = 2^{-\ell} h_0$)
Require: $(u_0^1, q_0^1) \in U_{h_1} \times Q_{h_1}$
  for $\ell = 1$ to L **do**
    $h = h_\ell$
    Perform the Galerkin LMSQP method until the stopping criterion is satisfied.
    if $\ell == L$ **then**
      return
    end if
    Prolongate the iteration $(u_{k_*}^\ell, q_{k_*}^\ell)$ to the finer level $U_{h_{\ell+1}} \times Q_{h_{\ell+1}}$, which results in a new starting value $(u_0^{\ell+1}, q_0^{\ell+1})$.
  end for

Up to now we did not talk about the choice of the nested spaces. Of course, they can be chosen in advance. In the finite element community it is well known, that the accuracy of the

solution can be improved by using a-posteriori error estimators. They provide information which elements shall be refined to obtain a more precise solution. This information can be used to construct appropriate fine grid spaces. For an overview of a-posteriori error estimation see e.g. VERFÜRTH [50]. In the context of optimization the concept of adaptivity and a-posteriori error estimation is not as well-known as in the finite element community. An example in the context of optimization is presented in BECKER, KAPP, AND RANNACHER [5], in the context of optimal control problems see e.g. BECKER, KAPP, AND RANNACHER [4] or LIU AND YAN [34].

4 Numerical solution of the KKT-system

In the following we will discuss the numerical solution of the discretized KKT-system (40) for fixed iteration number k. We have seen above that the Galerkin-type approximation (40) of the original KKT-system is well-posed, now we discuss some of its structural properties, which are important for the application of iterative solution methods and for the construction of preconditioners.

Choosing bases

$$\Phi = (\phi_1, \ldots, \phi_m)^T \in U_h, \qquad \Sigma = (\sigma_1, \ldots, \sigma_n)^T \in Q_h, \tag{53}$$

of the finite-dimensional subspaces U_h and Q_h, we may represent $(u_h, q_h, \lambda_h) \in U_h \times Q_h \times U_h$ via

$$u_h = \mathbf{u}^T \Phi, \qquad q_h = \mathbf{q}^T \Sigma, \qquad \lambda_h = \boldsymbol{\lambda}^T \Phi, \tag{54}$$

with coordinate vectors $\mathbf{u}, \boldsymbol{\lambda} \in \mathbb{R}^m$ and $\mathbf{q} \in \mathbb{R}^n$. In order to transform (40) into a linear system for $\mathbf{u}$, $\mathbf{q}$ and $\boldsymbol{\lambda}$, we define the matrices

$$\mathbf{G} := (\langle E\phi_j, E\phi_i \rangle_Z)_{i,j=1,\ldots,m} \qquad \mathbf{H} := (\langle \sigma_j, \sigma_i \rangle_Q)_{i,j=1,\ldots,n} \tag{55}$$

$$\mathbf{K} := (\langle K_k \phi_j, \phi_i \rangle)_{i,j=1,\ldots,m} \qquad \mathbf{L} := (\langle L_k \sigma_j, \phi_i \rangle)_{i=1,\ldots,m;j=1,\ldots,n} \tag{56}$$

and the vectors

$$\mathbf{f}_1 := (\langle z^{\delta,\eta} - Eu_k, E\varphi_i \rangle_Z)_{i=1,\ldots,m}, \qquad \mathbf{f}_3 := (\langle f - e(u_k, q_k), \varphi_i \rangle)_{i=1,\ldots,m}. \tag{57}$$

This allows us to rewrite the discretized KKT-system (with penalty parameter $\beta = \beta_k$) as

$$\begin{pmatrix} \mathbf{G} & 0 & \mathbf{K}^T \\ 0 & \beta\mathbf{H} & \mathbf{L}^T \\ \mathbf{K} & \mathbf{L} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{q} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ 0 \\ \mathbf{f}_3 \end{pmatrix}, \tag{58}$$

respectively as

$$\mathbf{M}\,\mathbf{X} = \mathbf{F}, \tag{59}$$

with

$$\mathbf{M} = \begin{pmatrix} \mathbf{G} & 0 & \mathbf{K}^T \\ 0 & \beta\mathbf{H} & \mathbf{L}^T \\ \mathbf{K} & \mathbf{L} & 0 \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} \mathbf{u} \\ \mathbf{q} \\ \boldsymbol{\lambda} \end{pmatrix}, \qquad \mathbf{F} = \begin{pmatrix} \mathbf{f}_1 \\ 0 \\ \mathbf{f}_3 \end{pmatrix}.$$

The structural properties of $\mathbf{M}$ and its sub-matrices will be examined in the following subsection.

4.1 The system matrix M

Due to the well-posedness result of the discretized KKT-system (40) (cf. Theorem 4), we may conclude that the system matrix $\mathbf{M}$ is regular. In order to obtain further insight into the structure of $\mathbf{M}$, we investigate the properties of the sub-matrices $\mathbf{G}$, $\mathbf{H}$, $\mathbf{K}$ and $\mathbf{L}$.

Proposition 6. *The matrices $\mathbf{K} \in \mathbb{R}^{m\times m}$ and $\mathbf{H} \in \mathbb{R}^{n\times n}$ are symmetric positive definite, and the matrix $\mathbf{G} \in \mathbb{R}^{m\times m}$ is symmetric positive semi-definite. If in addition the operator E is injective on U_h, then $\mathbf{G}$ is regular, too.*

The matrix $\mathbf{L} \in \mathbb{R}^{m\times n}$ is difficult to analyze, it is neither symmetric nor regular in general (in particular if $n \neq m$). However, some fundamental properties of $\mathbf{M}$ (such as its regularity) rely rather on $\mathbf{G}$, $\mathbf{H}$ and $\mathbf{K}$ than on $\mathbf{L}$. Moreover, the classical splitting of a symmetric saddle-point problem as

$$\begin{pmatrix} \mathbf{G} & 0 & \mathbf{K}^T \\ 0 & \mathbf{H}_\beta & \mathbf{L}^T \\ \mathbf{K} & \mathbf{L} & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{I} & 0 \\ \mathbf{K}\mathbf{G}^{-1} & \mathbf{L}\mathbf{H}_\beta^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{G} & 0 & 0 \\ 0 & \mathbf{H}_\beta & 0 \\ 0 & 0 & -\mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{I} & 0 & \mathbf{G}^{-1}\mathbf{K}^T \\ 0 & \mathbf{I} & \mathbf{H}_\beta^{-1}\mathbf{L}^T \\ 0 & 0 & \mathbf{I} \end{pmatrix},$$

where $\mathbf{H}_\beta := \beta\mathbf{H}$ and $\mathbf{C}$ is the Schur-complement

$$\mathbf{C} := \mathbf{K}\mathbf{G}^{-1}\mathbf{K}^T + \beta^{-1}\mathbf{L}\mathbf{H}_\beta^{-1}\mathbf{L}^T, \tag{60}$$

is only possible if both $\mathbf{G}$ and $\mathbf{H}_\beta$ are regular. In particular, we may conclude that $\mathbf{M}$ has $n + m$ positive and m negative eigenvalues.

4.2 Reduced SQP approaches

The basic idea of reduced SQP-methods is the a-priori elimination of the equality constraint, which can be written in matrix form as

$$\mathbf{K}\mathbf{u} + \mathbf{L}\mathbf{q} = \mathbf{f}_3, \tag{61}$$

which is equivalent to an elimination of $\mathbf{u}$ and $\boldsymbol{\lambda}$ in (58).

Due to Proposition 6, $\mathbf{K}$ is a regular, symmetric matrix and thus, we may compute

$$\mathbf{u} = \mathbf{K}^{-1}(\mathbf{f}_3 - \mathbf{L}\mathbf{q}), \tag{62}$$

$$\boldsymbol{\lambda} = \mathbf{K}^{-T}(\mathbf{f}_1 - \mathbf{G}\mathbf{u}), \tag{63}$$

which yields after some calculations the $n \times n$-system

$$\mathbf{M}_r\mathbf{q} = \mathbf{g} \tag{64}$$

with

$$\mathbf{M}_r := \mathbf{H}_\beta + \mathbf{L}^T\mathbf{K}^{-T}\mathbf{G}\mathbf{K}^{-1}\mathbf{L} \tag{65}$$

$$\mathbf{g} := \mathbf{L}^T\mathbf{K}^{-T}(\mathbf{G}\mathbf{K}^{-1}\mathbf{f}_3 - \mathbf{f}_1). \tag{66}$$

The reduced SQP-approach seems of particular interest if $n \ll m$, which is a frequently used discretization strategy for parameter identification and optimal control problems (cf. e.g. Sachs [43], Schulz and Bock [45], or Schulz [44]). The original matrix $\mathbf{M}$ is an

indefinite matrix of size $(2m+n) \times (2m+n)$, while the reduced system matrix $\mathbf{M}_r$ in (64) is of size $n \times n$. However, $\mathbf{M}_r$ is not a sparse matrix even if all the sub-matrices of $\mathbf{M}$ are sparse, since it involves the inverse of $\mathbf{K}$. Moreover, the evaluation of a matrix-vector product using $\mathbf{M}_r$ is more expensive than a matrix-vector product using $\mathbf{M}$, since it involves the solution of two systems of the form

$$\mathbf{K}\mathbf{w} = \mathbf{g}, \tag{67}$$

with different right-hand sides $\mathbf{g}$, while for the evaluation of matrix-vector product with $\mathbf{M}$ only direct evaluations of $\mathbf{K}$ are needed, which are very cheap for typical finite element discretization of the state constraint. In practice, one usually tries to compensate this disadvantage of reduced SQP-methods by using a Broyden-type update for the reduced system matrix instead of the exact matrix $\mathbf{M}_r$, which leads to efficient optimization algorithms for small numbers of design parameters n.

4.3 Simultaneous solution of the KKT-system

Recently, the simultaneous solution of KKT-systems by iterative methods has been investigated, in particular in connection with optimal control problems (cf. Battermann and Heinkenschloss [2], Biros and Ghattas [6, 7] or Haber and Ascher [22]). Compared to the reduced SQP-approach, a simultaneous solution strategy has the obvious advantage that the allocation and evaluation of the system matrix $\mathbf{M}$ is much cheaper than of $\mathbf{M}_r$. The pay-off is that $\mathbf{M}$ is indefinite and larger than $\mathbf{M}_r$, which might cause additional effort. However, the main effort in the reduced SQP-approach is related to the evaluation or assembly of the system matrix $\mathbf{M}_r$, respectively, and therefore a simultaneous solution of the KKT-system can result in a tremendous speed-up of the SQP-method, in particular for fine discretizations.

At a first glance, it seems rather straight-forward to solve (59) by a standard iterative method for indefinite systems such as inexact Uzawa methods (cf. Bramble, Pasciak, and Vassilev [8], Elman and Golub [16], Langer and Queck [32, 33], or Queck [41]) or Krylov-subspace methods such as GMRES (cf. Saad and Schultz [42]), MINRES (cf. Paige and Saunders [39]) and QMR (cf. Freund and Nachtigal [21]). However, in the case of large-scale problems, we have to expect a large condition number (note that β is usually small and that $\mathbf{M}$ is singular for $\beta = 0$) and a complicated eigenvalue pattern of the matrix $\mathbf{M}$, which might cause iterative methods to diverge or to need a high number of iterations. Therefore, an appropriate preconditioning technique seems necessary for any of the methods.

In the following we distinguish two types of solvers that seem appropriate for the solution of the indefinite system (59) and discuss their basic properties with respect to the special structure of $\mathbf{M}$.

Inexact Uzawa iterations

Inexact Uzawa methods and similar iteration procedures have been developed for the solution of the classical Stokes system and similar problems (cf. Quarteroni and Valli [40] for an overview). The classical Uzawa method is just a gradient method for the dual of the corresponding Lagrange functional, the inexact Uzawa method can be interpreted as a preconditioned version (cf. Quarteroni and Valli [40]). Following the exposition by

ZULEHNER [52], we can write an inexact Uzawa method for a system of the form (58) as

$$\hat{\mathbf{A}} \begin{pmatrix} \mathbf{u}_{k+1} - \mathbf{u}_k \\ \mathbf{q}_{k+1} - \mathbf{q}_k \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 - \mathbf{G}\mathbf{u}_k - \mathbf{K}\boldsymbol{\lambda}_k \\ -\beta\mathbf{H}\mathbf{q}_k - \mathbf{L}\boldsymbol{\lambda}_k \end{pmatrix}, \tag{68}$$

followed by

$$\hat{\mathbf{C}}(\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k) = \mathbf{f}_3 - \mathbf{K}\mathbf{u}_{k+1} - \mathbf{L}\mathbf{q}_{k+1}, \tag{69}$$

where $\hat{\mathbf{A}}$ is a preconditioner for the diagonal matrix

$$\mathbf{A} := \begin{pmatrix} \mathbf{G} & 0 \\ 0 & \beta\mathbf{H} \end{pmatrix}, \tag{70}$$

$\hat{\mathbf{C}}$ is a preconditioner for the Schur-complement $\mathbf{C}$ defined by (60) and k denotes the iteration index. In terms of (59) we can write the inexact Uzawa iteration as

$$\mathbf{X}_{k+1} = (\mathbf{I} - \hat{\mathbf{M}}^{-1}\mathbf{M})\mathbf{X}_k + \hat{\mathbf{M}}^{-1}\mathbf{F}, s \tag{71}$$

where $\hat{\mathbf{M}}$ is a preconditioner for the system matrix, given by

$$\hat{\mathbf{M}} = \begin{pmatrix} \hat{\mathbf{A}} & 0 \\ \mathbf{B} & \hat{\mathbf{C}} \end{pmatrix}, \tag{72}$$

with $\mathbf{B} = \begin{pmatrix} \mathbf{K} & \mathbf{L} \end{pmatrix}$.

A convergence analysis of this method is available only in the case when $\mathbf{A}$ is a regular matrix (cf. BRAMBLE, PASCIAK, AND VASSILEV [8] or ZULEHNER [52]), which means that we have to assume that $\mathbf{G}$ is regular. The latter is true e.g. if the data z represent distributed data for the state, i.e., E is an embedding operator. In this case, the structure of $\mathbf{A}$ is rather simple and it is not a difficult task to construct a preconditioner, even exact preconditioning seems possible (note that $\mathbf{G}$ is just a mass matrix for a typical finite element discretization). Since the matrices $\mathbf{G}$ and $\mathbf{H}$ do not change during the SQP-iteration we may even compute decompositions in a preprocessing step. The construction of a preconditioner for the Schur-complement $\mathbf{C}$ is more difficult and must take into account the specific nature of the underlying state equation.

Krylov-subspace methods

The Krylov-subspace methods GMRES and QMR are variants of the CG-algorithm that are applicable to indefinite problems, too. The basic idea of such methods is a defect minimization in the Krylov-subspace

$$\mathcal{K}_k(\mathbf{M}; \mathbf{X}_1) = \quad \{\mathbf{X}_1, \mathbf{M}\mathbf{X}_1, \ldots, \mathbf{M}^{k-1}\mathbf{X}_1\}, \tag{73}$$

generated by $\mathbf{X}_1$, in the k-th iteration step. Since preconditioned CG-methods are probably the most successful class of iteration methods for positive definite systems, such methods seem very attractive also in the indefinite case, although additional difficulties may arise (cf. e.g. SAAD AND SCHULTZ [42]).

The convergence analysis in SAAD AND SCHULTZ [42] and FREUND AND NACHTIGAL [21] shows that the error bounds obtained for both methods are essentially the same, and mainly dependent on the eigenvalue distribution and the condition number of the system matrix $\mathbf{M}$. Therefore, appropriate preconditioning is again of high importance, in this case also with the possibility that $\mathbf{G}$ is singular.

Preconditioning

For the efficient solution of the KKT-system (58) it is necessary to use iterative solution procedures due to the size of the equation system. For these methods appropriate preconditioning strategies are needed to get fast convergence. Unfortunately, for symmetric indefinite equation systems by far fewer methods compared to the positive definite case are available.

The most popular class of methods are Uzawa type methods. Many publications can be found, especially in the field of fluid dynamics. Recently, ZULEHNER [52] presented a unified approach to many of these methods. The methodology presented in the previous subsection on inexact Uzawa methods can also be used as preconditioner.

A different class of preconditioners originates form reduced SQP methods and can be explained as follows: The KKT-matrix $\mathbf{M}$ can be factorized into

$$\mathbf{M} = \begin{pmatrix} \mathbf{G}\mathbf{K}^{-1} & 0 & \mathbf{I} \\ 0 & \mathbf{I} & \mathbf{L}^T\mathbf{K}^{-T} \\ \mathbf{I} & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{K} & \mathbf{L} & 0 \\ 0 & \mathbf{S_c} & 0 \\ 0 & -\mathbf{G}\mathbf{K}^{-1}\mathbf{L} & \mathbf{K}^T \end{pmatrix} \tag{74}$$

where $\mathbf{S_c}$ denotes the Schur-complement

$$\mathbf{S_c} = \beta\mathbf{H} + \mathbf{L}^T\mathbf{K}^{-T}\mathbf{G}\mathbf{K}^{-1}\mathbf{L}. \tag{75}$$

Replacing the matrix $\mathbf{K}^{-1}$ by a preconditioner $\hat{\mathbf{K}}^{-1}$ (e.g. a multigrid preconditioner) and the Schur complement $\mathbf{S_c}$ by an appropriate preconditioner $\hat{\mathbf{S}}_\mathbf{c}$ leads to a preconditioner for $\mathbf{M}$ of the form

$$\hat{\mathbf{M}} = \begin{pmatrix} \mathbf{G}\hat{\mathbf{K}}^{-1} & 0 & \mathbf{I} \\ 0 & \mathbf{I} & \mathbf{L}^T\hat{\mathbf{K}}^{-T} \\ \mathbf{I} & 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{K}} & \mathbf{L} & 0 \\ 0 & \hat{\mathbf{S}}_\mathbf{c} & 0 \\ 0 & -\mathbf{G}\hat{\mathbf{K}}^{-1}\mathbf{L} & \hat{\mathbf{K}}^T \end{pmatrix}. \tag{76}$$

It must be noted that for the preconditioning operation $\hat{\mathbf{M}}^{-1}$ only applications of $\hat{\mathbf{K}}^{-1}$ and $\hat{\mathbf{S}}_\mathbf{c}^{-1}$ are necessary and no applications of $\hat{\mathbf{K}}$ of $\hat{\mathbf{S}}_\mathbf{c}$. This preconditioner was used in our computations (see Section 5) , but also by HABER AND ASCHER [22] or BIROS AND GHATTAS [6, 7].

A similar preconditioner was presented by BATTERMANN AND SACHS [3]. They used

$$\hat{\mathbf{M}} = \begin{pmatrix} 0 & 0 & \hat{\mathbf{K}}^T \\ 0 & \hat{\mathbf{S}}_\mathbf{c} & \mathbf{L}^T \\ \hat{\mathbf{K}} & \mathbf{L} & 0 \end{pmatrix} \tag{77}$$

as a preconditioner for an all-at-once approach for an optimal control problem. Their paper contains also some analysis of the eigenvalue structure of the preconditioned system, which influences the convergence of the used iterative method to a large extent.

At least for elliptic state problems leading to positive definite matrices $\mathbf{K}$ it is clear how to choose appropriate preconditioners $\hat{\mathbf{K}}$ for the state equation for the previous two preconditioners. On the other hand, it is by far more difficult how to choose preconditioners for the Schur-complement. One approach is to exploit mapping properties of the underlying pseudo-differential operator. This approach was used e.g. by TA'ASAN [47, 49, 46, 48] in the context of shape design for fluid dynamics. He used Fourier transformation to get the symbol of the Schur-complement and exploited this for preconditioning. A completely different approach was presented by BRAMBLE, PASCIAK, AND VASSILEVSKI [9]. They developed a way for the construction of efficient preconditioners of pseudo-differential operators of positive and negative order, based on multi-level techniques.

5 Examples and numerical results

In order to illustrate the previously described methods, we carry out some numerical experiments with the problem described in Section 2. As the description there does not contain details on the observation as well as on the state equation, we will further restrict ourselves to two problem classes, namely the identification of a reaction coefficient and the identification of a conductivity.

5.1 The identification of a reaction coefficient

In our first example we want to identify the reaction coefficient $q \in H_0^1(\Omega)$ in the 1D potential equation

$$-u'' + qu = f \qquad \text{in } \Omega = (0,1), \tag{78}$$

$$u(0) = u(1) = 0 \tag{79}$$

The data z are an observation of $u \in L^2(\Omega)$, i.e. the observation operator E is the canonical embedding from $H^1(\Omega)$ into $L^2(\Omega)$. The right-hand side $f \in H^{-1}(\Omega)$ is given by

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

the exact reaction coefficient $\hat{q} \in H_0^1(\Omega)$ by

$$\hat{q}(x) = x(1-x).$$

In other words, we consider the parameter identification problem

$$\frac{1}{2} \|u - z^\delta\|^2_{L^2(\Omega)} \to \min_{(u,q) \in H_0^1(\Omega) \times H_0^1(\Omega)} \tag{80}$$

subject to

$$\int_0^1 u'(x)v'(x) + q(x)u(x)v(x)\,\mathrm{d}x = \int_0^1 f(x)v(x)\,\mathrm{d}x \qquad \forall\, v \in H_0^1(\Omega) \tag{81}$$

where z^δ denotes a noisy approximation of the data z.

We use the LMSQP method, for which the KKT-system of the quadratic subproblem (18) looks as follows:

$$\begin{pmatrix} I_{L^2(\Omega)} & 0 & K_k^* \\ 0 & \beta_k I_{H_0^1(\Omega)} & L_k^* \\ K_k & L_k & 0 \end{pmatrix} \begin{pmatrix} u \\ q \\ \lambda \end{pmatrix} = \begin{pmatrix} (z^\delta - u_k) \\ 0 \\ -r_k \end{pmatrix}, \tag{82}$$

with

$$K_k : H_0^1(\Omega) \to H^{-1}(\Omega), \qquad \langle K_k u, v\rangle = \int_0^1 u'(x)v'(x) + q_k(x)u(x)v(x)\,\mathrm{d}x \qquad \forall\, v \in H_0^1(\Omega)$$

$$L_k : L^2(\Omega) \to H^{-1}(\Omega), \qquad \langle L_k q, v\rangle = \int_0^1 q(x)u_k(x)v(x)\,\mathrm{d}x \qquad \forall\, v \in H_0^1(\Omega)$$

and r_k denoting the residual of the state equation, i.e.

$$\langle r_k, v\rangle = \int_0^1 u_k'(x)v'(x) + q_k(x)u_k(x)v(x)\,\mathrm{d}x - \int_0^1 f(x)v(x)\,\mathrm{d}x \qquad \forall\, v \in H_0^1(\Omega).$$

As the state equation is linear in u and q in our case, the definition of K_k and L_k is straightforward.

For the numerical realization, Ω is discretized uniformly using linear finite elements. The approximations of the state variable u and the Lagrangian multiplier λ has m degrees of freedom, for approximating the parameter q we use n degrees of freedom. The noisy data z^δ are generated by solving the state equation on a fine grid using the exact reaction coefficient, restricting the fine grid solution to a coarser grid and finally adding some high-frequency perturbation as noise.

The simple structure of our example implies a rather simple structure of the KKT-submatrices of the Galerkin-LMSQP method, in particular in (58) $\mathbf{G}$ is an L^2-mass matrix, i.e.

$$\mathbf{G} = (\langle \phi_j, \phi_i\rangle_{L^2(\Omega)})_{i,j=1,\ldots,m},$$

and $\mathbf{H}$ is an H^1-stiffness matrix, i.e.

$$\mathbf{H} = (\langle \sigma_j, \sigma_i\rangle_{H^1(\Omega)})_{i,j=1,\ldots,n}.$$

In the linearization of the state equation, $\mathbf{K}$ is defined via

$$\mathbf{K} = (\langle K_k\phi_j, \phi_i\rangle)_{i,j=1,\ldots,m}$$

and $\mathbf{L}$ via

$$\mathbf{L} = (\langle L_k\sigma_j, \phi_i\rangle)_{i=1,\ldots,m;j=1,\ldots,n}\ ,$$

where $\phi_i, i = 1,\ldots,m$ and $\sigma_j, j = 1,\ldots,m$ denote the according basis functions. Hardly any properties of the matrix $\mathbf{L}$ are known, except that $\mathbf{L}$ approximates a differential operator of order 0. We refer to (55), (56) for the definition of these matrices.

This problem is implemented in the software-system MATLAB. The KKT-system (58) is solved using a direct solver in this case, which is probably not the best choice with respect to the numerical effort for fine discretizations, but still yields reasonable results in our case. Figure 5.1 shows the results obtained with the LMSQP method for noise level $\delta = 5\%$ and $\delta = 20\%$. 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 is a typical phenomenon for inverse problems. That is why the iteration is not terminated according to the convergence criteria usually used in optimization, but due to an appropriate stopping rule (see e.g. Engl, Hanke, and Neubauer [17] for a general introduction). We used the so-called *discrepancy principle* as stopping criterion (the details can be found in Burger and Mühlhuber [12, 13]). 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.

We compare the numerical efficiency of the LMSQP-method with feasible path approaches, namely the Levenberg-Marquardt method (LM) (with the same Galerkin discretization as for

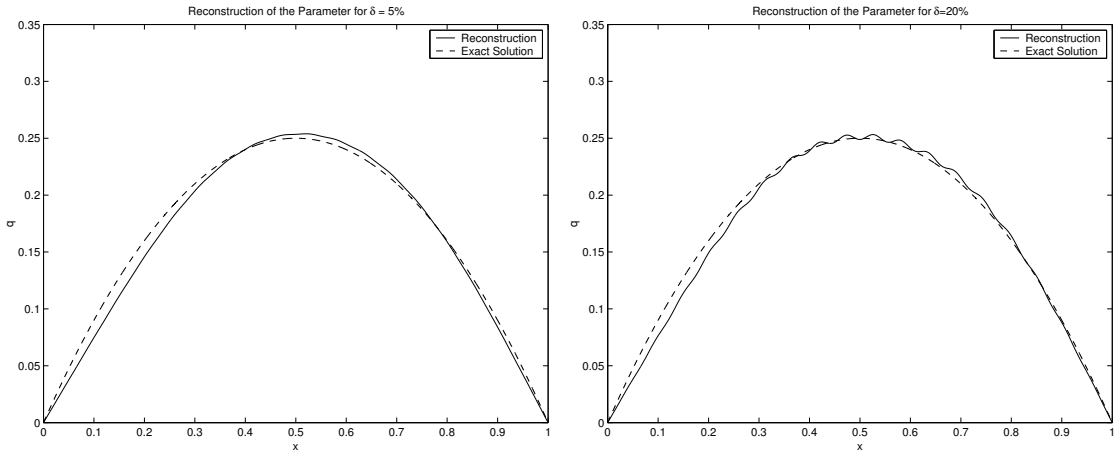

Figure 1: Reconstruction (solid) and exact solution (dashed) for noise level $\delta = 5\%$ (left) and $\delta = 20\%$ (right)

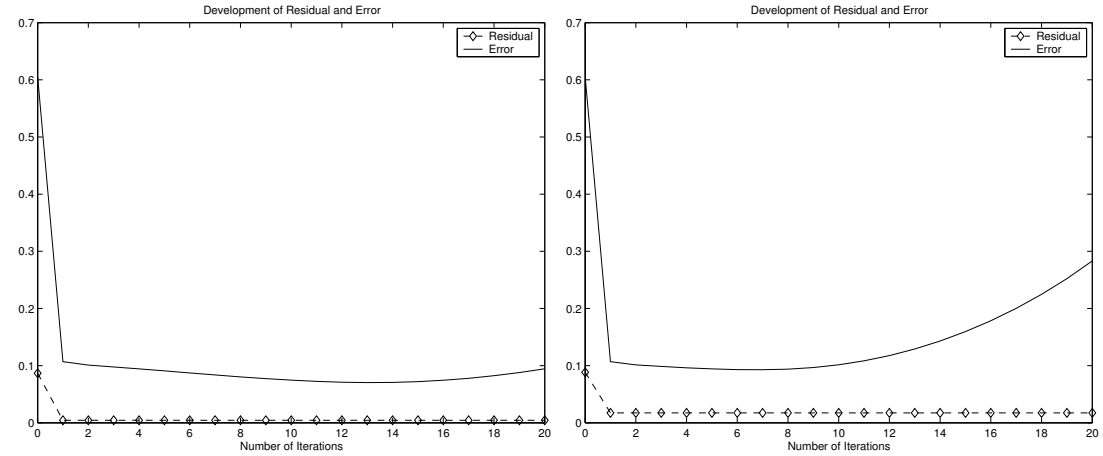

Figure 2: Development of the error $\|q_k - \hat{q}\|_{H^1(\Omega)}$ (solid) and the residual $\|u_k - z^\delta\|_{L^2(\Omega)}$ during the iteration for noise $\delta = 5\%$ (left) and $\delta = 20\%$ (right)

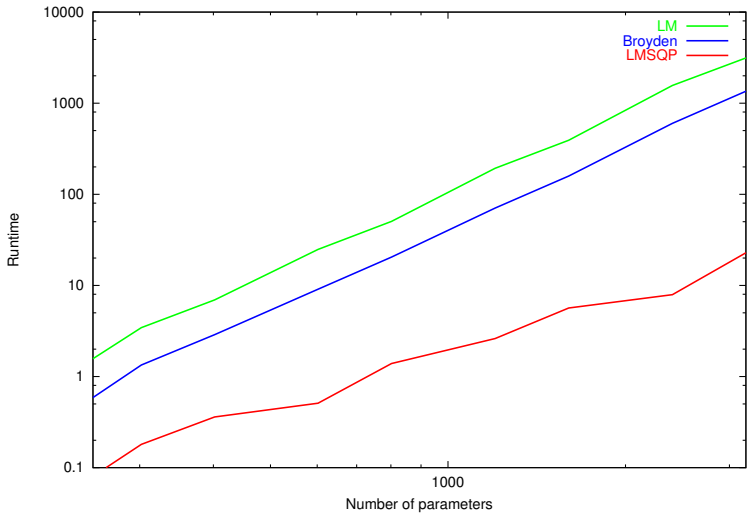

Figure 3: Comparison of the CPU-times for the LMSQP-method, the LM-Method and a Broyden-type variant of the LM-method

LMSQP) and a Broyden-type variant of the LM-method (cf. KALTENBACHER [29] for further details).

For this sake we choose different discretization levels (fixed during the iteration) and measure the CPU-time needed for the LMSQP-method, until the stopping rule is satisfied (for fixed noise level δ). From the results shown in Figure 3 one observes that the LMSQP-method with simultaneous solution of the KKT-system outperforms the feasible-path approaches for all different discretizations. Since the LMSQP and the LM-method need the same number of outer iterations, the difference in the numerical effort is caused by the fact that the effort for the evaluation of the system matrix in the LM-method is significantly higher than evaluation and preconditioning of the system matrix in the simultaneous LMSQP-method. Obviously, the gain in the numerical effort for the evaluation of the system matrix increases with the number of discretization points, which explains the extremely large CPU-time for the LM-method at the finest discretization level ($m = 1601$). For small m and n, the Broyden-variant is much faster than the LM-method, which is again caused by the fact that the evaluation of the system matrix can be carried out efficiently. However, the number of iterations needed for the Broyden-type variant is much larger than for the other two methods, which use the full information about the derivatives.

5.2 The identification of a conductivity

Our second numerical example is the identification of the conductivity $q \in L^\infty(\Omega) \subseteq L^2(\Omega)$ in

$$-\operatorname{div}(q \operatorname{grad} u) = f \qquad \text{in } \Omega, \tag{83}$$

$$u = 0 \qquad \text{on } \partial\Omega. \tag{84}$$

The data z are an observation of $u \in L^2(\Omega)$, i.e. the observation operator E is the canonical embedding of $H_0^1(\Omega)$ into $L^2(\Omega)$. The domain Ω is a ball in $\mathbb{R}^2$ with missing first quadrant

(see also Figure 4), i.e., in radial coordinates

$$\Omega = \{(r\cos\theta, r\sin\theta) \mid r \in [0,1), \theta \in (\pi/2, 2\pi)\}\,. \tag{85}$$

The exact parameter to be reconstructed is $\hat{q} \equiv 1$, the right-hand side $f \in H^{-1}(\Omega)$ in (83) is given by

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

The corresponding solution $\hat{u} \in H_0^1(\Omega)$ of the state equation is $\hat{u} = \cos(\frac{3\pi}{2}r)$. The noisy z^δ data are generated using the exact solution $\hat{u}$ perturbed by uniformly distributed random noise.

Summarizing, we consider the parameter identification problem

$$\frac{1}{2}\,\|u - z^\delta\|^2_{L^2\Omega} \to \min_{(u,q)\in H_0^1(\Omega)\times L^\infty(\Omega)} \tag{86}$$

subject to a weak formulation of the state problem (84).

We use the LMSQP-method for which the KKT-system of the quadratic subproblems looks as follows:

$$\begin{pmatrix} I_{L^2(\Omega)} & 0 & K_k^* \\ 0 & \beta_k I_{L^2(\Omega)} & L_k^* \\ K_k & L_k & 0 \end{pmatrix} \begin{pmatrix} u \\ q \\ \lambda \end{pmatrix} = \begin{pmatrix} (z^\delta - u_k) \\ 0 \\ -r_k \end{pmatrix}, \tag{87}$$

with

$$K_k : H_0^1(\Omega) \to H^{-1}(\Omega), \qquad \langle K_k u, v\rangle = \int_\Omega \langle q_k \operatorname{grad} u, \operatorname{grad} v\rangle \,\mathrm{d}x \qquad \forall\, v \in H_0^1(\Omega)$$

$$L_k : L^2(\Omega) \to H^{-1}(\Omega), \qquad \langle L_k q, v\rangle = \int_\Omega \langle q \operatorname{grad} u_k, \operatorname{grad} v\rangle \,\mathrm{d}x \qquad \forall\, v \in H_0^1(\Omega)$$

and r_k denoting the residual of the state equation, i.e.

$$\langle r_k, v\rangle = \int_\Omega \langle q_k \operatorname{grad} u_k, \operatorname{grad} v\rangle \,\mathrm{d}x - \int_\Omega f(x)v(x)\,\mathrm{d}x \qquad \forall\, v \in H_0^1(\Omega).$$

It is clear, that $L_k q$ does not exist for any $q \in L^2(\Omega)$, but only for $q \in L^\infty(\Omega)$. Thus, in the practical realization we have to introduce constraints on the parameter q. Usually, this can be done easily using a-priori information on the parameter, e.g. bounds on q.

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. This implies a rather simple structure of KKT-sub-matrices in (58), in particular $\mathbf{G}$ and $\mathbf{H}$ are mass matrices with $\mathbf{H}$ being diagonal due to the choice of piecewise constant shape functions. For the detailed definition of $\mathbf{G}$, $\mathbf{H}$, $\mathbf{K}$, and $\mathbf{L}$, see (55), (56).

In order to ensure $q \in L^\infty(\Omega)$ we added box-constraints for $\mathbf{q}$ to the discretized optimization problem which are included using a barrier method (Nocedal and Wright [38]). As they never become active, we will not go into detail here.

The results were calculated using the finite element code FEPP (see Kuhn, Langer, and Schöberl [31]), developed at the Institute of Computational Mathematics of the University of Linz.

Level	dim q	dim u	avg QMR it	SQP it	time
2	92	215	200	9	8 sec
3	368	797	200	4	15 sec
4	1472	3065	180	5	77 sec
5	5888	12017	142	6	450 sec

Table 1: CPU-time and number of inner (QMR) and outer (SQP) iterations for exact data

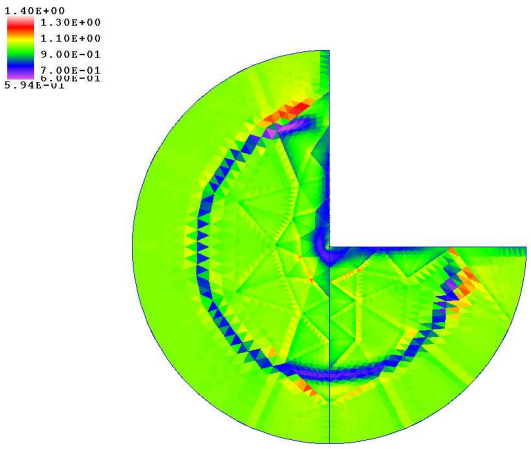

Figure 4: Parameter distribution for exact data at level 4, $q_{min} = 0.59$, $q_{max} = 1.4$

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 not identifiable along a level line in the interior, where u attains an extremum. This does not destroy the theoretical identifiability results, because it is a set of Lebesgue-measure zero, but it can be expected to create numerical difficulties.

The KKT-system (58) was solved using a preconditioned QMR method with the block-factorization type preconditioner (76) with a multi-grid preconditioner $\hat{\mathbf{K}}$ and no preconditioning of the Schur-complement $\mathbf{S_c}$. Results for exact data can be found in Table 1. The good performance of the method with respect to both, CPU time and number of outer iterations can be observed clearly. Especially for problems with fine discretizations of the parameter q, this method can still be realized efficiently, while classical approaches do not yield results in reasonable time. A plot of the finite dimensional approximation of the parameter q can be found in Figure 4, from which one observes that the parameter is reconstructed very well except in a neighborhood of the level curve $\{\operatorname{grad} u = 0\}$.

Additional speed-up can be gained using a multi-level approach as described in Subsection 3.3. We used nested spaces for approximating q and u by subdividing each triangular element into four smaller elements, when refining the mesh. Table 2 presents results for this approach. It can be seen that on fine discretization levels one SQP step is sufficient for fulfilling the stopping criterion, which corresponds very well to the theoretical predictions (for details see Burger and Mühlhuber [13]. A comparison of the results to the ones in Table 1 shows that for fixed discretization level, the solution of the identification problem on level 5 is only slightly faster than the identification of q on level 6 (with about the fourfold number

Level	dim q	dim u	avg QMR it	SQP it	time	acc. time
2	92	215	200	9	8 sec	8 sec
3	368	797	200	4	15 sec	23 sec
4	1472	3065	175	2	24 sec	47 sec
5	5888	12017	80	1	47 sec	94 sec
6	23552	47585	121	1	425 sec	520 sec

Table 2: CPU-time per level, accumulated time and number of inner (QMR) and outer (SQP) iterations for exact data using a nested multi-level approach

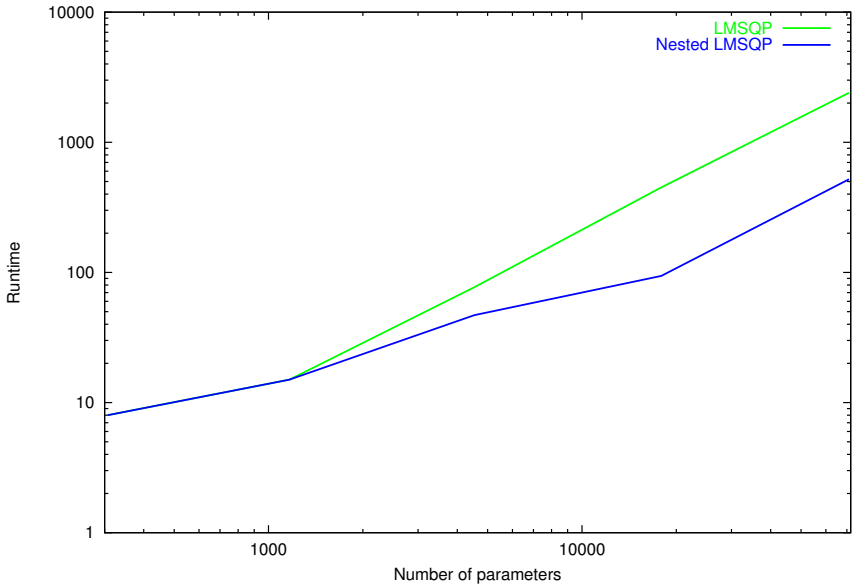

Figure 5: Comparison of the CPU-times for the LMSQP-method and its multi-level version using nested spaces

of parameters) using a multi-level approach (see also Figure 5.2).

A plot of the parameter can be found in Figure 6. Here the approximation of the parameter in the area where it can not identified is by far better than in the classical approach using only one discretization level (compare Figure 4). A possible explanation for this effect is the following: The influence of the level line $\{\operatorname{grad} u = 0\}$ where q can not be identified on the solution is smaller the coarser the discretization is. The prolongation from coarse levels to finer ones adds information to the region where the parameter is not identifiable from its surrounding region. As long as the parameter is smooth this helps to improve the quality of the numerical results where the parameter can not be identified.

Acknowledgements

This work has been supported by the Austrian National Science Foundation FWF under project grants SFB F013/08 and SFB F013/09.

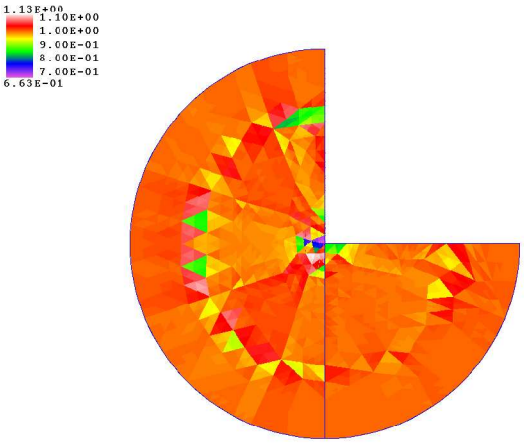

Figure 6: Parameter distribution for exact data at level 4 using a nested multi-level approach, $q_{min} = 0.66$, $q_{max} = 1.13$

References

[1] H. T. Banks and K. Kunisch, *Estimation techniques for distributed parameter systems*, Systems & Control: Foundations & Applications, vol. 1, Birkhäuser, Basel - Boston - Berlin, 1989.

[2] A. Battermann and M. Heinkenschloss, *Preconditioners for Karush-Kuhn-Tucker matrices arising in the optimal control of distributed systems*, Optimal Control of Partial Differential Equations (Vorau 1997) (W. Desch, F. Kappel, and K. Kunisch, eds.), Birkhäuser, Basel - Bosten - Berlin, 1998, pp. 15 – 32.

[3] A. Battermann and E. W. Sachs, *Block preconditioners for KKT systems in PDE-governed optimal control problems*, Fast solution of discretized optimization problems (Workshop held at the Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany) (H.-K. Hoffmann, R. H. W. Hoppe, and V. Schulz, eds.), ISNM, Internat. Series Numer. Math., vol. 138, Birkhäuser, Basel, 2001, pp. 1 – 18.

[4] R. Becker, H. Kapp, and R. Rannacher, *Adaptive finite element methods for optimal control of partial differential equations: Basic concepts*, SIAM J. Control Optim. **39** (2000), no. 1, 113 – 132.

[5] R. Becker, H. Kapp, and R. Rannacher, *Adaptive finite element methods for optimization problems*, Proceedings of the 18th Dundee biennial conference (University of Dundee, GB) (D. F. G. et al., ed.), CRC Res. Notes Math., no. 420, Chapman & Hall / CRC, Boca Raton, FL, 2000, pp. 21 – 42.

[6] G. Biros and O. Ghattas, *Parallel Lagrange-Newton-Krylov-Schwarz methods for PDE-constrained optimization problems. Part 1: The Krylov-Schur solver*, Preprint, Carnegie-Mellon University, 2000.

[7] G. Biros and O. Ghattas, *Parallel Lagrange-Newton-Krylov-Schwarz methods for PDE-constrained optimization problems. Part 2: The Lagrange-Newton solver and its appli-*

cation to optimal control of steady viscous flow, Preprint, Carnegie-Mellon University, 2001.

[8] J. H. Bramble, J. E. Pasciak, and A. T. Vassilev, *Analysis of the inexact Uzawa algorithm for saddle point problems*, SIAM J. Numer. Anal. **34** (1997), 1072 – 1092.

[9] J. H. Bramble, J. E. Pasciak, and P. S. Vassilevski, *Computational scales of Sobolev norms with application to preconditioning*, Math. Comp. **69** (2000), 463 – 480.

[10] F. Brezzi, *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers*, RAIRO Anal. Numér. **8** (1974), 129 – 151.

[11] F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, Springer Series in Computational Mathematics, vol. 15, Springer, New York - Berlin - Heidelberg, 1991.

[12] M. Burger and W. Mühlhuber, *Iterative regularization of parameter identification problems by SQP methods*, Inverse Problems **18** (2002), 943–970.

[13] M. Burger and W. Mühlhuber, *Numerical approximation of an SQP-type method for parameter identification*, SIAM J. Numer. Anal. **40** (2002), 1775–1797.

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

[15] D. Colton, R. Ewing, and W. Rundell (eds.), *Inverse problems in partial differential equations*, Philadelphia, SIAM, 1990.

[16] H. C. Elman and G. H. Golub, *Inexact and preconditioned Uzawa algorithms for saddle point problems*, SIAM J. Numer. Anal. **31** (1994), 1645 – 1661.

[17] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Mathematics and Its Applications, vol. 375, Kluwer Academic Publishers, Dordrecht - Boston - London, 1996.

[18] H. W. Engl and P. Kügler, *Identification of a temperature dependent heat conductivity by Tikhonov regularization*, J. Inverse Ill-Posed Probl. **10** (2002), no. 1, 67 – 90.

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

[20] H. W. Engl and W. Rundell (eds.), *Inverse problems in diffusion processes. Proceedings of the GAMM-SIAM symposium*, SIAM, 1995.

[21] R. W. Freund and N. M. Nachtigal, *QMR: a quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math. **60** (1991), 315 – 339.

[22] E. Haber and U. Ascher, *Preconditioned all-at-once methods for large, sparse parameter estimation problems*, Tech. report, University of British-Columbia, Vancouver, Canada, July 2001.

[23] E. Haber, U. Ascher, and D. Oldenburg, *On optimization techniques for solving nonlinear inverse problems*, Inverse Problems **16** (2000), no. 5, 1263 – 1280.

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

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

[26] M. Heinkenschloss, *The numerical solution of a control problem governed by a phase field model*, Optim. Methods Softw. **7** (1997), 211 – 263.

[27] B. Heise, *Nonlinear field calculations with multigrid Newton methods*, Impact Comput. Sci. Engrg. **5** (1993), 75 – 110.

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

[29] B. Kaltenbacher, *On Broyden's method for the regularization of non-linear ill-posed problems*, Numer. Funct. Anal. Optim. **19** (1998), 807 – 833.

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

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

[32] U. Langer and W. Queck, *On the convergence factor of Uzawa's algorithm*, J. Comput. Appl. Math. **15** (1986), 191 – 202.

[33] . Langer and W. Queck, *Preconditioned Uzawa-type iterative methods for solving mixed finite element equations. theory – applications – software*, Wissenschaftliche Schriftenreihe 3, Techn. Univ. Karl-Marx-Stadt, 1987.

[34] W. Liu and N. Yan, *A posteriori error estimates for control problems governed by Stokes equations*, SIAM J. Numer. Anal. **40** (2002), 1850–1869.

[35] D. Lukáš, *Shape optimization of homogeneous electromagnets*, SFB-Report 00-30, SFB F013, University of Linz, Austria, September 2000.

[36] D. Lukáš, *Shape optimization of homogeneous electromagnets*, Proceedings of SCEE, Lecture Notes in Computational Science and Engineering, Springer, 2000, submitted.

[37] A. Neubauer and 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. Anwendungen **14** (1995), 369 – 377.

[38] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer Series in Operations Research, Springer, New York, 1999.

[39] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite linear systems of linear equations*, SIAM J. Numer. Anal. **12** (1975), 617 – 629.

[40] A. Quarteroni and A. Valli, *Numerical approximation of partial differential equations*, Springer Series in Computational Mathematics, vol. 23, Springer, Berlin - Heidelberg - New York, 1994.

[41] W. Queck, *The convergence factor of preconditioned algorithms of the Arrow-Hurwicz type*, SIAM J. Numer. Anal. **26** (1989), no. 4, 1016 – 1030.

[42] Y. Saad and M. H. Schultz, *GMRES: a generalized minimal residual algorithm for solving non-symmetric linear systems*, SIAM J. Sci. Statist. Comput. **7** (1986), 856 – 869.

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

[44] V. Schulz, *Solving discretized optimization problems by partially reduced SQP methods*, Comput. Visualization Sci. **1** (1998), 83 – 96.

[45] V. Schulz and H.-G. Bock, *Partially reduced SQP methods for large-scale nonlinear optimization problems*, Nonlinear Anal. (1997), 4723 – 4734.

[46] S. Ta'asan, *Infinite dimensional preconditioners for optimal design problems*, Tech. report, Carnegie Mellon University, Pittsburgh, PA, 1997.

[47] S. Ta'asan, *Introduction to shape design and control*, Tech. report, Carnegie Mellon University, Pittsburgh, PA, 1997.

[48] S. Ta'asan, *Multigrid one-shot methods and design strategy*, Tech. report, Carnegie Mellon University, Pittsburgh, PA, 1997.

[49] S. Ta'asan, *Theoretical tools for problem setup*, Tech. report, Carnegie Mellon University, Pittsburgh, PA, 1997.

[50] R. Verfürth, *A review of a-posteriori error estimation and adaptive mesh-refinement techniques*, Wiley-Teubner Series on Advances in Numerical Mathematics, John Wiley & Sons, Chinchester, 1996.

[51] E. Zeidler, *Nonlinear functional analysis and its application*, vol. III, Variational Methods and Optimization, Springer, New York, 1985.

[52] W. Zulehner, *Analysis of iterative methods for saddle point problems: A unified approach*, Math. Comp. **71** (2002), 479 – 505.