Mesh Independence for Nonlinear Least Squares Problems with Norm Constraints

M. Heinkenschloss

CRPC-TR90067
August, 1990

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

Mesh Independence for Nonlinear Least Squares Problems with Norm Constraints

Matthias Heinkenschloss *

Rice University
Department of Mathematical Sciences and
Center for Research on Parallel Computation
P.O. Box 1892
Houston, Texas 77251–1892

August 9, 1990

Abstract

If one solves an infinite dimensional optimization problem by introducing discretizations and applying a solution method to the resulting finite dimensional problem, one often observes a very stable behavior of this method with respect to varying discretizations. The most striking observation is the constancy of the number of iterations needed to satisfy a certain stopping criteria. In this paper we give give an analysis of this phenomena, the so called mesh independence, for nonlinear least squares problems with norm constraints (NCNLLS). A Gauss–Newton method for the solution of NCNLLS is discussed and a convergence theorem is given. The mesh independence is proven in its sharpest formulation. Sufficient conditions for the mesh independence to hold are related to conditions guaranteeing convergence of the Gauss–Newton method. The results are demonstrated on a two point boundary value problem.

*On leave from Universität Trier, FB IV – Mathematik, Postfach 3825, D–5500 Trier, West Germany. This research was partially supported by Gottlieb-Daimler-und Karl-Benz-Stiftung, Ladenburg and NSF, Cooperate of Agreement No. CCR–8809615.

1 Introduction

This paper is concerned with the behavior of discretized Gauss–Newton methods for infinite dimensional nonlinear least squares problems of the following type:

$$\min_{||x||_X \leq R} \quad ||F(x)||_Y^2 . \tag{1.1}$$

F is a sufficiently smooth, weakly continuous function, which acts between the two Hilbert spaces X, Y. Problems of this kind for example arise in parameter identification, see e.g. [5], [16], [20]. The constraint $||x||_X \leq R$ reflects a priori information on the sought parameter and guarantees the solvability of (1.1).

If residual and nonlinearity of F are of moderate size, the Gauss–Newton method is an appropriate technique to solve (1.1). In the Gauss–Newton method the function F is linearized around a given approximation x_ℓ, whereas the constraint is retained. The approximation is improved by solving the resulting linear least squares problem. This yields the following Algorithm (here and in the subsequent chapters $B_r(x)$ will be the closed ball around x with radius r):

Algorithm

(0) Given an initial point $x_0 \in B_R(0)$, set $\ell = 0$.

(1) Solve the linearized problem

$$\min_{||x||_X \leq R} \quad ||F(x_\ell) + F'(x_\ell)(x - x_\ell)||_Y^2 \tag{1.2}$$

Let $x_{\ell+1}$ denote the solution of (1.2) and $\mu_{\ell+1}$ the corresponding Lagrange multiplier

(2) Test for convergence. If test succeeded, take $x_{\ell+1}$ as an approximation of the solution. Else

(3) Set $\ell = \ell + 1$ and goto (1)

Reviewing the convergence theorems for Gauss–Newton methods for unconstrained problems (see e.g. [7], [10], [11]), one expects a linear convergence rate for this Algorithm if the starting point is sufficiently close to the solution of (1.1). Moreover the speed of convergence should depend on the nonlinearity and size of the residual of F. A detailed convergence analysis confirming these considerations is given in chapter 2. Hence, if a good initial point is available, the problem (1.1) can theoretically be solved with the Gauss–Newton method as the outer iteration and an inner iteration scheme , e.g. the Newton or Hebden–Reinsch iteration ([19], p.273), for the solution of (1.2).

For a globalization of the convergence one can add a line search or trust region strategy. The latter leads to minimization problems with two normconstraints instead of (1.2). Utilizing the special structure of this subproblem, it can be solved using efficient methods designed for the solution of minimization problems with quadratic objective and simple norm constraint as in (1.2).

But in this paper we are only concerned with the local analysis and assume that a good estimation for the solution is available.

For the numerical solution, one has to introduce some sort of discretization for the parameter space X and the output space Y.

It is important to study the behavior of the solution method under varying discretizations. The continuous dependence of the method on changes in the discretization would guarantee the successful application of adaptive mesh refinement strategies, which are proven to be a powerful tool to solve infinite dimensional problems. Such strategies are presented in [1] for Newton's method, in [15] for Quasi–Newton methods and in [13] for the Gauss–Newton method.

The theoretical justification for mesh refinement strategies is the so called mesh independence of the method, which can roughly be described as the continuous dependence of solutions, iterates and convergence behavior of the discretized problem, respectively the method onto the discretizations.

Mesh independence in its sharpest form was developed in [2] for Newton's method. The influence of discretizations on Broyden's method was studied in [14]. There, a weaker mesh independent property was proven, which does not guarantee uniform bounds on the error between infinite and finite dimensional iterates.

In this paper we extend the mesh independence results in [2] to the normconstraint Gauss–Newton method, but we will use a somewhat different discretization scheme. We will assume, that

$$X_M \subset X \text{ and } Y_N \subset Y \text{ are finite dimensional linear subspaces}$$

and

$$F_N : X \longrightarrow Y_N \text{ is a suitable approximation for } F.$$

Although F_N is defined on the whole space X, it is evaluated only for some $x_M \in X_M$ during the numerical calculation. The discretized problem is then given as

$$\min_{\substack{\|x^M\|_X \le R \\ x^M \in X_M}} \quad \|F_N(x^M)\|_Y^2 \tag{1.3}$$

and in the ℓ–th iteration of the Gauss–Newton method we have to solve for given x_ℓ^{MN}

$$\min_{\substack{\|x^M\|_X \le R \\ x^M \in X_M}} \quad \|F_N(x_\ell^{MN}) + F_N'(x_\ell^{MN})(x^M - x_\ell^{MN})\|_Y^2 \tag{1.4}$$

instead of (1.2), where $x_\ell^{MN} \in B_R(0) \cap X_M$ is the current iteration point. Throughout the paper we will denote the iterates of the Gauss–Newton method applied to (1.3) by x_ℓ^{MN} and the corresponding Lagrange multipliers by μ_ℓ^{MN}. For the solution of (1.4) we have to compute the adjoints of $F_N'(x_\ell^{MN})$. Since we are working in the finite dimensional spaces, we define the 'adjoint' $F_N'(x)^* \in L(Y, X_M)$ through

$$< F_N'(\bar{x})^* y^N, x^M >_X = < y^N, F_N'(\bar{x}) x^M >_Y \quad \forall x^M \in X_M, y^N \in Y_N .$$

$F_N'(x)^*$ can be any extension of the $(X_M, \|\cdot\|_X), (Y_N, \|\cdot\|_Y)$ adjoint of $F_N'(x)$ onto Y. We need the extensions of F_N, $F_N'(x)$ and $F_N'(x)^*$ to apply these operators to points which are not contained in the finite dimensional subspaces. This allows us to compare infinite and finite dimensional terms without prolongation or restriction operators. For finite element discretizations these extensions are given in a natural way (see also chapter 4).

It is important to note, that $F_N'(x)^*$ is an extension of the $(X_M, \|\cdot\|_X), (Y_N, \|\cdot\|_Y)$ adjoint onto Y, but not the adjoint for the pair $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$, since in general we do not have, that

$$< F_N'(\bar{x})^* y, x >_X = < y, F_N'(\bar{x}) x >_Y \quad \forall x \in X, y \in Y\,.$$

A consequence of this fact is, that

$$\|F_N'(\bar{x})^* - F'(\bar{x})^*\|_{L(Y,X)} \neq \|F_N'(\bar{x}) - F'(\bar{x})\|_{L(X,Y)}\,.$$

Therefore we have to impose different approximation properties on the function and its derivative on one hand and its adjoint on the other. Since F_N is defined on X, it is evident, that the approximation properties of F_N and F_N' are affected only by the discretization of Y, whereas the quality of approximation of ${F_N'}^*$ is also influenced by the discretization of X. The assumptions we impose on X_M, Y_N and on the function F and its discretizations are :

Assumptions

(A1) $F \in C^1(B_R(0))$

(A2) $\|F^{(i)}(x) - F^{(i)}(y)\| \leq L_i \|x - y\| \qquad \forall x, y \in B_R(0), i = 0, 1.$

(A3) $F_N \in C^1(B_R(0))$

(A4) There exists uniformly bounded Lipschitz constants $L_i^N, i = 0, 1$ such that
$\|F_N^{(i)}(x) - F_N^{(i)}(y)\| \leq L_i^N \|x - y\| \qquad \forall x, y \in B_R(0), i = 0, 1.$
Without loss of generality we assume, that $L_i^N \leq L_i, i = 0, 1 \quad \forall N \in \mathbb{N}$.

(A5) There exists a bounded function $\rho_Y : [0,1] \longrightarrow \mathbb{R}^+$ which is continuous in 0 with $\rho_Y(0) = 0$ and which satisfies
$\|F^{(i)}(x) - F_N^{(i)}(x)\| \leq \rho_Y(1/N) \qquad \forall x \in B_R(0), i = 0, 1.$

(A6) For every x and $\delta > 0$ there exists M_δ, such that for all $M \geq M_\delta$ there exists $x_M \in X_M$ with $\|x - x_M\| \leq \delta$

(A7) There exists a bounded function $\rho_X : [0,1] \longrightarrow \mathbb{R}^+$ which is continuous in 0 with $\rho_X(0) = 0$, such that the adjoints of the original and discretized Fréchet derivatives can be estimated by
$\|F'(x)^* - F_N'(x)^*\| \leq \rho_Y(1/N) + \rho_X(1/M) \qquad \forall x \in B_R(0).$

This setting is suitable for finite element discretizations and, as already pointed out, allow us to compare the discretized and infinite dimensional terms without the incorporation of prolongation and restriction operators. Another, more important gain is, that we obtain uniform bounds for $\|x_\ell - x_\ell^{MN}\|$, which we would not get in the setting of [2] (see chapter 3). These uniform bounds enable us to deduce estimates for the error between the solution of (1.1) and the solutions of the discretized problems, which improve estimates derived from perturbation theory for infinite dimensional optimization problems. In this sense the Gauss–Newton method can be viewed as a tool for the analysis of (1.1) and its discretizations.

The sufficient conditions for mesh independence are strongly related to the conditions which are sufficient for the convergence of the Gauss–Newton method and throughout the paper we will use these conditions to formulate our mesh independent results. We do not need second order information of F.

The outline of the paper is as follows: In chapter 2 we give a convergence theorem for the algorithm stated above, which uses the special structure of (1.1) and extends results in [10]. Besides the convergence theorem we will give a result concerning the perturbation of solutions of (1.1) in presence of discretization. This result is based on perturbation theory for infinite dimensional optimization problems. In chapter 3 we will develop the mesh independence principle for the Gauss–Newton method and in chapter 4 we will discuss the application to a boundary value problem and present some numerical results.

2 Local Convergence

The main purpose of this chapter is to establish a convergence theorem for the algorithm presented in chapter 1. Gauss–Newton based algorithms for restricted nonlinear least squares problems are discussed for example in [6] and [21], but these algorithms, designed for more general problems, treat inequality constraints by active set strategies and are therefore not appropriate for our analysis. In [20], Vogel gives a convergence theorem for problem (1.2), but he uses second order information and does not utilize the least squares structure of (1.2). The analysis presented here incorporates the special structure of the problem and is a generalization of Theorem 10.2.1 in [10]. If the constraint is inactive, the assertions of both theorems are identical in the finite dimensional case.

It is well known, that the solutions of (1.2) can be characterized as solutions of the system of Kuhn–Tucker conditions:

$$\begin{aligned}(F'(x_\ell)^*F'(x_\ell)+\mu_{\ell+1}I)x_{\ell+1} &= -F'(x_\ell)^*(F(x_\ell)-F'(x_\ell)x_\ell)\\ \mu_{\ell+1}(\|x_{\ell+1}\|_X^2-R^2)&=0\\ \mu_{\ell+1}\geq 0,\quad \|x_{\ell+1}\|_X^2-R^2&\leq 0\,.\end{aligned} \tag{2.1}$$

For $\mu > 0$ let $x_\ell(\mu)$ be defined as the unique solution of

$$(F'(x_\ell)^*F'(x_\ell)+\mu I)x = -F'(x_\ell)^*(F(x_\ell)-F'(x_\ell)x_\ell) \tag{2.2}$$

and $x_\ell(0)$ the minimum norm solution of (2.2) with $\mu = 0$. If $\|x_\ell(0)\|_X > R$, the problem of finding a solution of the Kuhn–Tucker system is equivalent to the computation of a root of

$$g_\ell(\mu) := \|x_\ell(\mu)\|_X^2 - R^2\,.$$

g_ℓ is a convex and monotonically decreasing function with $g_\ell(\mu) \longrightarrow -R^2$ as $\mu \longrightarrow \infty$. Therefore the root is uniquely determined. Furthermore g_ℓ is continuously differentiable on $(0,\infty)$. The first derivative is given by

$$g_\ell'(\mu) = -2 < x_\ell(\mu), (F'(x_\ell)^*F'(x_\ell)+\mu I)^{-1}x_\ell(\mu) >_X\ . \tag{2.3}$$

Theorem 2.1 *Let X, Y be Hilbert spaces. Let $F : X \longrightarrow Y$ satisfy (A1), (A2) and x_* be a solution of*

$$\min_{\|x\|\leq R} \quad \|F(x)\|^2 \tag{2.4}$$

with $\mu_* \geq 0$ the Lagrange multiplier at x_*. Assume further, that for $\epsilon, \gamma_*, \sigma \geq 0$

$$\|F'(x_*)h\|^2 \;\geq\; \gamma_*\|h\|^2 \tag{2.5}$$

and

$$\|\,(F'(x)^* - F'(y)^*)\,F(y)\| \;\leq\; \sigma\|x-y\| \tag{2.6}$$

for all $x, y \in B_R(0) \cap B_\epsilon(x_*)$, *and for all* $h \in \{h \in X \,|\, x_* + h \in B_R(0)\}$.

If $\sigma < \gamma_* + \mu_*$, *then for all* $\alpha \in (1, (\gamma_* + \mu_*)/\sigma)$ *there exists* $\epsilon_* = \epsilon_*(\alpha)$, $\epsilon_* > 0$ *such that the solution* $x_{\ell+1}$ *of*

$$\min_{||x|| \le R} \quad ||F(x_\ell) + F'(x_\ell)(x - x_\ell)||^2 \tag{2.7}$$

obeys

$$||x_{\ell+1} - x_*|| \le \frac{\alpha\sigma}{\gamma_* + \mu_*}||x_\ell - x_*|| + \frac{\alpha L_1 \kappa}{2(\gamma_* + \mu_*)}||x_\ell - x_*||^2 \tag{2.8}$$

and

$$||x_{\ell+1} - x_*|| \le \frac{\gamma_* + \mu_* + \alpha\sigma}{2(\gamma_* + \mu_*)}||x_\ell - x_*|| < ||x_\ell - x_*|| \quad , \tag{2.9}$$

provided $x_\ell \in B_R(0) \cap B_{\epsilon_*}(x_*)$. *Here,* κ *is defined by* $\kappa := \sup_{x \in B_R(0)} ||F'(x)||$.

Moreover there exists θ, *such that if* $x_\ell \in B_R(0) \cap B_{\epsilon_*}(x_*)$ *the Lagrange multipliers can be estimated as follows*

$$|\mu_* - \mu_{\ell+1}| \le \frac{\theta\alpha}{\gamma_* + \mu_*}(\kappa \frac{L_1}{2}||x_* - x_\ell||^2 + \sigma||x_* - x_\ell||) \tag{2.10}$$

Proof: Let $c \in (1, (\gamma_* + \mu_*)/\sigma)$ be an arbitrary constant. Since F' is continuous, we obtain from (2.5), that for each $\tilde{\alpha} \in (1, \alpha)$ there exists $\epsilon_1 \in (0, \epsilon)$, such that for all $x \in B_R(0) \cap B_{\epsilon_1}(x_*)$ and for all $h \in \{h \in X \,|\, x_* + h \in B_R(0)\}$ the following inequality holds

$$< (F'(x)^* F'(x) + \mu_* I)\, h, h > \ge \frac{\gamma_* + \mu_*}{\tilde{\alpha}}||h||^2. \tag{2.11}$$

(i) In the first part of the proof we will derive the estimate for the Lagrange multipliers. If $\mu_{\ell+1}$ and μ_* are greater than zero, they are characterized as the roots of $g_\ell(\mu) := ||x_\ell(\mu)||^2 - R^2$ and $g_*(\mu) := ||x_*(\mu)||^2 - R^2$, respectively. We will utilize the convexity of these functions to give lower bounds for the Lagrange multipliers. For the development of the estimates it will be favorable to distinguish two cases:

Case 1: $\mu_* > \mu_{\ell+1}$

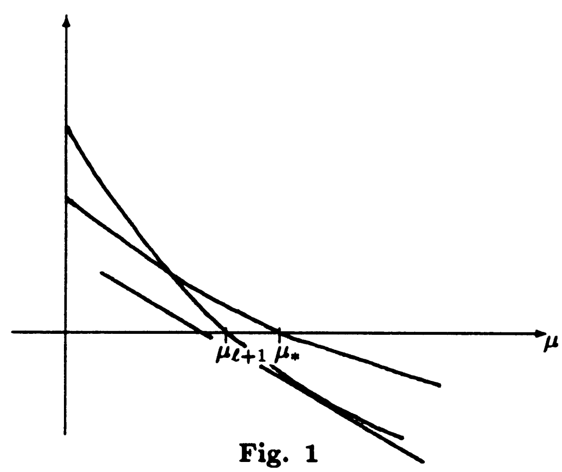

Fig. 1

From the definition of $x_\ell(\mu)$ we can conclude, that

$$\begin{aligned}||x_\ell(\mu_*)|| &\geq \frac{1}{\kappa^2+\mu_*}||F'(x_\ell)^*(F(x_\ell)-F'(x_\ell)x_\ell)|| \\ &\geq \frac{1}{\kappa^2+\mu_*}(||F'(x_*)^*(F(x_*)-F'(x_*)x_*)||-\tilde{L}||x_*-x_\ell||) \end{aligned} \tag{2.12}$$

(here, $\tilde{L}$ is a Lipschitz constant depending on $L_1, \kappa, \sup_{x\in B_R(0)}||F(x)||$ and R)
and

$$\begin{aligned}&||x_*-x_\ell(\mu_*)|| \\ &\leq ||(F'(x_\ell)^*F'(x_\ell)+\mu_*I)^{-1}|| \\ &\quad (||(F'(x_\ell)^*F'(x_\ell)+\mu_*I)(x_\ell-x_*)-(F'(x_\ell)^*F(x_\ell)+\mu_*x_\ell-F'(x_\ell)^*F(x_*)-\mu_*x_*)|| \\ &\quad +||(F'(x_*)-F'(x_\ell))^*F(x_*)+\mu_*x_*-\mu_*x_*||) \\ &\leq \frac{\tilde{\alpha}}{\gamma_*+\mu_*}(\kappa\frac{L_1}{2}||x_*-x_\ell||^2+\sigma||x_*-x_\ell||), \end{aligned} \tag{2.13}$$

provided $||x_*-x_\ell||\leq\epsilon_1$.
(2.13) yields the existence of c such that $||x_*||+||x_\ell(\mu_*)||\leq c$ independent of ℓ.
Since

$$\begin{aligned}g_\ell(\mu) &:= ||x_\ell(\mu)||^2-R^2 \\ &:= ||(F'(x_\ell)^*F'(x_\ell)+\mu I)^{-1}F'(x_\ell)^*(F(x_\ell)-F'(x_\ell)x_\ell)||^2-R^2\end{aligned} \tag{2.14}$$

is convex, we obtain

$$\begin{aligned}\mu_{\ell+1} &\geq \mu_*-\frac{g_\ell(\mu_*)}{g_\ell'(\mu_*)} \\ &= \mu_*+\frac{||x_\ell(\mu_*)||^2-R^2}{2<x_\ell(\mu_*),(F'(x_\ell)^*F'(x_\ell)+\mu_*I)^{-1}x_\ell(\mu_*)>} \\ &\geq \mu_*-\frac{\kappa^2+\mu_*}{2}\frac{R^2-||x_\ell(\mu_*)||^2}{||x_\ell(\mu_*)||^2} \\ &\geq \mu_*-\frac{\kappa^2+\mu_*}{2}\frac{||x_*||+||x_\ell(\mu_*)||}{||x_\ell(\mu_*)||^2}||x_*-x_\ell(\mu_*)|| \\ &\geq \mu_*-\frac{\kappa^2+\mu_*}{2}\frac{c}{||x_\ell(\mu_*)||^2}||x_*-x_\ell(\mu_*)||\,. \end{aligned} \tag{2.15}$$

If we choose

$$\epsilon_2 := \min\left\{\epsilon_1, \frac{||F'(x_*)^*(F(x_*)-F'(x_*)x_*)||}{2\tilde{L}}\right\}\quad ,$$

we obtain with (2.12), (2.13), (2.15), that for $||x_*-x_\ell||\leq\epsilon_2$

$$\mu_*-\mu_{\ell+1} \leq \frac{2(\kappa^2+\mu_*)^3c}{||F'(x_*)^*(F(x_*)-F'(x_*)x_*)||^2}\frac{\alpha}{\gamma_*+\mu_*}(\kappa\frac{L_1}{2}||x_*-x_\ell||^2+\sigma||x_*-x_\ell||). \tag{2.16}$$

Case 2: $\mu_* < \mu_{\ell+1}$
In this case we consider

$$\begin{aligned} g_*(\mu) &:= ||x_*(\mu)||^2 - R^2 \\ &:= ||(F'(x_*)^* F'(x_*) + \mu I)^{-1} F'(x_*)^*(F(x_*) - F'(x_*)x_*)||^2 - R^2. \end{aligned}$$

Applying the same considerations as in case 1 and using the inequalities

$$||(F'(x_*)^* F'(x_*) + \mu_{\ell+1} I)^{-1}|| \leq \frac{1}{\gamma_* + \mu_*}$$

and

$$||x_*(\mu_{\ell+1})|| \geq \frac{1}{\kappa^2 + \mu_{\ell+1}} ||F'(x_*)^*(F(x_*) - F'(x_*)x_*)||$$

yields

$$\mu_{\ell+1} - \mu_* \leq \frac{(\kappa^2 + \mu_{\ell+1})^3 c}{2||F'(x_*)^*(F(x_*) - F'(x_*)x_*)||^2} \frac{\alpha}{\gamma_* + \mu_*} (\kappa \frac{L_1}{2} ||x_* - x_\ell||^2 + \sigma ||x_* - x_\ell||). \quad (2.17)$$

In the derivation of (2.17) we have used without loss of generality, that $||x_{\ell+1}|| + ||x_*(\mu_{\ell+1})|| \leq c$ independent of ℓ. The boundedness of $||x_{\ell+1}|| + ||x_*(\mu_{\ell+1})||$ can be proven analogously to (2.13), since $x_{\ell+1} = x_\ell(\mu_{\ell+1})$.
From (2.14) it can be seen, that

$$\frac{||F'(x_\ell)^*(F(x_\ell) - F'(x_\ell)x_\ell)||}{R}$$

is an upper bound for $\mu_{\ell+1}$. Therefore the estimate (2.10) follows from (2.16) and (2.17), if we choose

$$\theta := \frac{2(\kappa^2 + s/R)^3 c}{||F'(x_*)^*(F(x_*) - F'(x_*)x_*)||^2},$$

where s is defined by $s := \sup_{x \in B_R(0)} ||F'(x)^*(F(x) - F'(x)x)||$.

(ii) For the proof of the first part of the theorem we will again distinguish two cases:

Case 1: $\mu_{\ell+1} \geq \mu_*$
From the equation

$$(F'(x_\ell)^* F'(x_\ell) + (\mu_* + \mu_{\ell+1} - \mu_*)I)x_{\ell+1} = -F'(x_\ell)^*(F'(x_\ell) - F(x_\ell)x_\ell)$$

it can be seen, that $x_{\ell+1}$ is a solution of

$$\min_{||x|| \leq R} \quad ||F(x_\ell) + F'(x_\ell)(x - x_\ell)||^2 + \mu_* ||x||^2 \quad (2.18)$$

with Lagrange multiplier $\mu_{\ell+1} - \mu_*$. Therefore the first order optimality condition

$$< F(x_\ell) + F'(x_\ell)(x_{\ell+1} - x_\ell), F'(x_\ell)(x - x_{\ell+1}) > + \mu_* < x_{\ell+1}, x - x_{\ell+1} > \geq 0 \quad \forall x \in B_R(0) \quad (2.19)$$

is satisfied. (2.19) together with the optimality condition for (2.4), $F'(x_*)^*F(x_*)+\mu_* x_* = 0$, yields

$$\begin{aligned}
&\mu_*(< x_{\ell+1}, x_{\ell+1} - x_* > - < x_*, x_{\ell+1} - x_* >) \\
&\quad \leq \; < F(x_\ell) + F'(x_\ell)(x_{\ell+1} - x_\ell), F'(x_\ell)(x_* - x_{\ell+1}) > + < F(x_*), F'(x_*)(x_{\ell+1} - x_*) > \\
&\quad \leq \; < F'(x_\ell)^* F(x_\ell) + F'(x_\ell)^* F'(x_\ell)(x_{\ell+1} - x_\ell) - F'(x_*)^* F(x_*), x_* - x_{\ell+1} > \\
&\quad \leq \; < F'(x_\ell)^* \left(F(x_*) - \int_0^1 F'(x_\ell + t(x_* - x_\ell))(x_* - x_\ell)\, dt \right) \\
&\qquad + F'(x_\ell)^* F'(x_\ell)(x_* - x_\ell) + F'(x_\ell)^* F'(x_\ell)(x_{\ell+1} - x_*) - F'(x_*)^* F(x_*), x_* - x_{\ell+1} > \qquad (2.20)
\end{aligned}$$

Hence, we obtain with (2.6), (2.11), (2.20)

$$\begin{aligned}
\frac{\gamma_* + \mu_*}{\alpha} \|x_{\ell+1} - x_*\| &\leq \frac{\gamma_* + \mu_*}{\tilde{\alpha}} \|x_{\ell+1} - x_*\| \\
&\leq \frac{1}{\|x_{\ell+1} - x_*\|} < \left(F'(x_\ell)^* F'(x_\ell) + \mu_* I\right)(x_{\ell+1} - x_*), x_{\ell+1} - x_* > \\
&\leq \|\left(F'(x_\ell)^* - F'(x_*)^*\right) F(x_*)\| \\
&\quad + \|F'(x_\ell)\| \int_0^1 \|F'(x_\ell) - F'(x_\ell + t(x_* - x_\ell))\| \, \|x_\ell - x_*\| \, dt \\
&\leq \sigma \|x_\ell - x_*\| + \frac{L_1 \kappa}{2} \|x_\ell - x_*\|^2 , \qquad (2.21)
\end{aligned}$$

provided, that $\|x_\ell - x_*\| \leq \epsilon_1$.

Case 2: $\mu_* > \mu_{\ell+1}$

x_* is the unique solution of

$$(F'(x_*)^* F'(x_*) + \mu_* I)x = -F'(x_*)^*(F'(x_*) - F(x_*)x_*)$$

and can therefore be characterized as the unique solution of

$$\min_{\|x\| \leq R} \; \|F(x_*) + F'(x_*)(x - x_*)\|^2 + \mu_{\ell+1}\|x\|^2$$

with Lagrange multiplier $\mu_* - \mu_{\ell+1}$. Similar to case 1 (replace μ_* by $\mu_{\ell+1}$ in (2.20)), it can be shown, that this yields

$$\frac{\gamma_* + \mu_*}{\tilde{\alpha}} \|x_{\ell+1} - x_*\| \;\leq\; \sigma\|x_\ell - x_*\| + \frac{L_1\kappa}{2}\|x_\ell - x_*\|^2 + (\mu_* - \mu_{\ell+1})\|x_{\ell+1} - x_*\|. \qquad (2.22)$$

with (2.10) we can conclude, that for $\|x_* - x_\ell\| < \epsilon_3$, where

$$\epsilon_3 \;:=\; \min\left\{1, \epsilon_2, \frac{2(\gamma_* + \mu_*)}{\theta\alpha\kappa L_1 + 2\theta\alpha\sigma} \frac{(\alpha - \tilde{\alpha})(\gamma_* + \mu_*)}{\alpha\tilde{\alpha}}\right\},$$

the following estimate holds

$$\mu_* - \mu_{\ell+1} \;\leq\; \frac{\gamma_* + \mu_*}{\tilde{\alpha}} - \frac{\gamma_* + \mu_*}{\alpha}. \qquad (2.23)$$

Inserting this into (2.22) yields

$$\frac{\gamma_* + \mu_*}{\alpha}\|x_{\ell+1} - x_*\| \;\leq\; \sigma\|x_\ell - x_*\| + \frac{L_1\kappa}{2}\|x_\ell - x_*\|^2 , \qquad (2.24)$$

provided $||x_\ell - x_*|| \leq \epsilon_3$.

If we choose $\epsilon_* := \min(\epsilon_3, \frac{(\gamma_*+\mu_*)-\alpha\sigma}{cL_1\kappa})$, we finally obtain from (2.21) and (2.24), that $||x_\ell - x_*|| \leq \epsilon_*$ implies

$$\begin{aligned} ||x_{\ell+1} - x_*|| &\leq \frac{\alpha\sigma}{\gamma_* + \mu_*}||x_\ell - x_*|| + \frac{\alpha L_1 \kappa}{2(\gamma_* + \mu_*)}||x_\ell - x_*||^2 \\ &\leq (\frac{\alpha\sigma}{\gamma_* + \mu_*} + \frac{\alpha L_1\kappa}{2(\gamma_* + \mu_*)}\frac{(\gamma_* + \mu_*) - \alpha\sigma}{cL_1\kappa})||x_\ell - x_*|| \\ &= \frac{(\gamma_* + \mu_*) + \alpha\sigma}{2(\gamma_* + \mu_*)}||x_\ell - x_*|| < ||x_\ell - x_*||. \end{aligned}$$

Hence, the assertion is proven. □

Remarks

(1) If one reviews the proof of (2.8), it can be seen immediately, that for the convergence of the iterates x_ℓ it is merely needed, that (2.6) holds with y replaced by x_*, i.e.

$$||\,(F'(x)^* - F'(x_*)^*)\,F(x_*)|| \leq \sigma||x - x_*|| \tag{2.25}$$

Requiring (2.25) instead of (2.6) yields qualitatively the same results . In this case one gets

$$|\mu_* - \mu_{\ell+1}| \leq \frac{\tau\alpha}{\gamma_* + \mu_*}(\kappa\frac{L_1}{2}||x_* - x_\ell||^2 + L_1 \sup_{x\in B_R(0)} ||F(x)||\,||x_* - x_\ell||) \tag{2.26}$$

instead of (2.10). In the case $\mu_* > \mu_{\ell+1}$ the constant σ can be retained in (2.26)

(2) Although inequality (2.5) is assumed to hold only for certain h, easy calculations show, that this requirement is equivalent to the condition $||F'(x_*)h||^2 \geq \gamma_*||h||^2 \quad \forall h \in X$. This property is due to the special shape of the admissible set.

(3) Since the errors in the Lagrange multipliers are dominated by the errors in the iterates, the theorem above also shows, that if x_* lies on the boundary and if $\mu_* > 0$, then it is also true for the iterates x_ℓ, for ℓ sufficiently large. In other words, the iterates approach the solution on the boundary. On the other hand, it is clear, that if the solution is an interior point of $B_R(0)$, then also the iterates x_ℓ lie in the interior up to finitely many.

Before we analyze the discretized problem, we will discuss the implications of assumption (2.6) in Theorem 2.1.

If F is two times Fréchet differentiable at the solution x_*, the property (2.6) leads to an estimate for the second order part of the second Fréchet derivative of $||F(x)||^2$ at x_*.

Lemma 2.2 *Let X, Y be Hilbert spaces and $F : X \longrightarrow Y$, $F \in C^1(B_R(0))$. If $F''(x_*)$ exists, the condition*

$$||\,(F'(x)^* - F'(x_*)^*)\,F(x_*)|| \leq \sigma||x - x_*|| \quad \forall x \in B_R(0) \cap B_\epsilon(x_*)$$

implies

$$||(F''(x_*)(\cdot, h))^* F(x_*)|| \leq \sigma||h|| \quad \forall h \in X$$

Proof: From the differentiability we obtain, that

$$||(F''(x_*)(\cdot,h))^* F(x_*)|| \leq (\sigma + \phi(h))||h|| \quad \forall h \in \{h \in X \,|\, x_* + h \in B_R(0)\},$$

where ϕ is continuous at the origin and fulfills $\phi(0) = 0$. For an arbitrary $n \in \mathbb{N}$ there exists $\delta_n > 0$ such that $\phi(h) < 1/n \quad \forall h \in B_{\delta_n}(0)$. This yields

$$||(F''(x_*)(\cdot,h))^* F(x_*)|| \leq (\sigma + \frac{1}{n})||h|| \quad \forall h \in \{h \in X \,|\, x_* + h \in B_R(0)\}.$$

Taking the limit $n \longrightarrow \infty$ gives

$$||(F''(x_*)(\cdot,h))^* F(x_*)|| \leq \sigma||h|| \quad \forall h \in \{h \in X \,|\, x_* + h \in B_R(0)\}.$$

Finally, we can apply the same considerations as in part (i) to show, that the inequality is valid for all $h \in X$. □

If we combine Theorem 2.1 and Lemma 2.2, we can conclude the following

Corollary 2.3 *Let the assumptions of Theorem 2.1 hold. Furthermore assume, that $F''(x_*)$ exists. Then the second Fréchet derivative of the Lagrangian at the solution is strictly positive:*

$$< F'(x_*)^* F'(x_*)h + (F''(x_*)(\cdot,h))^* F(x_*) + \mu_* h, h > \geq (\gamma_* + \mu_* - \sigma)||h||^2 \quad \forall h \in X$$

Hence the second order sufficient optimality criteria is satisfied at x_*. Especially we obtain, that x_* is an isolated minimizer and that the objective in (2.4) possesses local quadratic growth ([18] Theorem 5.6). This requirement seems to be inappropriate, since parameter identification problems are often 'rank–deficient' and ill–posed. But in presence of ill–posedness one has to employ regularization techniques to stabilize the problem, i.e. to guarantee continuous dependence of solutions of (2.4) upon input data. Such a technique may be the Tikhonov regularization, where a regularization term of the form $\alpha||x||^2$ is added to the objective, or a regularization by restriction of the admissible parameter set, i.e. an reduction of R. Hence, under suitable assumptions on F and on the regularization, the regularized parameter identification problem may fit the requirements of Theorem 2.1. In [8], [9] it was shown, that the output least squares formulation of elliptic parameter identification problems exhibit a quadratic growth for properly chosen regularization.

The quadratic growth of the objective function can also be used to derive an estimate for the error between the solution of the infinite dimensional problem and the solutions of the discretized ones. In the following theorem we will establish such a perturbation result without the requirement of twice Fréchet differentiable objective functions.

Theorem 2.4 *Let (A1)–(A6) and the assumptions of Theorem 2.1 are valid. Further assume, that F and F_N are weakly continuous functions. If there exists a continuous function g with $g(0) = 0$ and $g(t) \geq t \quad \forall t \in [0,1]$ such that $d(h_1,h_2) := g(\rho_Y(|h_1 - h_2|))$ defines a metric on $[0,1]$, then for all $\delta > 0$ there exists M_δ and N_δ such that for all $M \geq M_\delta$, $N \geq N_\delta$ the discretized problem*

$$\min_{\substack{||x^M|| \leq R \\ x^M \in X_M}} \quad ||F_N(x^M)||^2 \tag{2.27}$$

has a solution x_*^{MN} satisfying

$$||x_* - x_*^{MN}|| \leq \delta\,.$$

Proof: By (A2) and (A5) there exists $c > 0$ such that for all N sufficiently large and $x_1, x_2 \in B_R(0)$

$$||F(x_1)||^2 - ||F_N(x_2)||^2 \leq c(\rho_Y(1/N) + ||x_1 - x_2||) \leq c(d(0, 1/N) + ||x_1 - x_2||).$$

This shows, that the discretization of F defines a Lipschitzian perturbation. From (2.5), (2.6), (A1) and the complementary condition $\mu_*(||x_*||^2 - R^2) = 0$ we can conclude that for some Lipschitz constant $\tilde{L}$ and arbitrary $x \in B_R(0)$

$$\begin{aligned}
||F(x)||^2 &\geq ||F(x)||^2 + \mu_*(||x||^2 - R^2) \\
&\geq ||F(x_*) + \int_0^1 F'(x_* + t(x - x_*))dt||^2 + \mu_*(||x||^2 - R^2) \\
&\quad -2 < F'(x_*)F(x_*) + \mu_* x_*, x - x_* > \\
&\geq ||F(x_*)||^2 + \mu_*||x - x_*||^2 - \sigma||x - x_*||^2 + \mu_*(||x_*||^2 - R^2) \\
&\quad + ||\int_0^1 F'(x_* + t(x - x_*))dt||^2\, ||x - x_*||^2 + \gamma_*||x - x_*||^2 - ||F'(x_*)||^2\, ||x - x_*||^2 \\
&\geq ||F(x_*)||^2 + (\mu_* + \gamma_* - \sigma)||x - x_*||^2 - \tilde{L}||x - x_*||^3.
\end{aligned}$$

With $\alpha = (\mu_* + \gamma_* - \sigma)/2$ this yields the following growth condition for the infinite dimensional problem:

$$||F(x)||^2 \geq ||F(x_*)||^2 + \alpha||x - x_*||^2, \tag{2.28}$$

for all x with $||x - x_*|| \leq (\mu_* + \gamma_* - \sigma)/(2\tilde{L})$. Hence, the the results of Alt ([3], Theorem 4, 6) yield the existence of $\tilde{N}$, such that for each $N \geq \tilde{N}$ there exists a solution x_*^N of

$$\min_{||x|| \leq R} \quad ||F_N(x)||^2$$

with

$$||x_* - x_*^N|| \leq \tilde{c}\sqrt{d(0, \frac{1}{N})}, \tag{2.29}$$

where $\tilde{c}$ is independent of N.

For x with $||x - x_*|| \leq (\mu_* + \gamma_* - \sigma)/(2\tilde{L})$ we deduce from (2.28), (2.29) and (A5)

$$\begin{aligned}
&\rho_Y(\frac{1}{N})^2 + 2\rho_Y(\frac{1}{N})||F_N(x)|| + ||F_N(x)||^2 \\
&\geq ||F(x)||^2 \\
&\geq ||F(x_*)||^2 + \alpha||x - x_*||^2 \\
&\geq ||F(x_*^N)||^2 - 2L_0||x_* - x_*^N||\, ||F(x_*^N)|| + \alpha||x - x_*^N||^2 - 2\alpha||x - x_*^N||\, ||x_* - x_*^N|| \\
&\geq ||F_N(x_*^N)||^2 + \alpha||x - x_*^N||^2 - 2\rho_Y(\frac{1}{N})||F_N(x_*^N)|| - (2L_0\tilde{c}||F(x_*^N)|| - 4R\alpha\tilde{c})\sqrt{d(0, \frac{1}{N})}
\end{aligned}$$

define $\xi = \min\{\delta/2, (\mu_* + \gamma_* - \sigma)/(4\tilde{L})\}$. If we choose $N_\delta \geq \tilde{N}$ such that

$$\sqrt{d(0, \frac{1}{N})} \leq \frac{\xi}{\tilde{c}}$$

and

$$\rho_Y(\frac{1}{N})^2 + 2(||F_N(x)|| + ||F_N(x_*^N)||)\rho_Y(\frac{1}{N}) + (2L_0||F(x_*^N)|| + 4R\alpha)\tilde{c}\sqrt{d(0,\frac{1}{N})} \leq \frac{\alpha\xi^2}{2}$$

for all $N \geq N_\delta$, we obtain the following growth condition for the finite dimensional objective function

$$||F_N(x)||^2 \geq ||F_N(x_*^N)||^2 + \alpha||x - x_*^N||^2 - \frac{\alpha\xi^2}{2}, \tag{2.30}$$

provided $||x - x_*^N|| < (\mu_* + \gamma_* - \sigma)/(4\tilde{L})$.
Let $\tilde{L}_0 = 2L_0 \sup_{x \in B_R(0)} ||F_N(x)||$. Then we obtain from (A4), (A5) that

$$|\,||F_N(x_1)||^2 - ||F_N(x_2)||^2| \leq \tilde{L}_0||x_1 - x_2||.$$

By (A6) there exists M_δ such that for all $M \geq M_\delta$ there exits $x^M \in X_M$ with $||x_*^N - x^M|| \leq \min\{\xi, \tilde{\alpha}\xi^2/(4\tilde{L}_0)\}$. Let x_*^{MN} denote a solution of

$$\min_{\substack{||x|| \leq R \\ x^M \in B_\xi(x_*^N) \cap X_M}} ||F_N(x)||^2 .$$

In the next step we will show, that x_*^{MN} is a local solution of (2.27), which will be proven if we show, that $||x_*^N - x_*^{MN}|| < \xi$. Assume, that $||x_*^N - x_*^{MN}|| = \xi$. Then (2.30) yields

$$||F_N(x_*^{MN})||^2 \geq ||F_N(x_*^N)||^2 + \alpha\xi^2 - \frac{\alpha\xi^2}{2}.$$

On the other hand each $x^M \in B_R(0)$ with $||x_*^N - x^M|| \leq \alpha\xi^2/(4\tilde{L}_0)$ satisfies

$$||F_N(x_*^{MN})||^2 \leq ||F_N(x^M)||^2 \leq ||F_N(x_*^N)||^2 + \frac{\alpha\xi^2}{4}.$$

Hence

$$||F_N(x_*^{MN})||^2 \leq ||F_N(x^M)||^2 \leq ||F_N(x_*^{MN})||^2 - \frac{\alpha\xi^2}{4},$$

a contradiction.
This gives the assumption, since each local minimizer x_*^{MN} of (2.27) fulfills

$$||x_*^{MN} - x_*|| \leq ||x_*^{MN} - x_*^N|| + ||x_*^M - x_*|| < \delta/2 + \tilde{c}\sqrt{d(0,1/N)} \leq \delta.$$

□

If we have $\rho_Y(h) = c\,h^p$ with $p \geq 1$, which is usually the case for finite element discretizations, we can choose $g(t) = t^{1/p}$.

Theorem 2.4 gives a qualitative result on the perturbations of solutions, but does not give error estimates for the difference between x_* and x_*^{MN}, although the derivation of the theorem indicates, that $||x_*^M - x_*||$ is dominated by $\sqrt{d(0,1/N)}$ and $||x_*^{MN} - x_*^N||$ by $dist(X, X_M) = \sup_{x \in X} \inf_{x^M \in X_M} ||x^M - x||$. But note, that since M_δ in (A6) depends on δ and x, the distance $dist(X, X_M)$ may be infinite for fixed M. A detailed analysis of the Gauss–Newton method, which will be presented in the next section, will enable us to improve this theorem. We will derive error estimates related to the approximation properties of the discretization as well as uniqueness results for the minimizers of the discretized problems.

3 Mesh Independence

In this chapter we will investigate the behavior of the Gauss–Newton–Method for the discretized problem. Our goal is to develop estimates for the difference between the Gauss–Newton iterates of the infinite and finite dimensional problem.

In the sequel we will use some basic estimates, which are collected in the following lemma.

Lemma 3.1 *Assume, that (A1), (A2), (A3), (A5) and (A7) are valid. Define $\rho = \rho_X(1/M) + \rho_Y(1/N)$. Then there exist constants c_1, c_2 and c_3, independent of M and N, such that for all $x, x^M, y \in B_R(0)$ and for all $N \in \mathbb{N}$ the following inequalities hold.*

$$||F_N'(x^M)^* F_N'(x^M) - F'(x)^* F'(x)|| \le c_1(\rho + ||x - x^M||) \tag{3.1}$$

$$||F'(x)^*(F(x) - F'(x)x) - F_N'(x^M)^*(F_N(x^M) - F_N'(x^M)x^M)|| \le c_2(\rho + ||x - x^M||) \tag{3.2}$$

$$||F'(x)^* F'(x) - F'(y)^* F'(y)|| \le c_3 ||x - y|| \tag{3.3}$$

Proof: Define $\tilde{c}_4 := \sup_{x \in B_R(0)} ||F(x)||$ and $\tilde{c}_5 := \sup_{x \in B_R(0)} ||F'(x)||$. From (A5) we obtain that

$$\begin{aligned}
\sup_{x^M \in B_R(0)} ||F_N(x^M)|| \le \tilde{c}_4 + \rho_Y(1/N) &=: c_4^{MN}, \\
\sup_{x^M \in B_R(0)} ||F_N'(x^M)|| \le \tilde{c}_5 + \rho_Y(1/N) &=: c_5^{MN} \qquad \text{and} \\
\sup_{x^M \in B_R(0)} ||F_N'(x^M)^*|| \le \tilde{c}_5 + \rho_X(1/M) + \rho_Y(1/N) &=: c_5^{MN^*}.
\end{aligned}$$

By (A5), (A7) c_4^{MN}, c_5^{MN} and $c_5^{MN^*}$ are uniformly bounded. We set $c_4 = \max\{\tilde{c}_4, \sup_{M,N} c_4^{MN}\}$ and $c_5 = \max\{\tilde{c}_5, \sup_{M,N}(c_5^{MN}, c_5^{MN^*})\}$. This yields

$$||F_N'(x^M)^* F_N'(x^M) - F'(x)^* F'(x)|| \le 2(1 + L_1)c_5(\rho_X(1/M) + \rho_Y(1/N) + ||x - x^M||)$$

which proves (3.1).
(3.2) can be derived in a similar way. We have the following inequalities

$$\begin{aligned}
&||F'(x)^*(F(x) - F'(x)x) - F_N'(x^M)^*(F_N(x^M) - F_N'(x^M)x^M)|| \\
&\le \; ||F'(x)^*|| \, ||F(x) - F'(x)x - (F_N(x^M) - F_N'(x^M)x^M)|| \\
&\quad + ||F'(x)^* - F_N'(x^M)^*|| \, ||F_N(x^M) - F_N'(x^M)x^M|| \\
&\le \; c_5(L_0||x - x^M|| + \rho_Y(1/N) + c_3||x - x^M|| + R(L_1||x - x^M|| + \rho_Y(1/N))) \\
&\quad + (L_1||x - x^M|| + \rho_X(1/M) + \rho_Y(1/N)) \sup_{x^M \in B_R(0)} ||F_N(x^M) - F_N'(x^M)x^M|| .
\end{aligned}$$

The last inequality together with the estimate

$$\sup_{x^M \in B_R(0)} ||F_N(x^M) - F_N'(x^M)x^M|| \le c_4 + L_0||x - x^M|| + \rho_Y(\frac{1}{N}) + R(c_3 + L_1||x - x^M|| + \rho_Y(\frac{1}{N}))$$

yields the desired result.
(3.3) results with the choice $c_3 \equiv 2c_5L_1$. □

Now, we are able to derive our fundamental estimates for the iterates and the Lagrange multipliers. In the proofs of these results we will utilize the special representation of the iterates $x_{\ell+1}$, $x^{M}_{\ell+1}$. We set

$$\begin{aligned} x_\ell(\mu) &:= -(F'(x_\ell)^*F'(x_\ell)+\mu I)^{-1}F'(x_\ell)^*(F(x_\ell)-F'(x_\ell)x_\ell) \\ &= x_\ell-(F'(x_\ell)^*F'(x_\ell)+\mu I)^{-1}(F'(x_\ell)^*F(x_\ell)+\mu x_\ell) \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} x_\ell^{MN}(\mu) &:= -(F_N'(x_\ell^{MN})^*F_N'(x_\ell^{MN})+\mu I)^{-1}F_N'(x_\ell^{MN})^*(F_N(x_\ell^{MN})-F_N'(x_\ell^{MN})x_\ell^{MN}) \\ &= x_\ell^{MN}-(F_N'(x_\ell^{MN})^*F_N'(x_\ell^{MN})+\mu I)^{-1}(F_N'(x_\ell^{MN})^*F_N(x_\ell^{MN})+\mu x_\ell^{MN}) \end{aligned} \tag{3.5}$$

With these abbreviations we especially obtain, that $x_\ell(\mu_{\ell+1})=x_{\ell+1}$ and $x_\ell^{MN}(\mu_{\ell+1}^{MN})=x_{\ell+1}^{MN}$.

Lemma 3.2 *Assume, that (A1) - (A5) and (A7) are valid and that*

$$\|F'(x_\ell)h\|^2 \geq \gamma\|h\|^2 \quad \forall h\in X.$$

Define $\eta := \|x_\ell^{MN}-x_\ell\|$ *and* $\rho := \rho_X(\frac{1}{M})+\rho_Y(\frac{1}{N})$. *Let*

$$\|(F'(x_\ell)^*-F'(x_\ell^{MN})^*)F(x_\ell)\| \leq \sigma\|x_\ell^{MN}-x_\ell\|.$$

If $c_1(\rho_X(1/M)+\rho_Y(1/N)+\eta) < \gamma+\mu$ *, where c_1 is defined in Lemma 3.1, then there exists* $c_6>0$, *independent of M,N and ℓ, such that*

$$\|x_\ell^{MN}(\mu)-x_\ell(\mu)\| \leq \frac{c_6\eta^2+c_1(\rho+\eta)\|x_\ell(\mu)-x_\ell\|+\sigma\eta+c_6\rho}{\gamma+\mu-c_1(\rho+\eta)}. \tag{3.6}$$

Proof: From the definition of $x_\ell(\mu)$ and $x_\ell^{MN}(\mu)$ (see (3.4) and (3.5)), we obtain

$$\begin{aligned} &\|x_\ell^{MN}(\mu)-x_\ell(\mu)\| \\ &\leq \|(F_N'(x_\ell^{MN})^*F_N'(x_\ell^{MN})+\mu I)^{-1}\| && (3.7)\\ &\quad \{\|(F_N'(x_\ell^{MN})^*F_N'(x_\ell^{MN})+\mu I)(x_\ell^{MN}-x_\ell) \\ &\quad\quad -(F_N'(x_\ell^{MN})^*F_N(x_\ell^{MN})+\mu x_\ell^{MN}-F_N'(x_\ell^{MN})^*F_N(x_\ell)-\mu x_\ell)\| && (3.8)\\ &\quad +\|\left((F_N'(x_\ell^{MN})^*F_N'(x_\ell^{MN})+\mu I)-(F'(x_\ell)^*F'(x_\ell)+\mu I)\right) \\ &\quad\quad (F'(x_\ell)^*F'(x_\ell)+\mu I)^{-1}(F'(x_\ell)^*F(x_\ell)+\mu x_\ell)\| && (3.9)\\ &\quad +\|(F'(x_\ell)^*F(x_\ell)+\mu x_\ell)-(F_N'(x_\ell^{MN})^*F_N(x_\ell)+\mu x_\ell)\|\} && (3.10) \end{aligned}$$

Using the basic estimates of Lemma 3.1, the expressions (3.7) – (3.10) can be estimated as follows:

$$\begin{aligned} &\|(F_N'(x_\ell^{MN})^*F_N'(x_\ell^{MN})+\mu I)^{-1}\| \\ &\leq \frac{\|(F'(x_\ell)^*F'(x_\ell)+\mu I)^{-1}\|}{1-\|(F'(x_\ell)^*F'(x_\ell)+\mu I)^{-1}(F'(x_\ell)^*F'(x_\ell)-F_N'(x_\ell^{MN})^*F_N'(x_\ell^{MN}))\|} \\ &\leq \frac{\frac{1}{\gamma+\mu}}{1-\frac{1}{\gamma+\mu}c_1(\rho+\eta)} && (3.11) \end{aligned}$$

$$
\begin{aligned}
&||F_N'(x_\ell^{MN})^* F_N'(x_\ell^{MN})(x_\ell^{MN} - x_\ell) - (F_N'(x_\ell^{MN})^* F_N(x_\ell^{MN}) - F_N'(x_\ell^{MN})^* F_N(x_\ell))|| \\
&\quad \leq \ ||F_N'(x_\ell^{MN})^*|| \ \ ||F_N'(x_\ell^{MN})(x_\ell^{MN} - x_\ell) - F_N(x_\ell^{MN}) - F_N(x_\ell)|| \\
&\quad \leq \ c_5 L_1/2\, \eta^2
\end{aligned}
$$

$$
\begin{aligned}
&||(F_N'(x_\ell^{MN})^* F_N'(x_\ell^{MN} - F'(x_\ell)^* F'(x_\ell))(F'(x_\ell)^* F'(x_\ell) + \mu I)^{-1}(F'(x_\ell)^* F(x_\ell) + \mu x_\ell)|| \\
&\quad \leq \ c_1(\rho + \eta)||x_\ell(\mu) - x_\ell||
\end{aligned}
$$

$$
\begin{aligned}
&||F'(x_\ell)^* F(x_\ell) - F_N'(x_\ell^{MN})^* F_N(x_\ell)|| \\
&\quad \leq \ ||(F'(x_\ell)^* - F'(x_\ell^{MN})^*)F(x_\ell)|| + ||(F'(x_\ell^{MN})^* - F_N'(x_\ell^{MN})^*)F(x_\ell)|| \\
&\qquad + ||F_N'(x_\ell^{MN})^*|| \ ||F_N(x_\ell) - F(x_\ell)|| \\
&\quad \leq \ \sigma\eta + c_4\rho + c_5\rho_Y(1/N)
\end{aligned}
$$

Inserting these bounds into (3.7)–(3.10), we obtain the desired result by setting

$$
c_6 := \max\{c_5 L_1/2, c_5 + c_4\}
$$

□

For the derivation of the estimate for the Lagrange multipliers we will utilize the convexity of $||x_\ell(\mu)||^2 - R^2$ and its discretized analogue.

Lemma 3.3 *Let the assumptions of Lemma (3.2) are valid. Define* $\eta := ||x_\ell^{MN} - x_\ell||$ *and* $\rho := \rho_X(\frac{1}{M}) + \rho_Y(\frac{1}{N})$. *If* $||x_\ell^{MN}(\mu_{\ell+1}) - x_\ell(\mu_{\ell+1})|| < R$ *and* $c_1(\rho + \eta) < \gamma + \mu_{\ell+1}$, *then there exist* c_7, *independent of* M *and* N , *such that*

$$
\begin{aligned}
&|\mu_{\ell+1}^{MN} - \mu_{\ell+1}| \\
&\leq \ \frac{c_7(1 + ||x_\ell^{MN}(\mu_{\ell+1})||)}{(1 - ||x_\ell^{MN}(\mu_{\ell+1}) - x_\ell(\mu_{\ell+1})||/R)^2} \ \frac{c_6\eta^2 + c_1(\rho + \eta)||x_\ell(\mu_{\ell+1}) - x_\ell|| + \sigma\eta + c_6\rho}{\gamma + \mu_{\ell+1} - c_1(\rho + \eta)}
\end{aligned}
\tag{3.12}
$$

Proof: If $\mu_{\ell+1} = \mu_{\ell+1}^{MN}$ the assertion follows immediately. Therefore let us assume, that $\mu_{\ell+1} \neq \mu_{\ell+1}^{MN}$. Set

$$
\begin{aligned}
g_\ell(\mu) &:= ||x_\ell(\mu)||^2 - R^2, \\
g_\ell^{MN}(\mu) &:= ||x_\ell^{MN}(\mu)||^2 - R^2 .
\end{aligned}
\tag{3.13}
$$

From the definition of g_ℓ, g_ℓ^{MN} we obtain

$$
\begin{aligned}
|g_\ell(\mu) - g_\ell^{MN}(\mu)| &= (\,||x_\ell(\mu)|| + ||x_\ell^{MN}(\mu)||\,) \ \ |\,||x_\ell(\mu)|| - ||x_\ell^{MN}(\mu)||\,| \\
&\leq \ 2R||x_\ell(\mu) - x_\ell^{MN}(\mu)||,
\end{aligned}
\tag{3.14}
$$

provided $\mu \geq \max\{\mu_{\ell+1}, \mu_{\ell+1}^{MN}\}$ and (see (2.3))

$$
|g_\ell^{MN\prime}(\mu)| = 2 < x_\ell^{MN}(\mu), (F_N'(x_\ell^{MN})^* F_N'(x_\ell^{MN}) + \mu I)^{-1} x_\ell^{MN}(\mu) >
$$

Since $F_N'(x_\ell^{MN})^* F_N'(x_\ell^{MN})$ is selfadjoint on $(X_M, < \cdot, \cdot >_X)$ it holds, that

$$
< (F_N'(x_\ell^{MN})^* F_N'(x_\ell^{MN}) + \mu I)^{-1} h_M, h_M > \geq \frac{1}{||F_N'(x_\ell^{MN})||^2 + \mu} ||h_M||^2 \quad \forall h_M \in X_M .
$$

Hence

$$|g_\ell^{MN'}(\mu)| \;\geq\; \frac{2}{c_5^2+\mu}||x_\ell^{MN}(\mu)||^2 . \tag{3.15}$$

Now we will combine the estimates above, to develop the estimate for the error between the Lagrange multipliers. First let us consider the case $\mu_{\ell+1} < \mu_{\ell+1}^{MN}$:

For $\mu \in [\mu_{\ell+1}, \mu_{\ell+1}^{MN}]$ we obtain similar to (3.14) that

$$g_\ell^{MN}(\mu) - (R + ||x_\ell^{MN}(\mu)||)||x_\ell(\mu) - x_\ell^{MN}(\mu)|| \leq g_\ell(\mu) \leq 0 . \tag{3.16}$$

Since g_ℓ^{MN} is convex and $g_\ell^{MN}(\mu_{\ell+1}^{MN}) = 0$ (keep in mind, that $\mu_{\ell+1}^{MN} > 0$), we conclude

$$g_\ell^{MN}(\mu) \geq g_\ell^{MN}(\mu_{\ell+1}^{MN}) + |g_\ell^{MN'}(\mu_{\ell+1}^{MN})|\, |\mu - \mu_{\ell+1}^{MN}| .$$

With (3.15) and $||x_\ell^{MN}(\mu_{\ell+1}^{MN})|| = R$ this gives

$$g_\ell^{MN}(\mu) \geq \frac{2R^2}{c_5^2 + \mu_{\ell+1}^{MN}} |\mu - \mu_{\ell+1}^{MN}| . \tag{3.17}$$

Inserting (3.17) into (3.16) yields

$$\begin{aligned} 0 \;&\geq\; g_\ell^{MN}(\mu_{\ell+1}) - (R + ||x_\ell^{MN}(\mu)||)||x_\ell(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1})|| \\ &\geq\; \frac{2R^2}{c_5^2+\mu_{\ell+1}^{MN}} |\mu_{\ell+1}^{MN} - \mu_{\ell+1}| - (R + ||x_\ell^{MN}(\mu_{\ell+1})||)||x_\ell(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1})|| , \end{aligned}$$

respectively

$$|\mu_{\ell+1}^{MN} - \mu_{\ell+1}| \;\leq\; \frac{c_5^2 + \mu_{\ell+1}^{MN}}{2R^2}(R + ||x_\ell^{MN}(\mu_{\ell+1})||)||x_\ell(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1})|| .$$

With the estimates of Lemma 3.2 we finally obtain

$$|\mu_{\ell+1}^{MN} - \mu_{\ell+1}| \leq \frac{c_5^2+\mu_{\ell+1}^{MN}}{2R^2}(R + ||x_\ell^{MN}(\mu_{\ell+1})||)\, \frac{c_6\eta^2 + c_1(\rho+\eta)||x_\ell(\mu_{\ell+1}) - x_\ell|| + \sigma\eta + c_6\rho}{\gamma + \mu_{\ell+1} - c_1(\rho+\eta)} . \tag{3.18}$$

In the case $\mu_{\ell+1} > \mu_{\ell+1}^{MN}$ we can proceed as follows:
From the convexity of g_ℓ^{MN} we obtain

$$\begin{aligned} \mu_{\ell+1}^{MN} \;&\geq\; \mu_{\ell+1} - \frac{g_\ell^{MN}(\mu_{\ell+1})}{g_\ell^{MN'}(\mu_{\ell+1})} \\ &\geq\; \mu_{\ell+1} + \frac{g_\ell^{MN}(\mu_{\ell+1}) - g_\ell(\mu_{\ell+1})}{|g_\ell^{MN'}(\mu_{\ell+1})|} . \end{aligned}$$

In the last equality it was used, that $\mu_{\ell+1} > 0$ is the root of g_ℓ. Together with the estimates (3.14) and (3.15) we get

$$\begin{aligned} &|\mu_{\ell+1}^{MN} - \mu_{\ell+1}| \\ &\leq\; \frac{R(c_5^2 + \mu_{\ell+1})}{||x_\ell^{MN}(\mu_{\ell+1})||^2}||x_\ell^{MN}(\mu_{\ell+1}) - x_\ell(\mu_{\ell+1})|| \\ &\leq\; \frac{(c_5^2+\mu_{\ell+1})/R}{(1 - ||x_\ell^{MN}(\mu_{\ell+1}) - x_\ell(\mu_{\ell+1})||/R)^2}\, \frac{c_6\eta^2 + c_1(\rho+\eta)||x_\ell(\mu_{\ell+1}) - x_\ell|| + \sigma\eta + c_6\rho}{\gamma + \mu_{\ell+1} - c_1(\rho+\eta)} . \end{aligned} \tag{3.19}$$

From the definition of g_ℓ^{MN} (see (3.13) and (3.4)) it can be seen, that

$$\frac{||F_N'(x_\ell^{MN})^*(F_N(x_\ell^{MN}) - F_N'(x_\ell^{MN})x_\ell^{MN})||}{R}$$

and therefore (see (3.1))

$$\frac{||F'(x_\ell)^*(F(x_\ell) - F'(x_\ell)x_\ell)|| + c_2(\rho + \eta)}{R}$$

is an upper bound for $\mu_{\ell+1}^{MN}$. This gives the assertion, since $(c_5^2 + \mu_{\ell+1})/R$ and $(c_5^2 + \mu_{\ell+1}^{MN})/R$ in (3.18), (3.19) are uniformly bounded. □

Theorem 3.4 *Assume, that (A1) - (A5) and (A7) are valid and that the assumptions of Theorem 2.1 hold. Then there exists ϵ_1, c (both independent of M, N), M_1, N_1 and a function $\tau : \mathbb{N}^2 \to \mathbb{R}^+$, such that for all $x_0 \in B_R(0) \cap B_{\epsilon_1}(x_*)$, $M \geq M_1$, $N \geq N_1$ the condition $||x_0 - x_0^{MN}|| \leq \tau(M, N)$ implies*

$$\begin{aligned} ||x_\ell - x_\ell^{MN}|| &\leq c(\rho_X(\frac{1}{M}) + \rho_Y(\frac{1}{N})) \quad \forall \ell \qquad \textit{and} \qquad (3.20)\\ |\mu_\ell - \mu_\ell^{MN}| &\leq c(\rho_X(\frac{1}{M}) + \rho_Y(\frac{1}{N})) \quad \forall \ell \qquad (3.21) \end{aligned}$$

Proof: Throughout the proof let $\gamma_*, \mu_*, \epsilon_*, \sigma, \theta, \kappa, L_1$ and α denote the constants defined in Theorem 2.1 and its proof. For brevity we define $\rho := \rho_X(\frac{1}{M}) + \rho_Y(\frac{1}{N})$.
The proof of the theorem is somewhat technical and therefore will be split in two pieces. In the first part of the proof we will examine the unconstrained case and provide the essential estimates. In the second part we will treat the general case which requires to bound $|\mu_\ell - \mu_\ell^{MN}|$ and $||x_\ell(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1})||$ simultaneously. Although the second part is more extensive, it is based merely on the same estimates which will be applied in the first part.

(i) Assume, that $||x_*|| < R$ (, which implies $\mu_* = 0$). In the proof of Theorem 2.1 ϵ_* was chosen such that $||F'(x)||^2 \geq \frac{\gamma_*}{\alpha}||h||^2$ for all $x \in B_R(0) \cap B_{\epsilon_*}(x_*)$, moreover $\frac{\gamma_*}{\alpha} > \sigma$. Set $\gamma := \frac{\gamma_*}{\alpha}$ and choose

$$\epsilon_1 := \min\{\epsilon_*, \frac{\gamma - \sigma}{8c_1}, \frac{R - ||x_*||}{2}\}. \qquad (3.22)$$

We define $c_8 = \max\{1, 2\epsilon_1\}$. Further, we choose M_1, N_1 such that

$$\rho < \min\left\{\frac{\gamma}{c_1 + \frac{8(c_1+c_6)}{3(\gamma-\sigma)}c_1c_8}, \frac{\gamma - \sigma}{4c_1}, \frac{3(\gamma-\sigma)^2}{64c_8(c_6+c_1)^2}, \frac{3(\gamma-\sigma)(R - ||x_*||)}{16(c_6+c_1)c_8}\right\} \qquad (3.23)$$

$\forall M \geq M_1, N \geq N_1$, where c_1, c_6 are defined by Lemma 3.1,3.2. From Theorem 2.1 we obtain, that the sequence $\{x_\ell\}_{\mathbb{N}}$, generated by the Gauss–Newton Method with starting value $x_0 \in B_R(0) \cap B_{\epsilon_1}(x_0)$, converges q–linearly to x_*. Especially, we get $||x_{\ell+1} - x_\ell|| \leq 2\epsilon_1$. Define

$$\tau(M, N) \quad := \quad \left(\frac{\gamma - \sigma}{4(c_1 + c_6)} + \sqrt{\frac{(\gamma-\sigma)^2}{16(c_1+c_6)^2} - \frac{2c_1\epsilon_1 + c_6}{c_1 + c_6}\rho}\right)^{-1} \frac{2c_1\epsilon_1 + c_6}{c_1 + c_6}\rho.$$

From $\rho < 3(\gamma-\sigma)^2/(64c_8(c_6+c_1)^2)$ we obtain

$$\tau(M,N) \leq \frac{8(c_6+c_1)}{3(\gamma-\sigma)} c_8\rho\,. \tag{3.24}$$

Now the theorem can be proven by induction. We present the induction step. Assume, that $||x_\ell - x_\ell^{MN}|| \leq \tau := \tau(M,N)$.
From the definition of ϵ_1 is follows, that

$$\begin{aligned} ||x_{\ell+1}|| &\leq ||x_*|| + ||x_{\ell+1} - x_*|| \\ &\leq ||x_*|| + ||x_0 - x_*|| \quad < \quad R \end{aligned}$$

This shows, that $x_{\ell+1} = x_\ell(0)$. Therefore Lemma 3.2 yields

$$\begin{aligned} &||x_{\ell+1} - x_\ell^M(0)|| \\ \leq\ & \frac{c_6||x_\ell - x_\ell^M||^2 + c_1(\rho + ||x_\ell - x_\ell^{MN}||)||x_{\ell+1} - x_\ell|| + \sigma||x_\ell - x_\ell^{MN}|| + c_6\rho}{\gamma - c_1\rho - c_1||x_\ell - x_\ell^{MN}||} \\ \leq\ & \frac{c_6||x_\ell - x_\ell^{MN}||^2 + 2c_1(\rho + ||x_\ell - x_\ell^{MN}||)\epsilon_1 + \sigma||x_\ell - x_\ell^{MN}|| + c_6\rho}{\gamma - c_1\rho - c_1||x_\ell - x_\ell^{MN}||}\,. \end{aligned}$$

The denominator is greater than 0, since ρ is chosen less then $\gamma/(c_1 + \frac{8(c_1+c_6)}{3(\gamma-\sigma)}c_1c_8)$. Therefore the terms on the right hand side are well defined. From $\epsilon_1 \leq (\gamma-\sigma)/(8c_1)$ and $\rho < (\gamma-\sigma)/(4c_1)$ we find, that $\sigma - \gamma + 2c_1\epsilon_1 + c_1\rho < (\sigma-\gamma)/2$. Hence

$$||x_{\ell+1} - x_\ell^{MN}(0)|| \leq \frac{c_6\tau^2 + \frac{\sigma-\gamma}{2}\tau + 2c_1\rho\epsilon_1 + c_6\rho + (\gamma - c_1\rho)\tau}{\gamma - c_1\rho - c_1\tau} = \tau\,. \tag{3.25}$$

The last equality follows from that fact, that τ is the smallest root of

$$(c_1+c_6)\tau^2 + \frac{\sigma-\gamma}{2}\tau + 2c_1\rho\epsilon_1 + c_6\rho = 0\,.$$

(3.22), (3.24), (3.25) and $\rho \leq 3(\gamma-\sigma)(R - ||x_*||)/(16(c_6+c_1)c_8)$ yield

$$||x_\ell^{MN}(0)|| < ||x_*|| + ||x_{\ell+1} - x_*|| + ||x_{\ell+1} - x_\ell^{MN}(0)|| \leq ||x_*|| + \frac{1}{2}(R - ||x_*||) + \tau < R\,.$$

This shows, that $x_{\ell+1}^{MN} = x_\ell^{MN}(0)$. Now the assertion follows from (3.24) and (3.25).

(ii) To proof the general case we proceed as follows. In the first step we will use Lemma 3.3 to derive the bound for the Lagrange multipliers. This requires an estimate for $||x_\ell(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1})||$ to control the first term on the right hand side of (3.12). In the second step we will use Lemma 3.2 and perturbation results for linear equations together with the estimate for the Lagrange multipliers to verify the bound for the iterates.
Define

$$\gamma := \max\{\frac{\gamma_* + \mu_*}{\alpha} - \mu_*, 0\}\,,$$

It was shown in the proof of Theorem 2.1, that for $x \in B_{\epsilon_*}(x_*) \cap B_R(0)$ the following inequality is valid:

$$||F'(x)h||^2 \geq \gamma||h||^2\,. \quad \forall h \in X$$

Moreover α and ϵ_* were chosen such that $\sigma < \gamma + \mu_*$.
Let $\delta \in (0, \gamma + \mu_* - \sigma)$ be arbitrary. In this part of the proof δ will play the role which was played by $\gamma - \sigma$ in the first part. Set

$$\epsilon_2 := \min\{\epsilon_*, \left(\frac{\theta c}{\gamma_* + \mu_*}(\kappa \epsilon_* L_1/2 + \sigma)\right)^{-1} \frac{\delta}{8}, \frac{\delta}{8c_1}\}$$

and $c_9 = \max\{1, 2\epsilon_2\}$. Since the error in the Lagrange multipliers are dominated by the error in the iterates, we obtain from (2.10) in Theorem 2.1, that for $||x_* - x_0|| < \epsilon_2$

$$\begin{aligned} |\mu_* - \mu_{\ell+1}| &\leq \frac{\theta\alpha}{\gamma_* + \mu_*}(\kappa \epsilon_* L_1/2 + \sigma)||x_* - x_\ell|| \\ &\leq \frac{\theta\alpha}{\gamma_* + \mu_*}(\kappa \epsilon_* L_1/2 + \sigma)\epsilon_2 < \frac{\delta}{8}. \end{aligned}$$

Define M_2, N_2 such that

$$\rho \;<\; \min\left\{\frac{\gamma - \mu_* - \delta/8}{c_1 + c_1 \frac{8(c_1+c_6)}{3\delta} c_9}, \frac{\delta}{8c_1}, \frac{3\delta^2}{64 c_9 (c_6 + c_1)^2}\right\} \tag{3.26}$$

$\forall M \geq M_2, N \geq N_2$ and set

$$\tau_1(M,N) \;:=\; \left(\frac{\delta}{4(c_1+c_6)} + \sqrt{\frac{\delta^2}{16(c_1+c_6)^2} - \frac{2c_1\epsilon_2 + c_6}{c_1 + c_6}\rho}\right)^{-1} \frac{2c_1\epsilon_2 + c_6}{c_1 + c_6}\rho.$$

Then we obtain

$$\tau_1(M,N) \leq \frac{8(c_1+c_6)}{3\delta} c_9\, \rho. \tag{3.27}$$

With these arrangements, we obtain similar to the calculations in (i), that ($\tau_1 := \tau_1(M,N)$)

$$\begin{aligned} &||x_\ell(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1})|| \\ &\leq \frac{c_6||x_\ell - x_\ell^{MN}||^2 + 2c_1(\rho + ||x_\ell - x_\ell^{MN}||)\epsilon_2 + \sigma||x_\ell - x_\ell^{MN}|| + c_6\rho}{\gamma + \mu_* - |\mu_{\ell+1} - \mu_*| - c_1\rho - c_1||x_\ell - x_\ell^{MN}||} \\ &\leq \frac{c_6\tau_1^2 - \frac{\delta}{2}\tau_1 + 2c_1\rho\epsilon_2 + c_6\rho + (\gamma + \mu_* - |\mu_{\ell+1} - \mu_*| - c_1\rho)\tau_1}{\gamma + \mu_* - |\mu_{\ell+1} - \mu_*| - c_1\rho - c_1\tau_1} && (3.28) \\ &= \tau_1, && (3.29) \end{aligned}$$

provided $||x_\ell - x_\ell^{MN}|| \leq \tau_1(M,N)$.

Since $||x_\ell(\mu_{\ell+1})|| \leq R$ we obtain from (3.29) that $||x_\ell^{MN}(\mu_{\ell+1})||$ is bounded. Therefore there exists $\tilde{c}_7$, independent of M, N, such that $||x_\ell^{MN}(\mu_{\ell+1})|| \leq \tilde{c}_7$ and $c_7(1 + ||x_\ell^{MN}(\mu_{\ell+1})||) \leq \tilde{c}_7$, where c_7 is defined as in lemma 3.3. If we choose $M_3 \geq M_2, N_3 \geq N_2$ such that

$$\rho \leq \left(\frac{8(c_1+c_6)}{3\delta}c_9\right)^{-1} 2R \qquad \forall M \geq M_3, N \geq N_3, \tag{3.30}$$

(3.27) and (3.29) yield

$$\frac{\tilde{c}_7}{(1-\|x_\ell(\mu_{\ell+1})-x_\ell^{MN}(\mu_{\ell+1})\|/R)^2} \le 4\tilde{c}_7\,, \tag{3.31}$$

provided $M \ge M_3, N \ge N_3$ and $\|x_\ell - x_\ell^M\| \le \tau_1(M,N)$.

Since $|\mu_{\ell+1} - \mu_{\ell+1}^{MN}|$ is bounded by the same term as $\|x_\ell(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1})\|$ (up to the constant $\tilde{c}_7/(1-\|x_\ell(\mu_{\ell+1})-x_\ell^{MN}(\mu_{\ell+1})\|/R)^2)$) we obtain from (3.12), (3.31) and (3.29), that $\|x_\ell - x_\ell^{MN}\| \le \tau_1(M,N)$ implies

$$|\mu_{\ell+1} - \mu_{\ell+1}^{MN}| \le 4\tilde{c}_7\tau_1(M,N)\,. \tag{3.32}$$

Together with the 3.27 this gives the desired estimate for the Lagrange multipliers. To prove the estimate for the iterates, we have to combine the previous results. Lemma (3.2) yields

$$\begin{aligned}
&\|x_{\ell+1} - x_{\ell+1}^{MN}\| \\
&\quad = \|x_\ell(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1}^{MN})\| \\
&\quad \le \|x_\ell(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1})\| + \|x_\ell^{MN}(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1}^{MN})\| \\
&\quad \le \frac{c_6\|x_\ell - x_\ell^{MN}\|^2 + c_1(\rho + \|x_\ell - x_\ell^{MN}\|)\|x_{\ell+1} - x_\ell\| + \sigma\|x_\ell - x_\ell^{MN}\| + c_6\rho}{\gamma + \mu_* - |\mu_{\ell+1} - \mu_*| - c_1\rho - c_1\|x_\ell - x_\ell^{MN}\|} \\
&\qquad + \|x_\ell^{MN}(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1}^{MN})\|
\end{aligned} \tag{3.33}$$

If $A, \tilde{A} \in L(X,X)$ are continuously invertible with $\|A^{-1}\|\,\|A-\tilde{A}\| < 1$, then

$$\|A^{-1}b - \tilde{A}^{-1}b\| \le \frac{\|A^{-1}\|\,\|A-\tilde{A}\|}{1-\|A^{-1}\|\,\|A-\tilde{A}\|}\|A^{-1}b\|.$$

Together with (3.11) this yields

$$\begin{aligned}
&\|x_\ell^{MN}(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1}^{MN})\| \\
&\quad \le \frac{\|(F_N'(x_\ell^{MN})^*F_N'(x_\ell^{MN}) + \mu_{\ell+1}I)^{-1}\|\,|\mu_{\ell+1} - \mu_{\ell+1}^{MN}|\,\|x_\ell^{MN}(\mu_{\ell+1})\|}{1 - \|(F_N'(x_\ell^{MN})^*F_N'(x_\ell^{MN}) + \mu_{\ell+1}I)^{-1}\|\,|\mu_{\ell+1} - \mu_{\ell+1}^{MN}|} \\
&\quad \le \frac{|\mu_{\ell+1} - \mu_{\ell+1}^{MN}|\,\tilde{c}_7}{\gamma + \mu_* - |\mu_{\ell+1} - \mu_*| - c_1\rho - c_1\|x_\ell - x_\ell^{MN}\| - |\mu_{\ell+1} - \mu_{\ell+1}^{MN}|}
\end{aligned}$$

Define $c_{10} = c_1 + 4\tilde{c}_7\frac{8(c_1+c_6)}{3\delta}c_9$, then we conclude with (3.27) and (3.32)

$$\|x_\ell^{MN}(\mu_{\ell+1}) - x_\ell^{MN}(\mu_{\ell+1}^{MN})\| \le \frac{4\tilde{c}_7^2\rho}{\gamma + \mu_* - |\mu_{\ell+1} - \mu_*| - c_{10}\rho - c_1\|x_\ell - x_\ell^{MN}\|}\,, \tag{3.34}$$

provided $M \ge M_3, N \ge N_3$ and $\|x_\ell - x_\ell^{MN}\| \le \tau_1(M,N)$. If we insert (3.34) into (3.33), we observe that $\|x_{\ell+1} - x_{\ell+1}^{MN}\|$ is bounded by a term which has the same structure than the bound in (3.28) (replace in c_1 by c_{10} and c_6 by $c_{11} := c_6 + 4\tilde{c}_7^2$). Therefore, with the choices $M_1 \ge M_3, N_1 \ge N_3$ such that

$$\rho \le \frac{\delta^2}{64c_9(c_{10}+c_{11})} \quad \forall M \ge M_1, N \ge N_1$$

and

$$\tau(M,N) := \min\left\{\tau_1(M,N), \left(\frac{\delta}{4(c_{10}+c_{11})} + \sqrt{\frac{\delta^2}{16(c_{10}+c_{11})^2} - \frac{2c_{10}\epsilon_2 + c_{11}}{c_{10}+c_{11}}\rho}\right)^{-1} \frac{2c_{10}\epsilon_2 + c_{11}}{c_{10}+c_{11}}\rho\right\}$$

we finally obtain that $||x_\ell - x_\ell^{MN}|| \le \tau(M,N)$ implies $||x_{\ell+1} - x_{\ell+1}^{MN}|| \le \tau(M,N)$, which gives the assertion, since

$$\tau(M,N) \le \frac{8(c_{10}+c_{11})}{3\delta} c_9\,\rho\,.$$

□

To guarantee that the error between x_ℓ and x_ℓ^{MN} could be bounded by $\rho_X(1/M) + \rho_Y(1/N)$, we have to ensure that the starting point x_0^{MN} satisfies a certain approximation property, which is essentially $||x_0 - x_0^{MN}|| \le O(\rho_X(1/M) + \rho_Y(1/N))$. However, if the starting point for the infinite dimensional problem satisfies $x_0 \in X_M \quad \forall M$, we can choose $x_0^M = x_0$ for all M (and N). In this case we always have $||x_0 - x_0^{MN}|| \le \tau(M,N)$. Such situations occur for example if $X_M = \mathrm{span}\{\phi_1, \ldots, \phi_M\}$, where ϕ_i are splines and x_0 is a constant function.

The advantage of this approach is that we obtain uniform bounds between the infinite dimensional iterates x_ℓ and the corresponding finite dimensional x_ℓ^{MN}, whereas in the setting of [2] we would obtain uniform bounds between the restriction of the infinite dimensional iterates onto the finite dimensional space, $\Delta^M x_\ell$ and the iterates x_ℓ^{MN}. In the case of finite element discretizations, with $X = H^s$, Δ^M the spline interpolant, this would lead to estimates of the form (see [4] p.217)

$$\begin{aligned} ||x_\ell - x_\ell^{MN}||_{H^s} &\le ||x_\ell - \Delta^M x_\ell||_{H^s} + ||\Delta^M x_\ell - x_\ell^{MN}||_{H^s} \\ &\le c\frac{1}{M^{k+1-s}}||x_\ell||_{H^{k+1}} + c(\rho_X(1/M) + \rho_Y(1/N)). \end{aligned}$$

This bound involves the H^{k+1}–norm of x_ℓ, and therefore leads only to a pointwise estimate, since $||x_\ell||_{H^{k+1}}$ may not be bounded.

An immediate consequence of this mesh independent behavior is the fact, that independent of the meshsize an (almost) constant number of iterations is needed to satisfy an appropriate stopping criteria. Appropriate stopping criteria for the restricted Gauss–Newton method are either

$$\begin{aligned} ||x_\ell - P(x_\ell - F'(x_\ell)^* F(x_\ell))|| &< \mathrm{TOL} \qquad \text{or} \\ ||F'(x_\ell)^* F(x_\ell) + \mu_\ell x_\ell|| &< \mathrm{TOL}\,, \end{aligned}$$

where TOL is a given bound and P denotes the projection onto the feasible set. In our case

$$P(y) = \begin{cases} \frac{R}{||y||}y & \text{if } ||y|| > R \\ y & \text{else} \end{cases}.$$

If the iteration point x_ℓ is an interior point, both criteria reduces to $||F'(x_\ell)^* F(x_\ell)|| < \mathrm{TOL}$. We will use the abbreviation

$$t_\ell := ||x_\ell - P(x_\ell - F'(x_\ell)^* F(x_\ell))|| \qquad \text{or} \tag{3.35}$$

$$t_\ell := ||F'(x_\ell)^* F(x_\ell) + \mu_\ell x_\ell||\,, \tag{3.36}$$

depending on which criteria is used. With t_ℓ^{MN} we will denote the corresponding discretized values. We use the same notation for both terms, since we have the same type of estimates for $|t_\ell - t_\ell^{MN}|$ no matter if (3.35) or (3.36) is used. $\ell(\text{TOL})$ and $\ell^{MN}(\text{TOL})$ will be defined to be the smallest iteration counts for which the termination criteria is satisfied, i.e.

$$\begin{aligned} \ell(\text{TOL}) &:= \min\{\ell \mid t_\ell < \text{TOL}\} \\ \ell^{MN}(\text{TOL}) &:= \min\{\ell \mid t_\ell^{MN} < \text{TOL}\}. \end{aligned}$$

Now the uniform estimate, derived in Theorem 3.4 yields

Corollary 3.5 *Let the assumptions of Theorem (3.4) hold. If x_0 and x_0^M are given such that $x_0 \in B_{\epsilon_1}(x_*)$ and $||x_0 - x_0^{MN}|| \le \tau(M,N)$, ($\epsilon_1$ and $\tau(M,N)$ defined as in Theorem (3.4),) then for every $\text{TOL} > 0$ and $\delta > 0$ there exist M_2, N_2 such that*

$$\ell(\text{TOL} + \delta) \le \ell^{MN}(\text{TOL}) \le \ell(\text{TOL}) \qquad \forall M \ge M_2, N \ge N_2 .$$

If $t_{\ell(\text{TOL})-1} > \text{TOL}$ we obtain

$$\ell^{MN}(\text{TOL}) = \ell(\text{TOL}) \qquad \forall M \ge M_2, N \ge N_2 .$$

Proof: In the proof of Theorem (3.4) it was shown, that under the assumptions listed above $||x_\ell - x_\ell^{MN}|| \le c(\rho_X(1/M) + \rho_Y(1/N))$ for all ℓ and $M \ge M_1, N \ge N_1$. This yields, that there exists $\tilde{c}$, independent of M, N such that

$$|t_\ell - t_\ell^{MN}| \le \tilde{c}(\rho_X(1/M) + \rho_Y(1/N)) \quad \forall M \ge M_1, N \ge N_1 .$$

If we choose M_2, N_2 such that

$$|t_{\ell(\text{TOL})} - t_{\ell(\text{TOL})}^{MN}| < \text{TOL} - t_{\ell(\text{TOL})} \quad \forall M \ge M_2, N \ge N_2 .$$

and

$$|t_\ell - t_\ell^{MN}| < \delta \quad \forall \ell, M \ge M_2, N \ge N_2 .$$

we obtain

$$t_{\ell(\text{TOL})}^{MN} \le t_{\ell(\text{TOL})} + |t_{\ell(\text{TOL})} - t_{\ell(\text{TOL})}^{MN}| < \text{TOL} \quad \forall M \ge M_2, N \ge N_2$$

and

$$t_{\ell^{MN}(\text{TOL})} \le t_{\ell^{MN}(\text{TOL})}^{MN} + |t_{\ell^{MN}(\text{TOL})} - t_{\ell^{MN}(\text{TOL})}^{MN}| < \text{TOL} + \delta \quad \forall M \ge M_2, N \ge N_2$$

If $t_{\ell(\text{TOL})-1} > \text{TOL}$ we can choose $\delta = 1/2(t_{\ell(\text{TOL})-1} - \text{TOL})$. This yields $\ell(\text{TOL} + \delta) = \ell(\text{TOL})$. Hence the assumption is proven. □

We conclude this section with results on the convergence rate of the Gauss–Newton method for the discretized problem and on perturbations of solutions and Lagrange multipliers. In addition to the assumptions (A1)–(A7) we need an assumption on the curvature of F and F_N:

(A8) There exists a sequence $\{\xi_{MN}\}, \xi_{MN} \to 0 (M, N \to \infty)$, such that for all $x, y \in B_R(0) \cap X_M$

$$||((F_N'(x)^* - F_N'(y)^*) - (F'(x)^* - F'(y)^*))F_N(y)|| \le \xi_{MN} ||x - y||$$

In the following theorem we will use the notation of Theorem 2.1 and its proof.

Theorem 3.6 *Let (A1)-(A8) and the assumptions of Theorems 2.1, and 3.4 hold. Then for all* $\alpha \in (1, (\gamma_*+\mu_*)/\sigma)$ *and all* $\epsilon \in (0, \epsilon_*(\alpha))$ *there exist* M_ϵ *and* N_ϵ *such that for all* $M \geq M_\epsilon$, $N \geq N_\epsilon$ *and all* $x_0^{MN} \in B_R(0) \cap B_\epsilon(x_*)$ *the Gauss–Newton method for the discretized problem with staring point* x_0^{MN} *converges to a solution* x_*^{MN} *of (1.3). Moreover,* x_*^{MN} *is the unique minimizer of (1.3) in* $B_\epsilon(x_*)$ *and the convergence rate is given by*

$$\begin{aligned} \|x_{\ell+1}^{MN} - x_*^{MN}\| &\leq \frac{\alpha\sigma^{MN}}{\gamma_*^{MN} + \mu_*^{MN}}\|x_\ell^{MN} - x_*^{MN}\| + \frac{\alpha L_1^{MN}\kappa^{MN}}{2(\gamma_*^{MN} + \mu_*^{MN})}\|x_\ell^{MN} - x_*^{MN}\|^2 \\ &< \|x_\ell^{MN} - x_*^{MN}\|, \end{aligned}$$

where $\kappa^{MN} := \sup_{x \in B_R(0)} \|F_N'(x)\|$ *and* $\{\gamma_*^{MN}\}_{\mathbb{N}}$, $\{\sigma^{MN}\}_{\mathbb{N}}$ *are sequences with*

$$|\gamma_*^{MN} - \gamma_*| = O(\rho_X(1/M) + \rho_Y(1/N)),\ |\sigma^{MN} - \sigma| = O(\xi_{MN} + \rho_X(1/M) + \rho_Y(1/N)).$$

The errors between x_* *and* x_*^{MN} *and between the Lagrange multipliers can be estimated by*

$$\begin{aligned} \|x_* - x_*^{MN}\| &\leq c(\rho_X(\frac{1}{M}) + \rho_Y(\frac{1}{N})), \\ |\mu_*^{MN} - \mu_*| &\leq c(\rho_X(\frac{1}{M}) + \rho_Y(\frac{1}{N})), \end{aligned} \tag{3.37}$$

where $c > 0$ *denotes a generic constant.*

Proof: We only give a sketch of the proof.
Theorem 2.4 yields, that there exists a sequence $\{x_*^{MN}\}_{\mathbb{N}}$ of minimizers of (1.3) such that $x_*^{MN} \to x_* \quad (M, N \to \infty)$. Theorem 3.4 yields the error estimate

$$\|x_* - x_*^{MN}\| \leq 2c(\rho_X(\frac{1}{M}) + \rho_Y(\frac{1}{N})),$$

since (3.20) holds for all ℓ.

From (2.6), (A5) and (A8) we obtain, that for all $x, y \in B_R(0)$

$$\begin{aligned} &\|(F_N'(x)^* - F_N'(y)^*)F_N(y)\| \\ &\leq \|(F'(x)^* - F'(y)^*)F(y)\| + \|(F'(x)^* - F'(y)^*)(F(y) - F_N(y))\| \\ &\quad + \|((F_N'(x)^* - F_N'(y)^*) - (F'(x)^* - F'(y)^*))F_N(y)\| \\ &\leq \sigma\|x-y\| + \rho_Y(1/N)L_1\|x-y\| + \xi_{MN}\,\|x-y\| \end{aligned}$$

Hence there exist $\sigma^{MN} \geq \sigma$ such that $|\sigma^{MN} - \sigma| = O(\xi_{MN} + \rho_X(1/M) + \rho_Y(1/N))$ and

$$\|(F_N'(x)^* - F_N'(y)^*)F_N(y)\| \leq \sigma^{MN}\|x-y\|.$$

(A4) and (2.5) yield

$$\begin{aligned} \|F_N'(x_*^{MN})h_M\|^2 &\geq (\|F'(x_*^{MN})h_M\| - \|F_N'(x_*^{MN})h_M - F'(x_*^{MN})h_M\|)^2 \\ &\geq (\sqrt{\gamma_*}\|h_M\| - \rho_Y(1/N)\|h_M\|)^2. \end{aligned}$$

Hence there exist a sequence $\{\gamma_*^{MN}\}_N$, $\gamma_*^{MN} = \gamma_* + O(\rho_X(1/M) + \rho_Y(1/N))$ such that

$$||F_N'(x_*^{MN})h_M||^2 \geq \gamma_*^{MN}||h_M||^2 \quad \forall h_M \in X_M \, .$$

If we denote the Lagrange multiplier corresponding to x_*^{MN} by μ_*^{MN}, one can show similar to Lemma 3.3,3.2 (note that $x_*(\mu_*) = x_*$ and use (3.37)), that for sufficiently large M, N there exists c independent of M, N such that

$$|\mu_* - \mu_*^{MN}| \leq c(\rho_X(1/M) + \rho_Y(1/N)) \, .$$

These preliminaries show, that we can choose $\bar{M}, \bar{N}$ such that

$$\alpha \in \left(1, \frac{\gamma_*^{MN} + \mu_*^{MN}}{\sigma^{MN}}\right) \quad \forall M \geq \bar{M}, N \geq \bar{N} \, .$$

If we apply Theorem 2.1 to x_*^{MN}, we obtain the existence of ϵ_*^{MN} such that the Gauss–Newton method for the discretized problem with arbitrary starting point $x_0^{MN} \in B_R(0) \cap B_{\epsilon_*^{MN}}(x_*^{MN})$ converges to x_*^{MN}:

$$\begin{aligned} ||x_{\ell+1}^{MN} - x_*^{MN}|| &\leq \frac{\alpha\sigma^{MN}}{\gamma_*^{MN} + \mu_*^{MN}}||x_\ell^{MN} - x_*^{MN}|| + \frac{\alpha L_1^{MN}\kappa^{MN}}{2(\gamma_*^{MN} + \mu_*^{MN})}||x_\ell^{MN} - x_*^{MN}||^2 \\ &< ||x_\ell^{MN} - x_*^{MN}|| \, . \end{aligned}$$

Moreover, the proof of Theorem 2.1 shows, that $\epsilon_*^{MN} \to \epsilon_*$.
The uniqueness of the solution x_*^{MN} follows from the fact, that the Gauss–Newton method with arbitrary starting point $x_0 \in B_\epsilon(x_*)$ converges towards x_*^{MN}. □

If F and F_N are twice Fréchet differentiable, a sufficient condition for (A8) to hold is

$$||F_N''(y)^* F_N(x) - F''(y)^* F_N(x)||_{L(X,X)} \leq (\rho_X(1/M) + \rho_Y(1/N)) \quad \forall x, y \in B_R(0) \, .$$

Since $F_N''(y)^*$ is applied to an element of Y_N, it could be the ordinary Y_N, X_M adjoint, $F_N''(y)^* \in L(Y_N, X_M \otimes X_M)$. In this case we obtain $\xi_{MN} = \rho_X(1/M) + \rho_Y(1/N)$ and $\sigma^{MN} = O(\rho_X(1/M) + \rho_Y(1/N))$.

4 Examples

In this section we will demonstrate, how the analysis of the previous sections can be applied to a certain parameter identification problems. Although we are considering the one dimensional problem, it should be mentioned , that our analysis can be extended to the multidimensional case. The parameter identification problem for the two point boundary value problem can be stated as follows:

For a given observation $z \in L^2(0,1)$ or $H_0^1(0,1)$ find $q \in H^1(0,1)$ with $||q||_{H^1(0,1)} \leq R$ and $q(x) \geq \gamma > 0$ a.e. on $(0,1)$, such that

$$u(q) \approx z.$$

Here $u(q) \in H_0^1(0,1)$ is defined to be the weak solution of the state equation

$$\begin{aligned} -(qu')' &= f \qquad \text{in } (0,1) \quad , \\ u(0) = u(1) &= 0 \end{aligned}$$

with $f \in L^2(0,1)$, i.e. $u(q)$ is defined through

$$< qu', v' > = < f, v > \quad \forall v \in H_0^1(0,1)\,. \tag{4.1}$$

(For the rest of the chapter we will drop the notation of the space $(0,1)$ and we will always use the notation $< \cdot, \cdot >$ for the L^2 - scalarproduct.) It is well known, that (4.1) always possess a solution $u(q) \in H_0^1$ and since $q \in H^1$, $f \in L^2$ one can even show, that $u(q) \in H_0^1 \cap H^2$ with

$$||u(q)||_{H^2} \leq c||f||_{L^2}\,, \tag{4.2}$$

where c is a constant depending on γ and R (see e.g. [5], p. 223). In the sequel, we will denote by $u(q)$ the solution of (4.1). For the solution of the parameter identification problem described above, we have to specify '$\approx$'. Here we will investigate the Output Least Squares formulation, i.e. we seek solutions of

$$\min_{\substack{||q||_{H^1} \leq R \\ q(x) \geq \gamma \text{ a.e. on } (0,1)}} ||u(q) - z||_{\mathcal{Z}}\,, \tag{4.3}$$

where $\mathcal{Z} = \mathcal{L}^{\epsilon}(\prime, \infty)$ or $H_0^1(0,1)$. It is well known, that (4.3) may be ill–posed in the sense, that small perturbations in the observation z may lead to large errors in the solution q of (4.3). In order to get a stable problem, for which it is possible to estimate the error between the computed solution of problem (4.3) with perturbed data z and the true, but unknown solution corresponding to the unperturbed data, one has to modify the problem. A possible remedy to remove this difficulty is the Tikhonov regularization. Here one adds a regularization term to the objective, so that (4.3) changes to

$$\min_{\substack{||q||_{H^1} \leq R \\ q(x) \geq \gamma \text{ a.e. on } (0,1)}} ||u(q) - z||_{\mathcal{Z}}^2 + \alpha^2 ||q||_{H^1}^2\,. \tag{4.4}$$

The Tikhonov regularization for nonlinear problems was studied by many authors (see e.g. [5], [16], [8], [9], [12]). In the following we assume, that q_* is a solution of (4.4) which satisfies $q_*(x) > \gamma$ a.e. on $(0,1)$. Since $||\cdot||_{H^1}$ dominates the infinity–norm and since we are doing a local analysis, we may drop the constraint '$q(x) > \gamma$ a.e. on $(0,1)$'. In the sequel it will always implicitely be assumed, that the considered parameter functions q $(, q_1, q_2, \ldots)$ are satisfying this constraint. In this case (4.4) fits our framework, if we set

$$X = H^1, Y = \mathcal{Z} \times H^1 \quad \text{(endowed with the product topology)}$$

and

$$F(q) = \begin{pmatrix} u(q) - z \\ \alpha q \end{pmatrix}.$$

(In this chaper we follow the conventional notation in parameter identification and denote the sought variable by q, whereas $x \in (0,1)$ denotes the space variable!) It can be shown, that F is infinitely often Fréchet differentiable. The first Fréchet derivative is given by

$$F'(q)h = \begin{pmatrix} \eta \\ \alpha h \end{pmatrix},$$

where $\eta = u_q(q)(h)$ is the solution of

$$< q\eta', v' >= - < hu', v' > \quad \forall v \in H_0^1 . \tag{4.5}$$

The variational equation

$$< q\xi', v' >= - < h_1\eta_2', v' > - < h_2\eta_1', v' > \quad \forall v \in H_0^1 , \tag{4.6}$$

where η_i is the solution of (4.5) with h_i instead of h, characterizes the second Fréchet derivative of F, which is given as

$$F''(q)(h_1, h_2) = \begin{pmatrix} \xi \\ 0 \end{pmatrix} .$$

For the numerical solution of (4.4) we choose piecewise linear splines. Let φ_i^M, ψ_j^N be the functions defined by

$$\varphi_i^M(x) = \begin{cases} M(x - \frac{i-1}{M}) & x \in [\frac{i-1}{M}, \frac{i}{M}] \\ M(\frac{i+1}{M} - x) & x \in (\frac{i}{M}, \frac{i+1}{M}] \\ 0 & \text{otherwise} \end{cases} , \quad \psi_j^N(x) = \begin{cases} N(x - \frac{j-1}{N}) & x \in [\frac{j-1}{N}, \frac{j}{N}] \\ N(\frac{j+1}{N} - x) & x \in (\frac{j}{N}, \frac{j+1}{N}] \\ 0 & \text{otherwise} \end{cases}$$

We set $X_M := span\{\varphi_0^M, \ldots, \varphi_M^M\}$, $V_N := span\{\psi_1^N, \ldots, \psi_{N-1}^N\}$ and $Y_N := V_N \times X_M$. The discretized solution of the state equation is given as the uniquely determined element $u^N = u^N(q)$ which satisfies

$$< qu^{N'}, v^{N'} >=< f, v^N > \quad \forall v^N \in V_N . \tag{4.7}$$

Now we choose the discretization of F as follows:

$$F_N(q) = \begin{pmatrix} u^N(q) - z^N \\ \alpha q \end{pmatrix} ,$$

where z^N is a discretization of z, for example the spline interpolant.

The Fréchet derivative of $u^N(q)$, $\eta^N := u_q^N(q)(h)$, is given as the unique solution of

$$< q\eta^{N'}, v^{N'} >=< hu^{N'}, v^{N'} > \quad \forall v^N \in V_N . \tag{4.8}$$

The second Fréchet derivative is given analogously to (4.6). This especially proves the validity of (A1) and (A3).

In the following we will denote by $u(q)$ the solution of (4.1) and by $u^N(q)$ its discretization, i.e. the solution of (4.7), for a given parameter function q. And we will use a similar notation for the Fréchet derivatives.

We will now verify, that F and its discretization satisfies the assumptions (A2) and (A4). Since $u(q_1) - u(q_2)$ satisfies the variational equation

$$< q_1(u(q_1) - u(q_2))', v' >=< (q_1 - q_2)u(q_2)', v' > \quad \forall v \in H_0^1$$

we immediately obtain from (4.2) and

$$||v||_{L^\infty} \leq c_1||v||_{H^1} \tag{4.9}$$

that

$$\begin{aligned}||u(q_1)-u(q_2)||_{H^2} &\le c||(q_1-q_2)u(q_2)'||_{L^2}\\ &\le c^2c_1||f||_{L^2}||q_1-q_2||_{H^1}\,. \end{aligned}\tag{4.10}$$

From the error analysis of finite element methods we get (see [4], p.152,217)

$$\begin{aligned}||u(q)-u^N(q)||_{H^1} &\le \sqrt{\frac{R}{\gamma}\frac{1}{\pi}}||u(q)||_{H^2}\frac{1}{N}\\ &\le \sqrt{\frac{R}{\gamma}\frac{1}{\pi}}c||f||_{L^2}\frac{1}{N}. \end{aligned}\tag{4.11}$$

Using the Aubin-Nitsche trick (see e.g. [4], p.229), this estimate can be improved for the L^2-norm to

$$||u(q)-u^N(q)||_{L^2} \le c_2\frac{1}{N^2}$$

The Fréchet derivatives $u_q(q)$ and $u_q^N(q)$ are defined through the same kind of elliptic differential equation. Therefore we can apply a similar analysis to derive continuity results for these functions. If we use the corresponding estimates to (4.2), (4.10) and inequality (4.9), we obtain

$$\begin{aligned}||u_q(q_1)-u_q(q_2)||_{H^2} &\le c_1c(||q_1-q_2||_{H^1}||u_q(q_2)(h)||_{H^1}+||h||_{H^1}||u(q_1)-u(q_2)||_{H^1})\\ &\le 2c_1^2c^3||f||_{L^2}||q_1-q_2||_{H^1}||h||_{H^1}\end{aligned}$$

Let $\zeta \in H_0^1$ denote the solution of

$$< q\zeta', v' >=< hu^{N'}, v' > \quad \forall v \in H_0^1\,.$$

Then the error between the discretized and infinite dimensional Fréchet derivative can be estimated through

$$\begin{aligned}||\eta-\eta^N||_{H^1} &\le ||\eta-\zeta||_{H^1}+||\zeta-\eta^N||_{H^1}\\ &\le c_1\frac{1}{\gamma}||h||_{H^1}||u(q)-u^N(q)||_{H^1}+\sqrt{\frac{R}{\gamma}\frac{1}{\pi}}c_1c||h||_{H^1}||u^N(q)||_{H^1}\frac{1}{N}\\ &\le 2\sqrt{\frac{R}{\gamma^3}\frac{1}{\pi}}c_1c||f||_{L^2}||h||_{H^1}\frac{1}{N}\end{aligned}$$

In the case of L^2 norms we can apply the Aubin-Nitsche trick twice and improve the L^2–bound to

$$||\eta-\eta^N||_{L^2} \le c_3\frac{1}{N^2}$$

The above techniques can obviously be applied to the second and even higher Fréchet derivatives. The calculations above show, that F satisfies the assumptions (A2) and (A4) for $\mathcal{Z} = H_0^1$ with $\rho_Y(1/N) = c_Y\,1/N$, provided $||z-z^N||_{H_0^1} \le c\,1/N$ and for $\mathcal{Z} = L^2$ with $\rho_Y(1/N) = c_Y\,1/N^2$, provided $||z-z^N||_{L^2} \le c\,1/N^2$.

Now we will investigate the computation of the adjoint of F'. From the structure of F' it is obvious, that it is sufficient to study the calculation of $(u_q(q))^*$. The adjoint of $u_q(q)$ applied to $g \in L^2$ can be computed in two steps

(1) Solve the adjoint equation for given q and g

$$< qw', v' >=< g, v > \quad \forall v \in H_0^1 \tag{4.12}$$

(2) Transition from the L^2 to the H^1 topology

$$< p, \varphi > + < p', \varphi' >=< u(q)'w', \varphi > \quad \forall \varphi \in H^1 \tag{4.13}$$

(In our example the adjoint equation is just the state equation, since the differential operator $D_x q(D_x \cdot)$ is formally selfadjoint.) If we solve the two equations, we obtain $p = (u_q(q))^*(g)$, which can be seen, if we set $v = u_q(q)(\varphi)$ in (4.12):

$$\begin{aligned} < g, u_q(q)(\varphi) > &= < qw', u_q(q)(\varphi) > \\ &= < w', q\, u_q(q)(\varphi) > \\ &= < \varphi, u(q)'w' > \quad = \quad < p, \varphi >_{H^1} \end{aligned}$$

(for the third equality we used the definition of the Fréchet derivative, see (4.5) with v replaced by w). The variational equation (4.13) is the weak formulation of the elliptic problem

$$p'' + p = u(q)'w' \quad \text{in } (0,1)$$

with Neumann boundary conditions

$$\frac{\partial p}{\partial n}(0) = \frac{\partial p}{\partial n}(1) = 0\,.$$

(4.13) yields

$$- < p', \varphi' >=< p, \varphi > - < u(q)'w', \varphi > \quad \forall \varphi \in C_0^\infty\,,$$

which shows, that p'' exists and equals $p - u(q)'w'$. Especially we obtain $p'' \in L^2$ and

$$||p''||_{L^2} \leq ||p||_{L^2} + c_1 ||u(q)||_{H^2}\, ||w||_{H^1}\,.$$

The Lax-Milgram Theorem and (4.9) yield

$$||p||_{H^1} \leq ||u(q)'w'||_{L^2} \leq ||u(q)'||_{L^\infty}\, ||w||_{H^1} \leq c_1 ||u(q)||_{H^2}\, ||w||_{H^1}\,.$$

Hence we obtain, that the weak solution of the Neumann problem obeys the regularity property $p \in H^2$ and

$$\begin{aligned} ||p||_{H^2} &\leq 2c_1 ||u(q)||_{H^2}\, ||w||_{H^1} \\ &\leq 2c^2 c_1 ||f||_{L^2} ||g||_{L^2}\,. \end{aligned} \tag{4.14}$$

This bound together with the techniques already applied to prove (A2) and (A4) can now be used to derive an estimate of type (A5). If we discretize the Neumann Problem (4.13) and solve

$$< \hat{p}^M, \varphi^M > + < \hat{p}^{M'}, \varphi^{M'} >=< u(q)'w', \varphi^M > \quad \forall \varphi^M \in X_M \tag{4.15}$$

the error between the solutions of (4.13) and (4.15) can be estimated by

$$||p - \hat{p}^M||_{H^1} \leq 2c_1 c^2 \frac{1}{\pi} ||f||_{L^2} ||g||_{L^2} \frac{1}{M} \tag{4.16}$$

((4.14) and [4] p.152,217). The adjoint of the discretized Fréchet derivative $u_q^N(q)$ is given through

(1) Solve the adjoint equation

$$< qw^{N'}, v^{N'} >=< g, v^N > \quad \forall v^N \in V_N \tag{4.17}$$

(2) Transition from the L^2 to the H^1 topology

$$< p^M, \varphi^M > + < p^{M'}, \varphi^{M'} >=< u^N(q)' w^{N'}, \varphi^M > \quad \forall \varphi^M \in X_M \tag{4.18}$$

At the end we obtain $p^M = (u_q^N(q))^*(g)$. The error between the infinite dimensional and discretized adjoints can be estimated by (see (4.2), (4.9), (4.11), (4.16))

$$\begin{aligned}
&||p - p^M||_{H^1} \\
\le\ & ||p - \hat{p}^M||_{H^1} + ||\hat{p}^M - p^M||_{H^1} \\
\le\ & 2c^2 c_1 \frac{1}{\pi} ||f||_{L^2} ||g||_{L^2} \frac{1}{M} + \sup_{||\varphi||_{H^1}=1} < u^N(q)' w^{N'} - u(q)' w', \varphi > \\
\le\ & 2c^2 c_1 \frac{1}{\pi} ||f||_{L^2} ||g||_{L^2} \frac{1}{M} + \sup_{||\varphi||_{H^1}=1} < w^{N'} - w', u(q)'\varphi > + < u^N(q)' - u(q)', w^{N'} \varphi > \\
\le\ & 2c^2 c_1 \frac{1}{\pi} ||f||_{L^2} ||g||_{L^2} \frac{1}{M} + c_1 ||w^{N'} - w'||_{H^1} ||u(q)||_{H^1} + c_1 ||u^N(q)' - u(q)||_{H^1} ||w^{N'}||_{H^1} \\
\le\ & 2c^2 c_1 \frac{1}{\pi} ||f||_{L^2} ||g||_{L^2} \frac{1}{M} + 2\sqrt{\frac{R}{\gamma^3}} \frac{1}{\pi} c c_1 ||f||_{L^2} ||g||_{L^2} \frac{1}{N}\,.
\end{aligned}$$

The last inequality proves, that (A5) is also valid with $\rho_X(1/M) = c_X\, 1/M$, but we have $\rho_Y(1/N) = c_Y\, 1/N$ no matter if $\mathcal{Z} = L^2$ or H_0^1.

We ran several test examples from the set of test problems in [17]. The test functions for the results we present below are given by:

Example 1
$$\begin{aligned} u(q_*) &= \sin(\pi x) \\ q_* &= 1/2 + \cos(x)\,, \quad ||q_*||^2_{H^1} = \frac{5}{4} + \sin(1) \end{aligned}$$

Example 2
$$\begin{aligned} u(q_*) &= \begin{cases} -9x^2 + 6x & x \in [0, 1/3] \\ 1 & x \in (1/3, 2/3] \\ -9x^2 + 12x - 3 & x \in (2/3, 1] \end{cases} \\ q_* &= 1/2 + \sin(\pi x)\,, \quad ||q_*||^2_{H^1} = \frac{3}{4} + \frac{2}{\pi} + \frac{\pi^2}{2} \end{aligned}$$

Example 3
$$\begin{aligned} u(q_*) &= \sin(\pi x) \\ q_* &= 1 + x\,, \quad ||q_*||^2_{H^1} = \frac{10}{3} \end{aligned}$$

The Gauss–Newton method was implemented using the Hebden–Reinsch method for the computation of $\mu_{\ell+1}^{MN}$ as the inner iteration. In all test runs we chose z^N to be the spline interpolant of z. The iterations were terminated if $t_\ell^{MN} \le TOL$ or $\ell > 15$. For all test runs we took $q_0 \equiv 0.2$ and incorporated either the Tikhonov regularization or the regularization by norm constraint. All computation were done on a SUN Sparcstation1 in double precision Fortran.

Tables 1 and 2 show the results for unperturbed observations. For small regularization parameter α the discretized problems have almost zero residual at the solution and the Gauss–Newton method convergences quadratically. Therefore there is no difference in the number of iterations for small α, except for Example 2, where regularization is needed to observe mesh independence.

Table 1

Number of Iterations

$q_0 \equiv 0.2 \quad z = u(q_*)$

	Example 1											
	$\mathrm{TOL} = 10^{-8}$						$\mathrm{TOL} = 10^{-6}$					
M	6	12	24	48	96	192	6	12	24	48	96	192
α N	12	24	48	96	192	384	12	24	48	96	192	384
0	7	7	7	7	7	7	7	7	7	7	7	7
10^{-8}	7	7	7	7	7	7	7	7	7	7	7	7
10^{-6}	7	7	7	7	7	7	7	7	7	7	7	7
10^{-4}	8	8	8	8	8	8	7	7	7	7	7	7
10^{-2}	10	10	10	10	10	10	8	8	8	8	8	8
	Example 2											
	$\mathrm{TOL} = 10^{-8}$						$\mathrm{TOL} = 10^{-6}$					
M	6	12	24	48	96	192	6	12	24	48	96	192
α N	12	24	48	96	192	384	12	24	48	96	192	384
0	11	> 15	8	6	7	7	10	> 15	8	7	7	7
10^{-8}	8	7	7	7	7	7	7	7	7	7	7	7
10^{-6}	7	7	7	7	7	7	6	6	6	6	6	6
10^{-4}	9	9	9	9	9	9	7	7	7	7	7	7
10^{-2}	9	9	9	9	9	9	7	7	7	7	7	7
	Example 3											
	$\mathrm{TOL} = 10^{-8}$						$\mathrm{TOL} = 10^{-6}$					
M	6	12	24	48	96	192	6	12	24	48	96	192
α N	12	24	48	96	192	384	12	24	48	96	192	384
0	8	8	8	8	8	8	7	7	7	7	7	7
10^{-8}	8	8	8	8	8	8	7	7	7	7	7	7
10^{-6}	8	8	8	8	8	8	7	7	7	7	7	7
10^{-4}	8	8	8	8	8	8	7	7	7	7	7	7
10^{-2}	10	10	10	10	10	10	8	8	8	8	8	8

In the norm constraint case, we obtain similar results, except for Example 2. Here we recognize an unstable behavior for $R = 1.5, 1.2$ and $TOL = 10^{-8}$. This might be due to the fact, that the Lagrange multipliers are computed approximately. If the constraint is active, we stop the inner iteration for the computation of μ_ℓ^{MN} if

$$\frac{|\,\|q_{\ell-1}^{MN}(\mu_\ell^{MN})\|_{H^1} - R|}{R} < 10^{-4}.$$

Therefore the projection is computed in the following way:

$$P(q_\ell^{MN} - F_N'(q_\ell^{MN})^* F_N(q_\ell^{MN})) = \begin{cases} \frac{\|q_\ell^{MN}\|_{H^1}}{\|q_\ell^{MN} - F_N'(q_\ell^{MN})^* F_N(q_\ell^{MN})\|_{H^1}} & \text{if } \frac{|\,\|q_\ell^{MN}\|_{H^1} - R|}{R} < 10^{-4} \\ \frac{R}{\|q_\ell^{MN} - F_N'(q_\ell^{MN})^* F_N(q_\ell^{MN})\|_{H^1}} & \text{else} \end{cases}.$$

Table 2

Number of Iterations

$$t_\ell^{MN} = \|q_\ell^{MN} - P(q_\ell^{MN} - F_N'(q_\ell^{MN})^* F_N(q_\ell^{MN}))\| \quad (= \|F_N'(q_\ell^{MN})^* F_N(q_\ell^{MN}) + \mu_\ell^{MN} q_\ell^{MN}\|)$$

$q_0 \equiv 0.2 \quad z = u(q_*)$

	Example 1											
	TOL = 10^{-8}						TOL = 10^{-6}					
M	6	12	24	48	96	192	6	12	24	48	96	192
R N	12	24	48	96	192	384	12	24	48	96	192	384
1.3	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)	7(7)	7(7)	7(7)	7(7)	7(7)	7(7)
1.0	10(10)	10(10)	10(10)	10(10)	10(10)	10(10)	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)
0.8	8(9)	8(9)	8(9)	8(9)	8(9)	8(9)	6(7)	6(7)	6(7)	6(7)	6(7)	6(7)
	Example 2											
	TOL = 10^{-8}						TOL = 10^{-6}					
M	6	12	24	48	96	192	6	12	24	48	96	192
R N	12	24	48	96	192	384	12	24	48	96	192	384
2.5	7(7)	7(7)	7(7)	7(7)	7(7)	7(7)	6(6)	7(7)	7(7)	7(7)	7(7)	7(7)
2.0	7(7)	7(7)	7(7)	7(7)	7(7)	7(7)	6(6)	6(6)	6(6)	6(6)	6(6)	6(6)
1.5	8(8)	8(8)	8(8)	8(8)	8(8)	7(7)	6(6)	6(6)	6(6)	6(6)	6(6)	6(6)
1.2	8(8)	8(8)	8(8)	8(8)	7(7)	7(7)	7(7)	6(6)	6(6)	6(6)	6(6)	6(6)
	Example 3											
	TOL = 10^{-8}						TOL = 10^{-6}					
M	6	12	24	48	96	192	6	12	24	48	96	192
R N	12	24	48	96	192	384	12	24	48	96	192	384
1.8	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)	7(7)	7(7)	7(7)	7(7)	7(7)	7(7)
1.3	10(10)	10(10)	10(10)	10(10)	10(10)	10(10)	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)
1.0	10(10)	10(10)	10(10)	10(10)	10(10)	10(10)	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)
0.8	9(9)	9(9)	9(9)	9(9)	9(9)	9(9)	7(7)	7(7)	7(7)	7(7)	7(7)	7(7)

Tables 3 and 4 show the results for perturbed observations. In the case of Tikhonov regularization mesh independence can be observed only for sufficiently large α. This behavior is theoretically justified through Theorems 2.1 and 3.4. The increase of α causes an increase of γ on one hand (for this problem we have $\gamma = \alpha$) and on the other hand an increase of the residual and therefore of σ. Our results indicate, that $\alpha = \gamma + \mu_* > \sigma$ for small, but sufficiently large α. If α is further increased, the difference of between α and σ gets smaller and for regularization parameters $\alpha \geq 1$ the method did not converge (a result which is not reported in our tables). For Examples 1 and 3, $\alpha = 0.1$, the criteria $\ell > 15$ is satisfied before the gradient reaches TOL, although the method converges.

Table 3

Number of Iterations $q_0 \equiv 0.2 \quad z = u(q_*) + 0.05\sin(10\pi x - 0.5\pi)$												
	Example 1											
	TOL $= 10^{-8}$						TOL $= 10^{-6}$					
M	6	12	24	48	96	192	6	12	24	48	96	192
α N	12	24	48	96	192	384	12	24	48	96	192	384
0	> 15	8	10	> 15	> 15	> 15	11	9	8	> 15	> 15	> 15
10^{-6}	10	10	11	11	11	11	7	7	8	8	8	8
10^{-4}	8	8	8	8	8	8	7	7	7	7	7	7
10^{-2}	10	10	10	10	10	10	8	8	8	8	8	8
10^{-1}	> 15	> 15	> 15	> 15	> 15	> 15	12	12	12	12	12	12
	Example 2											
	TOL $= 10^{-8}$						TOL $= 10^{-6}$					
M	6	12	24	48	96	192	6	12	24	48	96	192
α N	12	24	48	96	192	384	12	24	48	96	192	384
0	12	> 15	> 15	> 15	> 15	> 15	10	> 15	13	> 15	> 15	> 15
10^{-6}	8	10	> 15	> 15	10	10	7	7	> 15	> 15	7	7
10^{-4}	8	8	8	8	7	9	7	7	7	7	7	7
10^{-2}	9	9	9	9	9	9	7	7	7	7	7	7
10^{-1}	11	11	11	11	11	11	9	9	9	9	9	9
	Example 3											
	TOL $= 10^{-8}$						TOL $= 10^{-6}$					
M	6	12	24	48	96	192	6	12	24	48	96	192
α N	12	24	48	96	192	384	12	24	48	96	192	384
0	> 15	9	10	> 15	> 15	> 15	11	8	9	> 15	> 15	> 15
10^{-6}	11	11	11	13	13	14	7	8	8	9	9	9
10^{-4}	8	8	8	8	8	8	7	7	7	7	7	7
10^{-2}	10	10	10	10	10	10	8	8	8	8	8	8
10^{-1}	> 15	> 15	> 15	> 15	> 15	> 15	13	13	13	13	13	13

In the case of regularization by restriction, we choose a stronger perturbation, since the given constraints force a strong regularization. The numerical results for the weaker perturbation did not differ (much) from those given in the Table 2.

Table 4

Number of Iterations

$$t_\ell^{MN} = \|q_\ell^{MN} - P(q_\ell^{MN} - F_N'(q_\ell^{MN})^* F_N(q_\ell^{MN}))\| \quad (= \|F_N'(q_\ell^{MN})^* F_N(q_\ell^{MN}) + \mu_\ell^{MN} q_\ell^{MN}\|)$$

$$q_0 \equiv 0.2 \quad z = u(q_*) + 0.5\sin(10\pi x - 0.5\pi)$$

	Example 1											
	TOL = 10^{-8}						TOL = 10^{-6}					
M	6	12	24	48	96	192	6	12	24	48	96	192
R N	12	24	48	96	192	384	12	24	48	96	192	384
1.3	9(9)	8(8)	9(9)	9(9)	9(9)	9(9)	7(7)	7(7)	7(7)	7(7)	7(7)	7(7)
1.0	9(9)	9(10)	9(10)	9(10)	9(10)	9(10)	7(7)	7(7)	7(7)	7(7)	7(7)	7(7)
0.8	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)	6(6)	6(6)	6(6)	6(6)	6(6)	6(6)
	Example 2											
	TOL = 10^{-8}						TOL = 10^{-6}					
M	6	12	24	48	96	192	6	12	24	48	96	192
R N	12	24	48	96	192	384	12	24	48	96	192	384
2.5	11(11)	9(9)	10(10)	10(10)	11(11)	11(11)	8(8)	7(7)	8(8)	8(8)	8(8)	8(8)
2.0	11(11)	10(10)	10(10)	9(9)	9(9)	9(9)	8(8)	7(7)	7(7)	7(7)	7(7)	7(7)
1.5	11(11)	10(10)	10(10)	10(10)	9(9)	9(9)	8(8)	8(8)	7(7)	7(7)	7(7)	7(7)
1.2	11(11)	10(10)	10(10)	9(9)	9(9)	9(9)	8(8)	7(7)	7(7)	7(7)	7(7)	7(7)
	Example 3											
	TOL = 10^{-8}						TOL = 10^{-6}					
M	6	12	24	48	96	192	6	12	24	48	96	192
R N	12	24	48	96	192	384	12	24	48	96	192	384
1.8	9(9)	9(9)	9(9)	9(9)	9(9)	9(9)	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)
1.3	10(10)	10(10)	10(10)	10(10)	10(10)	10(10)	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)
1.0	10(10)	10(10)	10(10)	10(10)	10(10)	10(10)	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)
1.2	9(9)	9(9)	9(9)	9(9)	9(9)	9(9)	7(7)	7(7)	7(7)	7(7)	7(7)	7(7)

Acknowledgements

The author would like to thank Prof. E.W. Sachs, Universität Trier, for valuable suggestions and Prof. J.E. Dennis, Jr., Rice University, Houston, for his invitation to visit Rice University.

References

[1] E. L. Allgower and K. Böhmer. Application of the mesh independence principle to mesh refinement strategies. *SIAM J. Numer. Anal.*, 24:1335–1351, 1987.

[2] E. L. Allgower, K. Böhmer, F. A. Potra, and W. C. Rheinboldt. A mesh-independence principle for operator equations and their discretizations. *SIAM J. Numer. Anal.*, 23:160–169, 1986.

[3] W. Alt. Lipschitzian perturbations of infinite optimization problems. In A.V. Fiacco, editor, *Mathematical Programming with Data Perturbations II*, pages 7–21. Marcel Dekker, New-York, Basel, 1983.

[4] O. Axelsson and V.A. Barker. *Finite Element Solution of Boundary Value Problems.* Academic Press, Orlando, San Diego, New-York, London,..., 1984.

[5] H.T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems.* Systems & Control: Foundations & Applications. Birkhäuser-Verlag, Boston, Basel, Berlin, 1989.

[6] H.G. Bock. Randwertprobleme zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen. Preprint Nr. 442, Universität Heidelberg, Institut für Angewandte Mathematik, SFB 123, D–6900 Heidelberg, West Germany, 1988.

[7] K.M. Brown and J.E. Dennis, Jr. Derivative free analogues of the Levenberg–Marquardt and Gauss algorithms for nonlinear least squares approximations. *Numerische Mathematik*, 18:289–297, 1972.

[8] F. Colonius and K. Kunisch. Stability for parameter estimation in two point boundary value problems. *J. reine und angewandte Mathematik*, 370:1–29, 1986.

[9] F. Colonius and K. Kunisch. Output least squares stability in elliptic systems. *Appl. Math. Optim.*, 19:33–63, 1989.

[10] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Nonlinear Equations and Unconstrained Optimization.* Prentice-Hall, Englewood Cliffs, N.J, 1983.

[11] P. Deuflhard and G. Heindl. Affine invariant convergence theorems for Newton's method and extensions to related methods. *SIAM J. Numer. Anal.*, 16:1–10, 1979.

[12] H.W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularization of non-linear ill-posed problems. *Inverse Problems*, 5:523–540, 1989.

[13] M. Heinkenschloss, M. Laumen, and E. W. Sachs. Gauss–Newton methods with grid refinement. in preparation.

[14] C. T. Kelley and E. W. Sachs. Broyden's method for approximate solution of nonlinear integral equations. *J. Integral Eqs.*, 9:25–44, 1985.

[15] C. T. Kelley and E. W. Sachs. Approximate Quasi–Newton methods. *Math. Prog.*, 48:41–70, 1990.

[16] C. Kravaris and J.H. Seinfeld. Identification of parameters in distributed parameter systems by regularization. *SIAM J. Control and Optimization*, 23:217–241, 1985.

[17] M. Kroller and K. Kunisch. A numerical study of an augmented lagrangian method for the estimation of parameters in a two point boundary value problem. Technical Report 87, Technical University of Graz, Austria, 1987.

[18] H. Maurer and J. Zowe. First and second–order necessary and sufficient optimality conditions for infinite–dimensional programming problems. *Math. Prog.*, 16:98–110, 1979.

[19] J.J. Moré. Recent developments in algorithms and software for trust region methods. In B. Korte A. Bachem, M. Grötschel, editor, *Mathematical Programming, The State of The Art*, pages 258–287. Springer Verlag, Berlin, Heidelberg, New-York, 1983.

[20] C.R. Vogel. A constrained least squares regularization method for nonlinear ill-posed problems. *SIAM J. Control and Optimization*, 28:34–49, 1990.

[21] P-Å Wedin and P. Lindström. Methods and software for nonlinear least squares problems. Report UMINF–133.87, University of Umeå, Institute of Information Processing, S–90187 Umeå, Sweden, 1987. (revised 1988).