8 PDE Boundary Value Problems

This chapter deals with Newton methods for boundary value problems (BVPs) in nonlinear partial differential equations (PDEs). There are two principal approaches: (a) finite dimensional Newton methods applied to given systems of already discretized PDEs, also called *discrete Newton methods*, and (b) function space oriented inexact Newton methods directly applied to continuous PDEs, at best in the form of *inexact Newton multilevel methods.*

Before we discuss the two principal approaches in detail, we study the underlying feature of *asymptotic mesh independence* that connects the finite dimensional and the infinite dimensional Newton methods, see Section 8.1. In Section 8.2, we assume the standard situation in industrial technology software, where the grid generation module is strictly separated from the solution module. Consequently, nonlinear PDEs there arise as discrete systems of nonlinear equations with fixed finite, but usually high dimension n and large sparse ill-conditioned Jacobian (n, n)-matrix. This is the domain of applicability of finite dimensional inexact Newton methods. More advanced, but typically less convenient in a general industrial environment, are *function space* oriented inexact Newton methods, which additionally include the adaptive manipulation of discretization meshes within a multilevel or multigrid solution process. This situation is treated in Section 8.3 and compared there with *finite dimensional* inexact Newton techniques.

We will *not* treat 'multilevel Newton methods' here (often also called 'Newton multilevel methods'), which are in between discrete Newton methods and inexact Newton methods in function space; they have been extensively treated in the classical textbook [113] by W. Hackbusch or in the synoptic study [135] by R. Kornhuber, who uses an affine conjugate Lipschitz condition.

8.1 Asymptotic Mesh Independence

The term 'mesh independence' characterizes the observation that finite dimensional Newton methods, when applied to a nonlinear PDE on successively finer discretizations with comparable initial guesses, show roughly the same convergence behavior on all sufficiently fine discretizations. In this section, we want to analyze this experimental evidence from an abstract point of view.

P. Deuflhard, *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*, Springer Series in Computational Mathematics 35,
DOI 10.1007/978-3-642-23899-4_8, © Springer-Verlag Berlin Heidelberg 2011

Let a general nonlinear operator equation be denoted by

$$F(x) = 0\,, \tag{8.1}$$

where $F : D \to Y$ is defined on a convex domain $D \subset X$ of a Banach space X with values in a Banach space Y. We assume the existence of a unique solution x^* of this operator equation. The corresponding ordinary Newton method in Banach space may then be written as

$$F'(x^k)\Delta x^k = -F(x^k)\ , \quad x^{k+1} = x^k + \Delta x^k\ , \quad k = 0, 1, \ldots\,. \tag{8.2}$$

In each Newton step, the linearized operator equation must be solved, which is why this approach is often also called *quasilinearization*. We assume that an affine covariant version of the classical Newton-Mysovskikh theorem holds—like Theorem 2.2 for the finite dimensional case. Let ω denote the affine covariant Lipschitz constant characterizing the mapping F. Then the quadratic convergence of Newton's method is governed by the relation

$$|x^{k+1} - x^k| \le \tfrac{1}{2}\omega|x^k - x^{k-1}|^2\,,$$

where $|\cdot|$ is a norm in the domain space X.

In actual computation, we can only solve discretized nonlinear equations of finite dimension, at best on a sequence of successively finer mesh levels, say

$$F_j(x_j) = 0\ , \quad j = 0, 1, \ldots\ ,$$

where $F_j : D_j \to Y_j$ denotes a nonlinear mapping defined on a convex domain $D_j \subset X_j$ of a finite-dimensional subspace $X_j \subset X$ with values in a finite dimensional subspace Y_j. The corresponding finite dimensional ordinary Newton method reads

$$F_j'(x_j^k)\Delta x_j^k = -F_j(x_k^k)\ , \quad x_j^{k+1} = x_j^k + \Delta x_j^k\ , \quad k = 0, 1, \ldots\,.$$

In each Newton step, a system of linear equations must be solved, which may be a quite challenging task of its own in discretized PDEs. The above Newton system can be interpreted as a discretization of the linearized operator equation (8.2) and, at the same time, as a linearization of the discretization of the nonlinear operator equation (8.1). Again we assume that Theorem 2.2 holds, this time for the finite dimensional mapping F_j. Let ω_j denote the corresponding affine covariant Lipschitz constant. Then the quadratic convergence of this Newton method is governed by the relation

$$\|x_j^{k+1} - x_j^k\| \le \tfrac{1}{2}\omega_j\|x_j^k - x_j^{k-1}\|^2\,, \tag{8.3}$$

where $\|\cdot\|$ is a norm in the finite dimensional space X_j.

Under the assumptions of Theorem 2.2 there exist unique discrete solutions x_j^* on each level j. Of course, we want to choose appropriate discretization schemes such that

$$\lim_{j \to \infty} x_j^* = x^* \,.$$

From the synopsis of discrete and continuous Newton method, we immediately see that any comparison of the convergence behavior on different discretization levels j will direct us toward a comparison of the affine covariant Lipschitz constants ω_j. Of particular interest is the connection with the Lipschitz constant ω of the underlying operator equation.

Consistent norms. An important issue for any comparison of affine covariant Lipschitz constants ω_j on different discretization levels j is the choice of *consistent* norms. In the mathematical treatment of Galerkin methods, we will identify the norm $|\cdot|$ in X with the norm $\|\cdot\|$ in $X_j \subset X$. Moreover, the needs of algorithmic adaptivity strongly advise to choose *smooth* norms. These considerations bring us to Sobolev H^p-norms to be properly selected in each particular problem.

For non-Galerkin methods such as finite difference methods, the easiest way to construct consistent norms is to discretize the function space norm $|\cdot|$ appropriately, which directs us toward *discrete H^p-norms.* For example, in one-dimensional BVPs we may naturally use discrete L^2-norms (7.9) to treat highly nonuniform meshes—see also their application in (7.16). For uniform one-dimensional meshes, the discrete L^2-norms on level j differ from the Euclidean vector norms in $\mathbb{R}^{n_j}$ only by some dividing factor $\sqrt{n_j}$. Insertion of the discrete L^2-norm instead of the Euclidean vector norm into the Lipschitz condition (8.3) shows that this same factor would now multiply ω_j. As long as merely a single finite dimensional system were to be analyzed, this change would not make a substantial difference, but only affect the interpretation. A synoptic analysis of a *sequence* of nonlinear mappings, however, will be reasonable only, if consistent discrete norms are used.

In what follows we will consider the phenomenon of mesh independence of Newton's method along two lines. First, we will show that the discrete Newton sequence tracks the continuous Newton sequence closely, with a maximal distance bounded in terms of the mesh size; both of the Newton sequences behave nearly identically until, eventually, a small neighborhood of the solution is reached. Second, we prove the existence of affine covariant Lipschitz constants ω_j for the discretized problems, which approach the Lipschitz constant ω of the continuous problem in the limit $j \longrightarrow \infty$; again, the distance can be bounded in terms of the mesh size. Upon combining these two lines, we finally establish the existence of locally unique discrete solutions x_j^* in a vicinity of the continuous solution x^*.

To begin with, we prove the following nonlinear perturbation lemma.

Lemma 8.1 *Consider two Newton sequences $\{x^k\}$, $\{y^k\}$ starting at initial guesses x^0, y^0 and continuing as*

$$x^{k+1} = x^k + \Delta x^k \ , \quad y^{k+1} = y^k + \Delta y^k \ ,$$

where $\Delta x^k, \Delta y^k$ are the corresponding ordinary Newton corrections. Assume that an affine covariant Lipschitz condition with Lipschitz constant ω holds. Then the following propagation result holds:

$$\|x^{k+1} - y^{k+1}\| \le \omega \left(\frac{1}{2} \|x^k - y^k\| + \|\Delta x^k\| \right) \|x^k - y^k\| \, . \tag{8.4}$$

Proof. Dropping the iteration index k we start with

$$\begin{aligned}
&x + \Delta x - y - \Delta y \\
&= x - F'(x)^{-1}F(x) - y + F'(y)^{-1}F(y) \\
&= x - F'(x)^{-1}F(x) + F'(x)^{-1}F(y) - F'(x)^{-1}F(y) - y + F'(y)^{-1}F(y) \\
&= x - y - F'(x)^{-1}(F(x) - F(y)) + F'(x)^{-1}(F'(y) - F'(x))F'(y)^{-1}F(y) \\
&= F'(x)^{-1}\Big(F'(x)(x-y) - \int\limits_{t=0}^{1} F'(y + t(x-y))(x-y)\, dt \Big) \\
&\quad + F'(x)^{-1}(F'(y) - F'(x))\Delta y.
\end{aligned}$$

Upon using the affine covariant Lipschitz condition, we arrive at

$$\begin{aligned}
\|x^{k+1} - y^{k+1}\| &\le \int\limits_{t=0}^{1} \|F'(x^k)^{-1}\big(F'(x^k) - F'(y^k + t(x^k - y^k))\big)(x^k - y^k)\|\, dt \\
&\quad + \|F'(x^k)^{-1}(F'(y^k) - F'(x^k))\Delta y^k\| \\
&\le \frac{\omega}{2}\|x^k - y^k\|^2 + \omega\|x^k - y^k\| \, \|\Delta y^k\|,
\end{aligned}$$

which confirms (8.4). □

With the above auxiliary result, we are now ready to study the relative behavior of discrete versus continuous Newton sequences.

Theorem 8.2 *Notation as introduced. Let $x^0 \in \bigcap X_j \subset X$ denote a given starting value such that the assumptions of Theorem 2.2 hold including*

$$h_0 = \omega \|\Delta x^0\| < 2 \ .$$

For the discrete mappings F_j and all arguments $x_j \in S(x^0, \rho + \frac{2}{\omega}) \cap X_j$ define

$$F_j'(x_j)\Delta x_j = -F_j(x_j)\ , \quad F'(x_j)\Delta x = -F(x_j)\ .$$

Assume that the discretization is fine enough such that

$$\|\Delta x_j - \Delta x\| \le \delta_j \le \frac{1}{2\omega}\ . \tag{8.5}$$

Then the following cases occur:

I. *If* $h_0 \le 1 - \sqrt{1-2\omega\delta_j}$, *then*

$$\|x_j^k - x^k\| < 2\delta_j \le \frac{1}{\omega}\ , \quad k = 0, 1, \dots\ .$$

II. *If* $1 - \sqrt{1-2\omega\delta_j} < h_0 \le 1 + \sqrt{1-2\omega\delta_j}$, *then*

$$\|x_j^k - x^k\| \le \frac{1}{\omega}(1 + \sqrt{1-2\omega\delta_j}) < \frac{2}{\omega}\ , \quad k = 0, 1, \dots\ .$$

In both cases I and II, the asymptotic result

$$\limsup_{k\to\infty} \|x_j^k - x^k\| \le \frac{1}{\omega}(1 - \sqrt{1-2\omega\delta_j}) < 2\delta_j \le \frac{1}{\omega}$$

can be shown to hold.

Proof. In [114, pp. 99, 160], E. Hairer and G. Wanner introduced ‘Lady Windermere’s fan’ as a tool to prove discretization error results for evolution problems based on some linear perturbation lemma. We may copy this idea and exploit our nonlinear perturbation Lemma 8.1 in the present case. The situation is represented graphically in Figure 8.1.

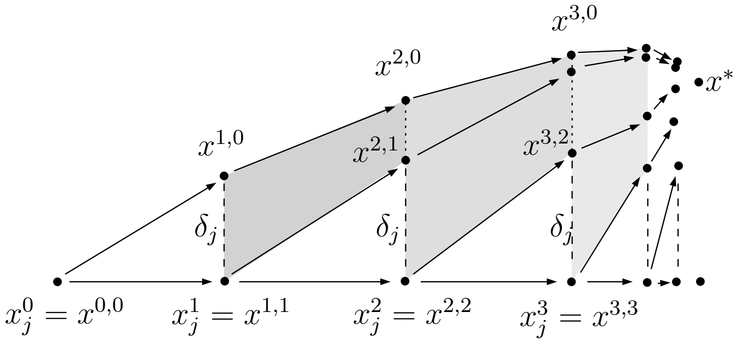

Fig. 8.1. Lady Windermere’s fan: continuous versus discrete Newton iterates.

The discrete Newton sequence starting at the given initial point $x_j^0 = x^{0,0}$ is written as $\{x^{k,k}\}$. The continuous Newton sequence, written as $\{x^{k,0}\}$, starts at the same initial point $x^0 = x^{0,0}$ and runs toward the solution point x^*. In between we define further continuous Newton sequences, written as $\{x^{i,k}\}, k = i, i+1, \ldots$, which start at the discrete Newton iterates $x_j^i = x^{i,i}$ and also run toward x^*. Note that the existence or even uniqueness of a discrete solution point x_j^* is not clear yet.

For the purpose of repeated induction, we assume that

$$\|x_j^{k-1} - x^0\| < \rho + \frac{2}{\omega},$$

which certainly holds for $k = 1$. In order to characterize the deviation between discrete and continuous Newton sequences, we introduce the two majorants

$$\omega\|\Delta x^k\| \le h_k \ , \quad \|x_j^k - x^{k,0}\| \le \epsilon_k \ .$$

Recall from Theorem 2.2 that

$$h_{k+1} = \tfrac{1}{2} h_k^2 \ . \tag{8.6}$$

For the derivation of a second majorant recursion, we apply the triangle inequality in the form

$$\|x^{k+1,k+1} - x^{k+1,0}\| \le \|x^{k+1,k+1} - x^{k+1,k}\| + \|x^{k+1,k} - x^{k+1,0}\|.$$

The first term can be treated using assumption (8.5) so that

$$\|x^{k+1,k+1} - x^{k+1,k}\| = \|x_j^k + \Delta x_j^k - \left(x^{k,k} + \Delta x^{k,k}\right)\| = \|\Delta x_j^k - \Delta x^{k,k}\| \le \delta_j \ .$$

For the second term, we may apply our nonlinear perturbation Lemma 8.1 (see the shaded regions in Fig. 8.1) to obtain

$$\|x^{k+1,k} - x^{k+1,0}\| \le \omega\Big(\tfrac{1}{2}\|x^{k,k} - x^{k,0}\| + \|\Delta x^{k,0}\|\Big)\|x^{k,k} - x^{k,0}\| \, .$$

Combining these results then leads to

$$\|x^{k+1,k+1} - x^{k+1,0}\| \le \delta_j + \frac{\omega}{2}\epsilon_k^2 + h_k\epsilon_k \ .$$

The above right side may be defined to be ϵ_{k+1}. Hence, together with (8.6), we arrive at the following set of majorant equations

$$h_{k+1} = \frac{1}{2}h_k^2 \ , \quad \epsilon_{k+1} = \delta_j + \frac{\omega}{2}\epsilon_k^2 + h_k\epsilon_k \ .$$

If we introduce the quantities $\alpha_k = \omega\epsilon_k + h_k$ and $\delta = \omega\delta_j$, we may obtain the decoupled recursion

$$\alpha_{k+1} = \delta + \tfrac{1}{2}\alpha_k^2 \ , \tag{8.7}$$

which can be started with $\alpha_0 = h_0$, since $\epsilon_0 = 0$. Upon solving the equation

$$\delta = \hat{\alpha} - \tfrac{1}{2}\hat{\alpha}^2 \ ,$$

we get the two equilibrium points

$$\hat{\alpha}_1 = 1 - \sqrt{1 - 2\delta} < 1 \ , \quad \hat{\alpha}_2 = 1 + \sqrt{1 - 2\delta} > 1 \ .$$

Insertion into the recursion (8.7) then leads to the form

$$\alpha_{k+1} - \hat{\alpha} = \tfrac{1}{2}(\alpha_k - \hat{\alpha})(\alpha_k + \hat{\alpha}) \ .$$

For $\alpha_k < \hat{\alpha}_2$ we see that

$$\tfrac{1}{2}(\alpha_k + \hat{\alpha}_1) < \tfrac{1}{2}(\hat{\alpha}_2 + \hat{\alpha}_1) = 1 \ ,$$

which implies that

$$|\alpha_{k+1} - \hat{\alpha}_1| < |\alpha_k - \hat{\alpha}_1| \ . \tag{8.8}$$

Hence, the fixed point $\hat{\alpha}_1$ is attractive, whereas $\hat{\alpha}_2$ is repelling. Moreover, since $\alpha_k + \hat{\alpha}_1 > 0$, we immediately obtain the result

$$\operatorname{sign}(\alpha_{k+1} - \hat{\alpha}) = \operatorname{sign}(\alpha_k - \hat{\alpha}) \ .$$

Therefore, we have the following cases:

$$\begin{array}{lccc} \text{I.} & \alpha_0 \le \hat{\alpha}_1 & \Longrightarrow & \alpha_k \le \hat{\alpha}_1 \ , \\ \text{II.} & \hat{\alpha}_1 < \alpha_0 < \hat{\alpha}_2 & \Longrightarrow & \hat{\alpha}_1 \le \alpha_k < \hat{\alpha}_2 \ . \end{array}$$

Insertion of the expressions for the used quantities then shows that cases I,II directly correspond to cases I,II of the theorem. Its last asymptotic result is an immediate consequence of (8.8). Finally, with application of the triangle inequality

$$\|x_j^{k+1} - x^0\| \le \epsilon_{k+1} + \|x^{k+1} - x^0\| < \frac{2}{\omega} + \rho$$

the induction and therefore the whole proof is completed. □

We are, of course, interested whether a discrete solution point x_j^* exists. The above tracking theorem, however, only supplies the following result.

Corollary 8.3 *Under the assumptions of Theorem* 8.2*, there exists at least one accumulation point*

$$\hat{x}_j \in S\left(x^*, 2\delta_j\right) \cap X_j \ \subset \ S\left(x^*, \frac{1}{\omega}\right) \cap X_j \ ,$$

which need not be a solution point of the discrete equations $F_j(x_j) = 0$.

Proof. This is just the last asymptotic result of Theorem 8.2. □

In order to prove more, Theorem 2.2 directs us to study the question of whether an affine covariant Lipschitz condition holds for the finite dimensional mapping F_j, too.

Lemma 8.4 *Let Theorem* 2.2 *hold for the mapping $F : X \to Y$. For collinear $x_j, y_j, y_j + v_j \in X_j$, define quantities $w_j \in X_j$ and $w \in X$ according to*

$$F_j'(x_j)w_j = \big(F_j'(y_j + v_j) - F_j'(y_j)\big)\, v_j\,,\ F'(x_j)w = \left(F'(y_j + v_j) - F'(y_j)\right) v_j\,.$$

Assume that the discretization method satisfies

$$\|w - w_j\| \le \sigma_j \|v_j\|^2\,. \tag{8.9}$$

Then there exist constants

$$\omega_j \le \omega + \sigma_j \tag{8.10}$$

such that the affine invariant Lipschitz condition

$$\|w_j\| \le \omega_j \|v_j\|^2$$

holds for the discrete Newton process .

Proof. The proof is a simple application of the triangle inequality

$$\|w_j\| \le \|w\| + \|w_j - w\| \le \omega\|v_j\|^2 + \sigma_j\|v_j\|^2 = (\omega + \sigma_j)\,\|v_j\|^2.$$

□

Finally, the existence of a unique solution x_j^* is now an immediate consequence.

Corollary 8.5 *Under the assumptions of Theorem* 8.2 *and Lemma* 8.4 *the discrete Newton sequence $\{x_j^k\}, k = 0, 1, \ldots$ converges quadratically to a unique discrete solution point*

$$x_j^* \in S\left(x^*, 2\delta_j\right) \cap X_j \ \subset\ S\left(x^*, \frac{1}{\omega}\right) \cap X_j\,.$$

Proof. We just need to apply Theorem 2.2 to the finite dimensional mapping F_j with the starting value $x_j^0 = x^0$ and the affine invariant Lipschitz constant ω_j from (8.10). □

Remark 8.1 In the earlier papers [3, 4] two assumptions of the kind

$$\|F_j'(x_j)^{-1}\| \le \beta_j\ , \quad \|F_j'(x_j + v_j) - F_j'(x_j)\| \le \gamma_j \|v_j\|$$

have been made in combination with the *uniformity* requirements

$$\beta_j \leq \beta\,, \quad \gamma_j \leq \gamma\,. \tag{8.11}$$

Obviously, these assumptions lack any affine invariance. More important, however, and as a consequence of the noninvariance, these conditions are phrased in terms of *operator norms*, which, in turn, depend on the relation of norms in the domain and the image space of the mappings F_j and F, respectively. For typical PDEs and norms we would obtain

$$\lim_{j\to\infty} \beta_j \to \infty\,,$$

which clearly contradicts the uniformity assumption (8.11). Consequently, an analysis in terms of β_j and γ_j would not be applicable to this important case. The situation is different with the affine covariant Lipschitz constants ω_j: they only depend on the choice of *norms in the domain space.* It is easy to verify that

$$\omega_j \leq \beta_j \gamma_j\,.$$

From the above Lemma 8.4 we see that the ω_j remain bounded in the limit $j \longrightarrow \infty$, as long as ω is bounded—even if β_j or γ_j blow up. Moreover, even when the product $\beta_j\gamma_j$ remains bounded, the Lipschitz constant ω_j may be considerably lower, i.e.

$$\omega_j \ll \beta_j \gamma_j\,.$$

For illustration, just compare the simple $\mathbb{R}^2$-example in Exercise 2.3.

Summarizing, we come to the following conclusion, at least in terms of upper bounds: If the asymptotic properties

$$\lim_{j\to\infty} \delta_j = 0\,, \quad \lim_{j\to\infty} \sigma_j = 0$$

can be shown to hold, then the convergence speed of the discrete ordinary Newton method is asymptotically just the one for the continuous ordinary Newton method—compare Exercises 8.3 and 8.4. Moreover, if related initial guesses x^0 and x^0_j and a common termination criterion are chosen, then even the number of iterations will be nearly the same.

Bibliographical Remark. The 'mesh independence' principle has been reported and even exploited for mesh design in papers by E.L. Allgower and K. Böhmer [3] and S.F. McCormick [148]. Further theoretical investigations of the phenomenon have been given in the paper [4] by E.L. Allgower, K. Böhmer, F.A. Potra, and W.C. Rheinboldt; that paper, however, lacked certain important features, which have been discussed in Remark 8.1 above. A first affine covariant theoretical study has later been worked out by P. Deuflhard and F.A. Potra in [82]; from that analysis, the modified term 'asymptotic mesh independence' naturally emerged. The presentation here follows the much simpler and more intuitive treatment [198] of the topic as given recently by M. Weiser, A. Schiela, and the author.

8.2 Global Discrete Newton Methods

In the present section we regard BVPs for nonlinear PDEs as given in already discretized form on a fixed mesh, to be briefly called *discrete PDEs* here. In what follows, we report about the comparative performance of exact and inexact Newton methods in solving such problems. Part of the results are from a recent paper by P. Deuflhard, U. Nowak, and M. Weiser [80].

In the exact Newton methods, we use either band mode LU-decomposition or a sparse solver provided by MATLAB. Failure exits in the various numerical tests are characterized by

- OUTMAX: the Newton iteration (outer iteration in inexact Newton methods) does not converge within 75 iterations,
- ITMAX: the inner iteration per inexact Newton step does not converge within ITMAX iterations,
- λ-fail: the adaptive damping strategy suggests some 'too small' damping factor $\lambda_k < 10^{-4}$.

8.2.1 General PDEs

This section documents the comparative performance of residual based (or affine contravariant) Newton methods versus error oriented (or affine covariant) Newton methods, both for the exact and the inexact versions, at a common set of discrete PDE test problems.

Common test set. We consider a subset of the discrete PDE problems given in [160]. In order to be able to compare exact and inexact methods, we selected examples in only two space dimensions. This choice leads to moderate system dimensions n that still permit a direct solution of the arising linear equations. Throughout the examples, we use the usual second order, centered finite differences on tensor product grids. Neumann boundary conditions are included by simple one-sided differences, as usual.

Example 8.1 *Artificial test problem (atp1).* This problem comprises the simple scalar PDE

$$\Delta u - (0.9\exp(-q) + 0.1u)(4x^2 + 4y^2 - 4) - g = 0\,,$$

where

$$g = \exp(u) - \exp(\exp(-q)) \;\text{ and } q = x^2 + y^2\,,$$

with boundary conditions $u|_{\partial\Omega} = 0$ on the domain $\Omega = [-3,3]^2$. The analytical solution is known to be $u(x,y) = \exp(-q)$.

Example 8.2 *Driven cavity problems (dcp1000, dcp5000).* This problem involves the steady stream-function/vorticity equations

$$\Delta\omega + \mathrm{Re}(\psi_x\omega_y - \psi_y\omega_x) = 0\ ,\quad \Delta\psi + \omega = 0\,,$$

where ψ is the stream-function and ω the vorticity. For the domain $\Omega = [0,1]^2$ the following discrete boundary conditions are imposed (with $\Delta x, \Delta y$ the mesh sizes in x, y-direction)

$$\begin{aligned}
&\tfrac{\partial\psi}{\partial y}(x,1) = -16x^2(1-x)^2\ ,\\
&\omega(x,0) = -\tfrac{2}{\Delta y^2}\psi(x,\Delta y)\ ,\\
&\omega(x,1) = -\tfrac{2}{\Delta y^2}[\psi(x,1-\Delta y) + \Delta y\tfrac{\partial\psi}{\partial y}(x,1)]\ ,\\
&\omega(0,y) = -\tfrac{2}{\Delta x^2}\psi(\Delta x,y)\ ,\\
&\omega(1,y) = -\tfrac{2}{\Delta x^2}\psi(1-\Delta x,y)\,.
\end{aligned}$$

Problems *dcp1000, dcp5000* correspond to Reynolds numbers $\mathrm{Re}=1000, 5000$, respectively. For both cases the default initial guess is $\psi^0 = \omega^0 = 0$.

As will be seen below, the residual based Newton strategy was unable to solve problems *dcp1000* and *dcp5000* with this initial guess. That is why we added problems *dcp1000a* and *dcp5000a* with the better initial guesses $\omega^0 = y^2\sin(\pi x)$, $\psi^0 = 0.1\sin(\pi x)\sin(\pi y)$.

Example 8.3 *Supersonic transport problem (sst2).* The four model equations for the chemical species O, O_3, NO, NO_2, represented by the unknown functions (u_1, u_2, u_3, u_4), are

$$\begin{aligned}
0 &= D\Delta u_1 + k_{1,1} - k_{1,2}u_1 + k_{1,3}u_2 + k_{1,4}u_4 - k_{1,5}u_1u_2 - k_{1,6}u_1u_4\ ,\\
0 &= D\Delta u_2 + k_{2,1}u_1 - k_{2,2}u_2 + k_{2,3}u_1u_2 - k_{2,4}u_2u_3\ ,\\
0 &= D\Delta u_3 - k_{3,1}u_3 + k_{3,2}u_4 + k_{3,3}u_1u_4 - k_{3,4}u_2u_3 + 800.0 + SST\ ,\\
0 &= D\Delta u_4 - k_{4,1}u_4 + k_{4,2}u_2u_3 - k_{4,3}u_1u_4 + 800.0\ ,
\end{aligned}$$

where $D = 0.5\cdot 10^{-9}$,
$k_{1,1},\ldots,k_{1,6} = 4\cdot 10^5, 272.443800016, 10^{-4}, 0.007, 3.67\cdot 10^{-16}, 4.13\cdot 10^{-12}$,
$k_{2,1},\ldots,k_{2,4} = 272.4438, 1.00016\cdot 10^{-4}, 3.67\cdot 10^{-16}, 3.57\cdot 10^{-15}$,
$k_{3,1},\ldots,k_{3,4} = 1.6\cdot 10^{-8}, 0.007, 4.1283\cdot 10^{-12}, 3.57\cdot 10^{-15}$,
$k_{4,1},\ldots,k_{4,3} = 7.000016\cdot 10^{-3}, 3.57\cdot 10^{-15}, 4.1283\cdot 10^{-12}$, and

$$SST = \begin{cases} 3250 & \text{if } (x,y)\in[0.5,0.6]^2 \\ 360 & \text{otherwise.} \end{cases}$$

Homogeneous Neumann boundary conditions are imposed on the unit square. For the initial guess we take

$$u_1^0(x,y) = 10^9 \,, u_2^0(x,y) = 10^9 \,, u_3^0(x,y) = 10^{13} \,, u_4^0(x,y) = 10^7 \,.$$

Again we consider better initial guesses to allow for convergence in the residual based Newton methods:

$$u_i^0 \rightarrow (1 + 100(\sin(\pi x)\sin(\pi y))^2)u_i^0 \,.$$

Name	Grid	Dim n	OrdNew
atp1	31×31	961	4
dcp1000	31×31	1922	OUTMAX
dcp1000a	31×31	1922	9
dcp5000	63×63	7983	OUTMAX
dcp5000a	63×63	7983	OUTMAX
sst2	51×51	10404	OUTMAX
sst2a	51×51	10404	OUTMAX

Table 8.1. Test set characteristics.

Characteristics of the selected test set are arranged in Table 8.1. In order to give some idea about the complexity of the individual problems, we first applied an exact ordinary Newton method (uncontrolled)—see the last column of the table. All of its failures are due to ‘too many’ Newton (outer) iterations (recall OUTMAX= 75).

Exact Newton methods. Recall that *exact* Newton methods require the direct solution of the arising linear subsystems for the Newton corrections. Hence, adaptivity only shows up through affine invariant trust region (or damping) strategies. From the code family NLEQ we compare the following variants:

- NLEQ-RES requiring monotonicity in the residual norm $\|F\|_2$, as discussed in Section 3.2.2,
- NLEQ-RES/L requiring monotonicity in the preconditioned residual norm $\|C_L F\|_2$, also discussed in Section 3.2.2; the preconditioner C_L comes from incomplete LU-decomposition with fill-in only accepted within the block pentadiagonal structure (compare, e.g., [184]), and
- NLEQ-ERR requiring monotonicity in the natural level function, as discussed in Section 3.3.3.

The residual based methods realize the restricted monotonicity test (3.32). For termination, the (possibly preconditioned) criterion (2.70) with FTOL $= 10^{-8}$ has been taken, except for the badly scaled problems *sst*, which required FTOL $= 10^{-5}$ to terminate within a tolerable computing time. The error oriented methods realize the restricted monotonicity test (3.47) and the (scaled) termination criterion (2.14) with XTOL $= 10^{-8}$.

Name	`NLEQ-RES`	`NLEQ-RES/L`	`NLEQ-ERR`
atp1	4 (0)	4 (0)	4 (0)
dcp1000	OUTMAX	10 (5)	8 (4)
dcp1000a	21 (17)	8 (2)	8 (2)
dcp5000	OUTMAX	OUTMAX	11 (7)
dcp5000a	42 (39)	λ-fail	8 (2)
sst2	λ-fail	12 (11)	13 (8)
sst2a	38 (33)	15 (13)	19 (14)

Table 8.2. Exact Newton codes: adaptive control via residual norm (`NLEQ-RES`), preconditioned residual norm (`NLEQ-RES/L`), and error norm (`NLEQ-ERR`).

In Table 8.2 we compare the residual based versus the error oriented exact Newton codes in terms of Newton steps (in parentheses: damped). As can be seen, there is striking evidence that the error oriented adaptive Newton methods are clearly preferable to the residual based ones, at least for the problem class tested here.

The main reason for this phenomenon is certainly that the arising discrete Jacobian matrices are bound to be ill-conditioned, the more significant the finer the mesh is. For this situation, the limitation of residual monotonicity has been described at the end of Section 3.3.1. Example 3.1 has given an illustration representative for a class of ODE boundary value problem. The experimental evidence here seems to indicate that the limitation carries over to PDE boundary value problems as well.

Inexact Newton methods. Finite dimensional *inexact* Newton methods contain some inner iterative solver, which induces the necessity of an accuracy matching between inner and outer iteration. The implemented ILU-preconditioning [184] is the same as in the exact Newton codes above. In the code family `GIANT`, various affine invariant damping and accuracy matching strategies are realized—according to the selected affine invariance class. The failure exit ITMAX was activated at 2000 inner iterations.

Residual based methods. For this type of inexact Newton method, we chose the codes `GIANT-GMRES/R` and `GIANT-GMRES/L` with right (R) or left (L) preconditioning.

As a first test, we selected the standard convergence mode from Sections 2.2.4 and 3.2.3, prescribing $\eta_k \leq \bar{\eta}$ with threshold values $\bar{\eta} = 0.1$ and $\bar{\eta} = 0.001$.

Name	NLEQ-RES	GIANT-GMRES/R/0.001		GIANT-GMRES/R/0.1	
atp1	4 (0)	4 (0)	34	4 (1)	28
dcp1000	OUTMAX	OUTMAX		OUTMAX	
dcp1000a	21 (17)	21 (17)	788	28 (23)	605
dcp5000	OUTMAX	OUTMAX		OUTMAX	
dcp5000a	42 (39)	43 (39)	5021	58 (53)	3208
sst2	λ-fail	15 (10)	1376	ITMAX	
sst2a	38 (33)	ITMAX		ITMAX	

Table 8.3. Residual based Newton codes: exact version NLEQ-RES versus inexact version GIANT-GMRES/R for threshold values $\bar{\eta} = 0.001$ and $\bar{\eta} = 0.1$.

Name	NLEQ-RES/L	GIANT-GMRES/L/0.001		GIANT-GMRES/L/0.1	
atp1	4 (0)	4 (0)	31	4 (1)	25
dcp1000	10 (5)	10 (5)	380	16 (10)	309
dcp1000a	8 (2)	8 (1)	279	12 (3)	229
dcp5000	OUTMAX	ITMAX		OUTMAX	
dcp5000a	λ-fail	24 (15)	1700	OUTMAX	
sst2	12 (10)	15 (12)	252	OUTMAX	
sst2a	15 (13)	18 (15)	465	OUTMAX	

Table 8.4. Preconditioned residual based Newton codes: exact version NLEQ-RES/L versus inexact version GIANT-GMRES/L for threshold values $\bar{\eta} = 0.001$ and $\bar{\eta} = 0.1$.

In Table 8.3, we compare exact versus inexact Newton methods, again at the common test set, in terms of Newton steps (in parentheses: damped Newton steps) and inner iterations. For comparison, the first column is identical to the first one from Table 8.2. In Table 8.4, the performance of two GIANT-GMRES/L versions is documented. This time, the first column is the second one from Table 8.2.

As can be seen from both tables, the inexact Newton codes behave very much like their exact counterparts in terms of outer iterations, with erratic discrepancies now and then. In view of the anyway poor behavior of the residual based Newton methods in this problem class, we did not realize the fully adaptive accuracy matching strategy (linear or quadratic convergence mode) in the frame of residual based inexact Newton methods.

Error oriented Newton methods. For this type of inexact Newton method, we chose the codes GIANT-CGNE/L and GIANT-GBIT/L, both with left (L) preconditioning. Adaptive matching strategies as worked out in Sections 2.1.5 and 3.3.4 have been realized. Initial values for the arising inner iterations were chosen according to the nested suggestions (3.59) and (3.60). Note that these inexact codes realize a damping strategy and a termination criterion slightly different from those in NLEQ-ERR. In view of a strict comparison, we constructed an exact variant `NLEQ-ERR/I`, which realizes just these modifications, i.e. which is an inexact Newton-ERR code with exact inner solution.

Name	`NLEQ-ERR/I`	`GIANT-CGNE/L`		`GIANT-GBIT/L`	
atp1	5 (0)	5 (0)	237	5 (0)	122
dcp1000	10 (5)		ITMAX	10 (5)	1388
dcp1000a	9 (3)		ITMAX	9 (3)	2241
dcp5000	13 (8)		ITMAX	14 (8)	5943
dcp5000a	10 (3)		ITMAX	10 (3)	9504
sst2	15 (11)	16 (11)	23084	16 (11)	1549
sst2a	20 (15)	20 (15)	39889	20 (15)	2399

Table 8.5. Error oriented Newton codes: exact version `NLEQ-ERR/I` versus inexact versions `GIANT-CGNE/L` and `GIANT-GBIT/L` for threshold values $\bar{\delta} = 10^{-3}$.

In Table 8.5, we give results for the 'standard convergence mode', imposing the (scaled) condition $\delta_k \leq \bar{\delta}$ for the threshold value $\bar{\delta} = 10^{-3}$ throughout, as defined in (2.61) for the local Newton method and (3.55) for the global Newton method, the latter via $\rho = 2\bar{\delta}/(1 - 2\bar{\delta})$. As can be seen, the first column for `NLEQ-ERR/I` and the third column of Table 8.2 for `NLEQ-ERR` differ only marginally.

From these numerical experiments, we may keep the following information:

- The error estimator (1.31) for `CGNE` is more reliable than (1.37) for `GBIT`.
- Nevertheless `CGNE` is less efficient than `GBIT`—compare Remark 1.2.
- The code `GIANT-GBIT/L` essentially reproduces the outer iteration pattern of the exact Newton code `NLEQ-ERR`*.

More insight into `GIANT-GBIT/L` can be gained from Table 8.6 where we compare the 'standard convergence mode' (SM) with the 'quadratic convergence mode' (QM), again in terms of outer (damped) and inner iterations. Different sets of control parameters are applied. The parameter ρ defines $\bar{\delta} = \frac{1}{2}\rho/(1+\rho)$ via (3.55). As a default, the parameter $\widetilde{\rho}$ is fixed to $\widetilde{\rho} = \frac{1}{2}\rho$, which, in turn, defines $\overline{\rho}_{\max} = \widetilde{\rho}/(1 + \widetilde{\rho})$ via (3.70).

The first column, SM(.025, .05), presents results obtained over our common test set, when the accuracy matching strategies (3.66) with (3.71) and (3.50) with (3.55) are implemented; the values $(\widetilde{\rho}, \rho) = (.025, .05)$ represent the largest values, for which all problems from the common test set were still solvable. This column should be compared with the third column in Table 8.5, where `GIANT-GBIT/L` has been realized roughly in an SM(.001, .002) variant: considerable savings are visible.

Detailed examination of the numerical results has revealed that the weakest point of this algorithm is the rather poor error estimator (1.37) realized

Name	$\mathrm{SM}(.025, .05)$		$\mathrm{SM}^*(1/16, 1/8)$		$\mathrm{QM}(.025, .05)$	
atp1	5 (0)	91	5 (0)	66	5 (0)	97
dcp1000	11 (5)	904	12 (5)	817	10 (5)	852
dcp1000a	11 (3)	1272	12 (4)	1458	9 (3)	1180
dcp5000	19 (11)	3952	16 (8)	3417	16 (11)	3802
dcp5000a	16 (3)	4304	11 (1)	3475	10 (3)	3539
sst2	22 (13)	963	19 (12)	1037	18 (13)	842
sst2a	25 (16)	1429	24 (17)	1597	22 (16)	1365

Table 8.6. Comparison of different variants of error oriented inexact Newton code `GIANT-GBIT/L`. Accuracy matching strategies $\mathrm{SM}(\widetilde{\rho}, \rho)$ and $\mathrm{QM}(\widetilde{\rho}, \rho)$ for control parameters $(\widetilde{\rho}, \rho)$. SM^* realizes an exact computation of the inner iteration error.

within `GBIT`, which is often quite off scale. For an illustration of this effect, the second column presents results for version SM^*, which realizes a *precise* error estimator

$$\epsilon_i = \|\delta x_i^k - \Delta x^k\|$$

instead of the unsatisfactory estimator (1.37); in this case, the relaxed choice $(\widetilde{\rho}, \rho) = (1/16, 1/8)$ appeared to be possible. Consequently, savings of inner iterations can be observed.

These savings, however, are not too dramatic when compared with the version QM shown in the third column; here the quadratic accuracy matching rule (3.56) is realized, which does not differ too much from the rule (2.62). For the control parameters we again selected $(\widetilde{\rho}, \rho) = (.025, .05)$ to allow for a comparison with SM in the first column. Obviously, this column shows the best comparative numbers.

Summary. From our restricted set of numerical experiments, we may nevertheless draw certain practical conclusions:

- Among the inner iterations for an inexact Newton-ERR method, the algorithm `GBIT` is the clear ‘winner’—despite the rather poor computational estimator for the inner iteration error, which is presently realized.
- Linear preconditioning also plays a role in nonlinear preconditioning as realized in the inexact Newton-ERR codes; in particular, the better the linear preconditioner, the better the inner iteration error estimator, the better the performance of the whole inexact Newton-ERR method.
- The ‘quadratic convergence mode’ in the local convergence phase can save a considerable amount of computing time over the ‘standard convergence mode’.

8.2.2 Elliptic PDEs

This section documents the comparative performance of exact versus inexact affine conjugate Newton methods at a common set of nonlinear BVPs for elliptic discrete PDEs. Recall that elliptic PDEs are associated with underlying convex optimization problems—see Sections 2.3 and 3.4.

Test set. Below three discretized nonlinear elliptic PDE BVPs in two space dimensions are given. Of course, their corresponding discretized functional is also at hand. All discretizations are simple finite difference schemes on uniform meshes.

Example 8.4 *Simple elastomechanics problem (elas).* For $\Omega =]0,1[^2$, minimize the functional (total energy in Ogden material law)

$$\int_{\Omega} \big(\|F\|^2 + (\mathrm{det}F)^{-1} - M(1/2,-1)u\big)\, dx \quad \text{with } F = I + \nabla u\,.$$

Homogeneous Dirichlet conditions on the boundary part $\{0\}\times[0,1]$ and natural boundary conditions on the remaining boundary part are imposed. Physically speaking, $u(x) \in \mathbb{R}^2$ is the displacement of an elastic body. The volume force $(1/2,-1)^T$ acting on the body is scaled by M, which can be used to weight the 'nonlinearity' of the problem. As initial value we chose $u^0 = 0$ in agreement with the Dirichlet conditions.

Detailed examination reveals that the above functional is not globally convex on the whole domain of definition, but only in a neighborhood of the solution. Fortunately, for the given initial guesses, our Newton codes did not encounter any nonpositive second derivatives. The locally unique solution is depicted in Figure 8.2, right.

Example 8.5 *Minimal surface problem over convex domain (msc).* Given $\Omega =]0,1[^2$, minimize the surface area

$$\int_{\Omega} (1 + |\nabla u|^2)^{\frac{1}{2}}\, dx$$

subject to the Dirichlet boundary conditions

$$u(x_1, x_2) = M(x_1 + (1 - 2x_1)x_2) \text{ on } \partial\Omega\,.$$

Here $u(x) \in \mathbb{R}$ is the vertical position of the surface parametrized over Ω. The scaling parameter M of the boundary conditions allows to weight the 'nonlinearity' of the problem. The initial value u^0 is chosen as the bilinear interpolation of the boundary conditions. This problem has a unique well-defined solution depicted in Figure 8.2, left.

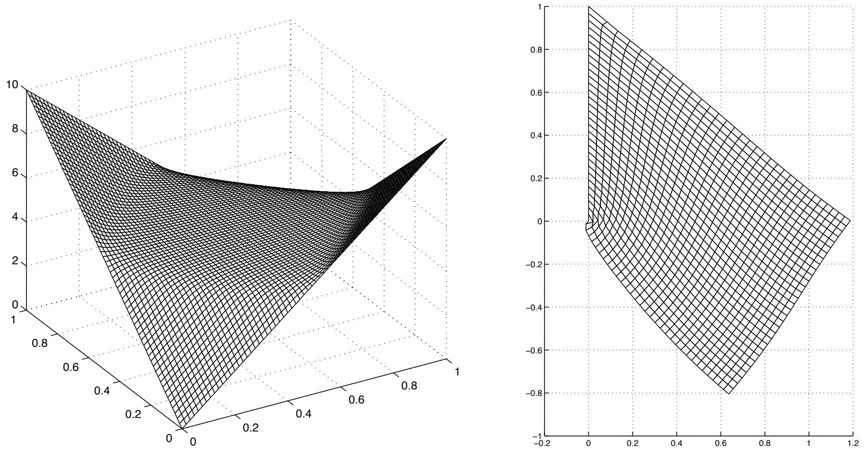

Fig. 8.2. *Left:* solution of problem *msc* ($M = 10, h = 1/63$). *Right:* solution of problem *elas* ($M = 2, h = 1/31$).

Example 8.6 *Minimal surface problem over nonconvex domain (msnc).* Given the domain $\Omega =]0,2[^2 \backslash]1,2[^2$, minimize the surface area

$$\int_\Omega (1 + |\nabla u|^2)^{\frac{1}{2}}\, dx$$

subject to the Dirichlet boundary conditions

$$u = 0 \text{ on } [0,2] \times \{0\} \cup \{0\} \times [0,2]\,, \quad u = M \text{ on } [1,2] \times \{1\} \cup \{1\} \times [1,2]\,.$$

On the remaining boundary parts, $[0,1] \times \{2\} \cup \{2\} \times [0,1]$, homogeneous Neumann boundary conditions $\partial_n u = 0$ are imposed. Here $u(x) \in \mathbb{R}$ is the vertical position of the surface parameterized over Ω. The scaling parameter M plays the same role as in problem *msc*. The initial value u^0 is chosen to be the linear interpolation of the Dirichlet boundary conditions on $[0,1] \times [1,2] \cup [1,2] \times [0,1]$ and the bilinear interpolation of the thus defined boundary values on $[0,1]^2$.

This problem has been deliberately constructed such that the underlying PDE does *not* have a unique *continuous* solution. Indeed, function space Newton multilevel methods (to be presented in Section 8.3 below) are able to detect this nonexistence: even though there exists a finite dimensional 'pseudosolution' on each mesh with size h, the local convergence domain of Newton's method shrinks when $h \to 0$. In the present setting of discrete PDEs, however, Newton's method will just supply a discrete solution on each of the meshes. As shown in Figure 8.3, these discrete solutions exhibit an interior 'discrete discontinuity', which is the sharper, the finer the mesh is.

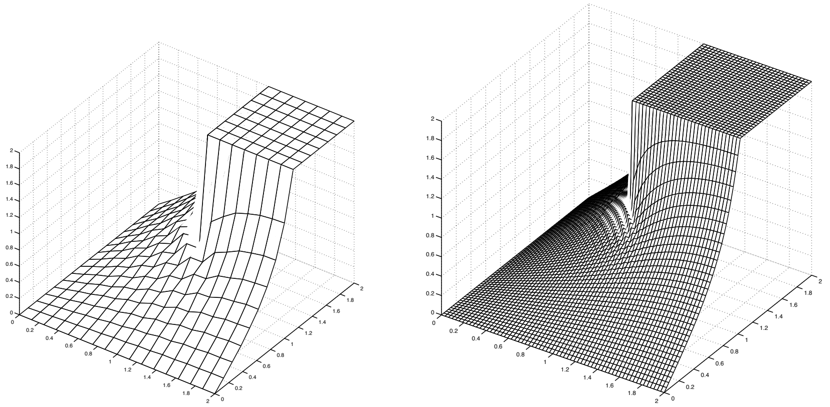

Fig. 8.3. Discrete solutions of problem *msnc*. *Left:* $M = 2, h = 1/8$, *right:* $M = 2, h = 1/32$.

Table 8.7 gives some 'measure' of the problem complexity for selected problem sizes of low dimension n: the value $M_{\max}$ in the last column indicates the maximal nonlinearity weight factor, for which the ordinary Newton method (uncontrolled) had still converged in our tests.

Name	Grid	Dim n	$M_{\max}$
msc	32×32	1024	6.2
elas	32×32	2048	1.0
msnc	32×32	3072	1.9

Table 8.7. Test set characteristics for special 2D grid.

Incidentally, below we also treat much larger problems, where dimensions up to $n \approx 200.000$ arise.

Exact versus inexact Newton methods. For the exact as well as the inexact Newton iteration, the energy error termination criterion (2.110) with ETOL $= 10^{-8}$ is taken. For Newton-PCG methods, we use the inner iteration termination criterion (1.25) and the accuracy matching strategy as worked out in Sections 2.3.3 and 3.4.3. As preconditioners we tested both the Jacobi and the incomplete Cholesky preconditioner (ICC) provided by MATLAB (with droptol $= 10^{-3}$). The failure exit ITMAX was activated at more than 500 inner iterations.

Local versus global Newton methods. In Table 8.8, we give comparative results for varying weight factor M at problem *msc.* Among the local Newton methods, we deliberately included the rather popular simplified Newton method (with initial Jacobian throughout the iteration)—see Section 2.3.2. The Newton-PCG algorithms are run in the *quadratic* convergence mode—see Section 2.3.3. The following features can be clearly observed:

- The local Newton methods converge only for the mildly nonlinear case (here: small M).
- Among the local Newton methods, the simplified variant behaves poorest.
- Exact and inexact Newton methods realize nearly the same number of (outer) Newton iterations, both local and global.

	local			global	
M	simplified	exact	inexact	exact	inexact
2	21	5	5	9	9
5	DIV	7	7	10	9
10	DIV	DIV	DIV	10	10

Table 8.8. Problem *msc*: comparative Newton steps (DIV: divergence).

Asymptotic mesh independence. In Table 8.9, we test different discrete Newton algorithms over the whole test set for decreasing mesh sizes. Asymptotic mesh independence as studied in Section 8.1 (see also Exercise 8.4) is clearly visible in problems *msc* and *elas*, but not in problem *msnc*, which does not have a unique continuous solution (see also Table 8.11 in Section 8.3.2 below). In the latter problem failures occur on the finest meshes—in agreement with the subsequent Example 8.9. The missing entries indicate the fact that the inexact codes were able to tackle much larger problems than the exact ones—both due to time and, even more pronounced, memory requirements of the direct solver on the finer meshes.

	msc($M = 10$)		*elas*($M = 2$)		*msnc*($M = 2$)	
N	exact	inexact	exact	inexact	exact	inexact
4	9	8	10	9	9	8
8	10	9	10	10	9	9
16	10	9	10	10	10	10
32	10	10	10	10	10	11
64	10	10		11		13
128		10				λ-fail
256		10				ITMAX

Table 8.9. Test set: Newton steps for decreasing mesh sizes $h = 1/N$.

Different preconditioners. In Table 8.10, the Jacobi preconditioner (Jac) and the incomplete Cholesky preconditioner (ICC) are compared for the quadratic and the linear convergence mode. For the linear convergence mode, $\bar{\Theta} = 0.5$ has been chosen. As can be seen, Jac is insufficient for fine meshes. ICC is more effective, at least for small up to moderate size meshes. The linear convergence mode is comparable to the quadratic convergence mode—as opposed to the behavior in the function space Newton method presented in Section 8.3 below.

		quadratic		linear	
	n	ICC	Jac	ICC	Jac
msc	4	7 (16)	7 (39)	7 (14)	8 (30)
(M=3.5)	8	6 (15)	6 (134)	6 (12)	15 (120)
	16	6 (19)	7 (385)	6 (12)	$\Theta \geq 1$
	32	6 (25)	8 (921)	7 (16)	$\Theta \geq 1$
	64	6 (35)	ITMAX	9 (24)	$\Theta \geq 1$
	128	6 (57)	ITMAX	12 (52)	$\Theta \geq 1$
	256	6 (103)	ITMAX	15 (96)	$\Theta \geq 1$
	512	6 (210)	ITMAX	19 (235)	$\Theta \geq 1$
elas	4	6 (18)	6 (174)	6 (12)	$\Theta \geq 1$
(M=0.2)	8	5 (19)	6 (479)	6 (12)	$\Theta \geq 1$
	16	5 (29)	ITMAX	7 (18)	$\Theta \geq 1$
	32	5 (44)	ITMAX	9 (36)	$\Theta \geq 1$
	64	5 (80)	ITMAX	11 (67)	$\Theta \geq 1$
	128	6 (176)	ITMAX	14 (144)	ITMAX

Table 8.10. Local inexact Newton-PCG method: comparative outer (inner) iterations. Quadratic versus linear convergence mode, Jacobi (Jac) versus incomplete Cholesky (ICC) preconditioning.

Summary. For elliptic discrete nonlinear PDEs both the exact and the inexact affine conjugate Newton methods perform efficiently and reliably, in close connection with the associated convergence theory. The inexact Newton code GIANT-PCG with ICC preconditioning seems to be a real competitor to so-called nonlinear PCG methods (for references see Section 2.3.3).

8.3 Inexact Newton Multilevel FEM for Elliptic PDEs

In this section we consider minimization problems of the kind

$$f(x) = \min ,$$

wherein $f : D \subset X \to \mathbb{R}$ is assumed to be a *strictly convex* C^2-functional defined on an open *convex* subset D of a Banach space X. Let X be endowed with a norm $\|\cdot\|$. In order to assure the *existence* of a minimum point $x^* \in D$, we assume that X is *reflexive*; in view of the subsequent finite element method (FEM), we choose $X = W^{1,p}$ for $1 < p < \infty$. Moreover, for given initial guess $x^0 \in D$, we assume that the level set $\mathcal{L}_0 := \{x \in D | f(x) \leq f(x^0)\}$ is nonempty, closed, and bounded. Under these assumptions the existence of a *unique* minimum point x^* is guaranteed. In this case the nonlinear minimization problem is equivalent to the nonlinear operator equation

$$F(x) := f'(x) = 0\,,\ x \in D\,. \tag{8.12}$$

In the present section, this equation is understood to be a nonlinear elliptic PDE problem. In order to guarantee the feasibility of Newton's method, we further assume that the PDE problem is *strictly* elliptic, which means that its symmetric Frechét-derivative $F'(x) = f''(x)$ is *strictly positive.*

In abstract notation, the ordinary Newton method for the mapping F reads $(k = 0, 1, \ldots)$

$$F'(x^k)\Delta x^k = -F(x^k)\,,\ x^{k+1} = x^k + \Delta x^k\,,$$

which just describes the successive linearization, often also called *quasilinearization*, of the above nonlinear operator equation. Since equation (8.12) is a nonlinear elliptic PDE in the Banach space $W^{1,p}$, the above Newton sequence consists of solutions of linear elliptic PDEs in some Hilbert space, say H_k, associated with each iterate $x^k \in W^{1,p}$. For reasonable arguments x, there exist energy products $\langle \cdot, F'(x)\cdot \rangle$, which induce energy norms $\langle \cdot, F'(x)\cdot \rangle^{1/2}$. The question of whether these energy norms are bounded for all arguments of interest needs to be discussed inside the proofs of the theorems to be stated below.

In the subsequent analysis, we will 'lift' these energy products and energy norms from H to $W^{1,p}$ (in the sense of dual pairing) defining the corresponding *local energy products* $\langle \cdot, F'(x)\cdot \rangle$ as symmetric bilinear forms and the induced *local energy norms* $\langle \cdot, F'(x)\cdot \rangle^{1/2}$ for arguments x in appropriate subsets of $W^{1,p}$. Moreover, motivated by the notation in Hilbert space, where the operator $F'(x)^{1/2}$ is readily defined, we also adopt the shorthand notation

$$\|F'(x)^{1/2} \cdot \| \equiv \langle \cdot, F'(x)\cdot \rangle^{1/2}$$

to be only used in connection with the local energy norms.

As already mentioned for space-like ODE BVPs in Section 7.4.2, *quasilinearization*, here for BVPs in more than one space dimension, cannot be realized without approximation errors. This means that we need to study *inexact* Newton methods

$$F'(x^k)\,\delta x^k = -F(x^k) + r^k\,,$$

equivalently written as

$$F'(x^k)\left(\delta x^k - \Delta x^k\right) = r^k \,.$$

Here the discretization errors show up either as residuals r^k or as the discrepancy between the inexact Newton corrections δx^k and the exact Newton corrections Δx^k. Among the discretization methods, we will focus on Galerkin methods, known to satisfy the *Galerkin condition*

$$\langle \delta x^k, F'(x^k)(\delta x^k - \Delta x_k)\rangle = \langle \delta x^k, r^k\rangle = 0\,. \tag{8.13}$$

Such a condition also holds in the finite dimensional inexact Newton-PCG method, where the residuals originate from the use of PCG as inner iterative solver—just look up condition (2.98) in Section 2.3.3. Recall that Newton-PCG methods are relevant for the discrete PDE situation, as presented in Section 8.2.2. Here, however, we want to treat the infinite dimensional exact Newton method approximated by an *adaptive* finite dimensional inexact Newton method. The benefit to be gained from adaptivity will become apparent in the following.

8.3.1 Local Newton-Galerkin methods

In this section we study the *ordinary* Newton-Galerkin method

$$x^{k+1} = x^k + \delta x^k\,,$$

where the iterates x^k are in $W^{1,p}$ and the inexact Newton corrections δx^k satisfy (8.13). With the above theoretical considerations we are ready to just modify the local convergence theorem for Newton-PCG methods (Theorem 2.20) in such a way that it covers the present infinite dimensional setting.

Theorem 8.6 *Let $f : D \to \mathbb{R}$ be a strictly convex C^2-functional to be minimized over some open convex domain $D \subset W^{1,p}$ endowed with the norm $\|\cdot\|$. Let $F'(x) = f''(x)$ be strictly positive. For collinear x, y, $z \in D$, assume the affine conjugate Lipschitz condition*

$$\left\|F'(z)^{-1/2}\big(F'(y) - F'(x)\big)v\right\| \le \omega \left\|F'(x)^{1/2}(y-x)\right\| \cdot \left\|F'(x)^{1/2}v\right\|$$

for some $0 \le \omega < \infty$. Consider an ordinary Newton-Galerkin method satisfying (8.13) *with approximation errors bounded by*

$$\delta_k := \frac{\|F'(x^k)^{1/2}(\delta x^k - \Delta x^k)\|}{\|F'(x^k)^{1/2}\delta x^k\|}\,.$$

At any well-defined iterate x^k, define the exact and inexact energy error norms by

$$\epsilon_k = \|F'(x^k)^{1/2}\Delta x^k\|^2 \,, \qquad \epsilon_k^\delta = \|F'(x^k)^{1/2}\delta x^k\|^2 = \frac{\epsilon_k}{1+\delta_k^2}$$

and the associated Kantorovich quantities as

$$h_k = \omega\,\|F'(x^k)^{1/2}\Delta x^k\| \,, \qquad h_k^\delta = \omega\,\|F'(x^k)^{1/2}\delta x^k\| = \frac{h_k}{\sqrt{1+\delta_k^2}} \,.$$

For given initial guess $x^0 \in D$ assume that the level set

$$\mathcal{L}_0 := \{x \in D \mid f(x) \le f(x^0)\}$$

is nonempty, closed, and bounded. Then the following results hold:

I. Linear convergence mode. *Assume that x^0 satisfies*

$$h_0 \le 2\overline{\Theta} < 2 \tag{8.14}$$

for some $\overline{\Theta} < 1$. Let $\delta_{k+1} \ge \delta_k$ throughout the inexact Newton iteration. Moreover, let the Galerkin approximation be controlled such that

$$\vartheta(h_k^\delta, \delta_k) = \frac{h_k^\delta + \delta_k\left(h_k^\delta + \sqrt{4+(h_k^\delta)^2}\right)}{2\sqrt{1+\delta_k^2}} \le \overline{\Theta} \,. \tag{8.15}$$

Then the iterates x^k remain in $\mathcal{L}_0$ and converge at least linearly to the minimum point $x^* \in \mathcal{L}_0$ such that

$$\|F'(x^{k+1})^{1/2}\Delta x^{k+1}\| \le \overline{\Theta}\,\|F'(x^k)^{1/2}\Delta x^k\| \tag{8.16}$$

and

$$\|F'(x^{k+1})^{1/2}\delta x^{k+1}\| \le \overline{\Theta}\,\|F'(x^k)^{1/2}\delta x^k\| \,. \tag{8.17}$$

II. Quadratic convergence mode. *Let, for some $\rho > 0$, the initial guess x^0 satisfy*

$$h_0 < \frac{2}{1+\rho} \tag{8.18}$$

and the Galerkin approximation be controlled such that

$$\delta_k \le \frac{\rho h_k^\delta}{h_k^\delta + \sqrt{4+(h_k^\delta)^2}} \,. \tag{8.19}$$

Then the inexact Newton iterates x^k remain in $\mathcal{L}_0$ and converge quadratically to the minimum point $x^* \in \mathcal{L}_0$ such that

$$\|F'(x^{k+1})^{1/2}\Delta x^{k+1}\| \le (1+\rho)\frac{\omega}{2}\|F'(x^k)^{1/2}\Delta x^k\|^2 \tag{8.20}$$

and

$$\|F'(x^{k+1})^{1/2}\delta x^{k+1}\| \le (1+\rho)\frac{\omega}{2}\|F'(x^k)^{1/2}\delta x^k\|^2 \,. \tag{8.21}$$

III. Functional descent. *The convergence in terms of the functional can be estimated by*

$$-\tfrac{1}{6}h_k^\delta \epsilon_k^\delta \leq f(x^k) - f(x^{k+1}) - \tfrac{1}{2}\epsilon_k^\delta \leq \tfrac{1}{6}h_k^\delta \epsilon_k^\delta \,.$$

The proof in the general Newton-Galerkin case is—mutatis mutandis—the same as the one for the more special Newton-`PCG` case in Theorem 2.20. In passing we mention that the above discussed boundedness of the local energy norms and, via the Cauchy-Schwarz inequality, also of the local energy products is actually guaranteed by (8.14), (8.16), and (8.17) in the linear convergence mode or by (8.18), (8.20), and (8.21) in the quadratic convergence mode.

For linear elliptic PDEs, we have *computational approximation error estimates* available, typically incorporated within adaptive multilevel FEM (Section 1.4.5), which are a special case of Galerkin methods. Hence, we may readily satisfy the above threshold criteria (8.15) or (8.19), respectively. Thus we are only left with the decision of whether to use the linear or the quadratic convergence mode in such a setting—an important algorithmic question that deserves special attention.

Computational complexity model. In order to get some insight, we study a rather simple, but nevertheless meaningful complexity model. It starts from the fact that at the final iterate, say x_q, we want to meet the prescribed energy error tolerance criterion (2.110), i.e.,

$$\epsilon_q \doteq \mathrm{ETOL}^2 \ .$$

If we replace the absolute error parameter $\mathrm{ETOL}^2 \ll \epsilon_0$ by a relative error parameter $\mathrm{EREL} \ll 1$ with $\mathrm{ETOL}^2 = \mathrm{EREL}^2 \cdot \epsilon_0$, then we may rewrite the above final accuracy requirement as

$$\Theta_0 \cdot \Theta_1 \cdots \Theta_{q-1} \doteq \mathrm{EREL}\,,$$

which is equivalent to

$$\sum_{k=0}^{q-1} \log \frac{1}{\Theta_k} \doteq \log \frac{1}{\mathrm{EREL}} \,. \tag{8.22}$$

The number q of Newton steps is unknown in advance. Let A_k denote the amount of work for step k. Then we will want to minimize the total amount of work, i.e.,

$$A_{total} = \sum_{k=0}^{q} A_k = \min$$

subject to the constraint (8.22). For the solution of this *discrete optimization* problem, there exists a quite efficient established heuristics, the popular *greedy algorithm*—see, e.g., Chapter 9.3 in the introductory textbook [2] by M. Aigner. From this, we obtain the prescription that, at Newton step k, the algorithm should maximize the *information gain per unit work*, i.e.,

$$I_k = \frac{1}{A_k} \log \frac{1}{\Theta_k} = \max \; . \tag{8.23}$$

In order to maximize this quantity with respect to the variable δ_k, the general relation (2.109) is applicable, which reads

$$\Theta_k \leq \vartheta(h_k^\delta, \delta_k) \, .$$

To simplify matters, we study the case $h_k \to 0$ here. Thus we arrive at the rough model

$$\Theta_k \doteq \vartheta(0, \delta_k) = \frac{\delta_k}{\sqrt{1+\delta_k^2}} \, ,$$

which, in view of (8.23), is equivalent to

$$\log \frac{1}{\Theta_k} \sim \log \left(1 + \frac{1}{\delta_k^2} \right) \, .$$

Next we compare two variants of Newton-Galerkin methods, the finite dimensional case (PCG) and the infinite dimensional case (FEM), which differ in the amount of work A_k as a function of δ_k.

Inexact Newton-`PCG` *method for discrete PDEs.* Assume that we attack a nonlinear discrete elliptic PDE by some inexact Newton method with PCG as inner iteration—as in the algorithm `GIANT-PCG`. This is exactly the situation treated in Section 8.2.2. For system dimension n, we have to consider

- the evaluation of the Jacobian matrix $J = F'(x^k)$, which is typically sparse, so that an amount $O(n)$ needs to be counted,
- the work per PCG step (evaluation of inner products), which for the sparse Jacobian J is also $O(n)$,
- the number m_k of PCG iterations at Newton step k: with preconditioner B we have (compare, e.g., Corollary 8.18 in the textbook [77])

$$m_k \sim \sqrt{\kappa(BJ)} \log 2 \left(1 + 1/\delta_k^2\right) \, .$$

Summing up, we arrive at the rough estimate

$$A_k \sim \left(c_1 + c_2 \log\left(1 + 1/\delta_k^2\right)\right) n \sim \text{const} + \log\left(1 + 1/\delta_k^2\right) \, ,$$

where 'const' represents some positive constant. So we finally end up with

$$I_k \sim \frac{\log(1 + 1/\delta_k^2)}{\text{const} + \log(1 + 1/\delta_k^2)} = \max \, .$$

The right hand side is a monotone *decreasing* function of δ_k, which directs us towards the smallest possible value of δ_k, i.e., to the *quadratic convergence mode.* It may be worth noting that the above analysis would lead to the same decision, if PCG were replaced by some linear multigrid method.

Inexact Newton multilevel FEM for continuous PDEs. For the inner iteration we now take an adaptive multilevel method for linear elliptic PDEs (such as the multiplicative multigrid algorithm `UG` by G. Wittum, P. Bastian et al. [22] or the additive multigrid algorithm `KASKADE` by P. Deuflhard, H. Yserentant et al. [78, 36, 23]). An example of such an algorithm is implemented in our code `Newton-KASKADE`. As a consequence of the adaptivity, the dimension n of subproblems to be solved at step k depends on δ_k. Let d denote the underlying spatial dimension. At iteration step k on refinement level j of the multilevel discretization, let n_k^j be the number of nodes and ϵ_k^j the local energy. With $l = l_k$ we mean the final discretization level, at which the prescribed final accuracy δ_k is achieved. On energy equilibrated meshes for linear elliptic PDEs, we have the following asymptotic theoretical result (see I. Babuška et al. [13])

$$\left(\frac{n_k^0}{n_k^l}\right)^{2/d} \sim \frac{\epsilon_k^\infty - \epsilon_k^l}{\epsilon_k^\infty} \leq \frac{\delta_k^2}{1+\delta_k^2} \,.$$

Any decent multigrid solver for linear elliptic PDEs will require an amount of work proportional to the number of nodes, i.e.

$$A_k \sim n_k^l \sim n_k^0 \left(1 + 1/\delta_k^2\right)^{d/2} \,.$$

Inserting this result into I_k, we arrive at the rough estimate

$$I_k \sim \left(1 + 1/\delta_k^2\right)^{-d/2} \log\left(1 + 1/\delta_k^2\right) = \max \,.$$

For variable space dimension d this scalar function has its maximum at

$$\delta_k = 1/\sqrt{\exp(2/d) - 1} \,,$$

which, with the help of (2.113), then leads to the choice

$$\overline{\Theta} = \exp(-1/d) \,.$$

We thus have the approximate values

$$d = 2: \quad \delta_k = 0.76,\ \overline{\Theta} = 0.61 \,, \qquad d = 3: \quad \delta_k = 1.03,\ \overline{\Theta} = 0.72 \,.$$

Even though our rough complexity model might not cover such large values of δ_k, these results may nevertheless be taken as a clear indication to favor the *linear* over the quadratic convergence mode in an *adaptive* multilevel setting. Empirical tests actually suggested to use $\delta_k \approx 1$ corresponding to $\overline{\Theta} \approx 0.7$ as default values.

Example 8.7 By modification of a problem given by R. Rannacher [175], we consider the convex functional in two space dimensions (with x, y Euclidean coordinates here)

$$f(u) = \int_{\Omega} \left(1 + |\nabla u|^2\right)^{p/2} - gu\, dx\,,\ p > 1\,,\ x \in \Omega \subset \mathbb{R}^2\,,\ u \in W^{1,p}(\Omega)$$

for the specification $p = 1.4,\ g \equiv 0$. The functional gives rise to the first and second order expressions

$$\begin{aligned}\langle F(u), v\rangle &= \int_{\Omega} \bigl(p(1 + |\nabla u|^2)^{p/2-1}\langle \nabla u, \nabla v\rangle - gv\bigr)\, dx\ ,\\ \langle w, F'(u)v\rangle &= \int_{\Omega} p\bigl(\ (p-2)(1 + |\nabla u|^2)^{p/2-2}\langle \nabla w, \nabla u\rangle\langle \nabla u, \nabla v\rangle\\ &\qquad +(1 + |\nabla u|^2)^{p/2-1}\langle \nabla w, \nabla v\rangle\bigr)\, dx\ .\end{aligned}$$

With $\langle\cdot,\cdot\rangle$ the Euclidean inner product in $\mathbb{R}^2$, the term $\langle v, F'(u)v\rangle$ is strictly positive for $p \geq 1$.

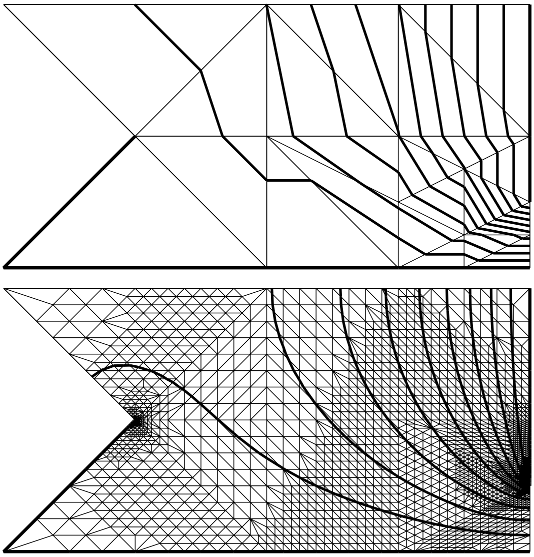

Fig. 8.4. Example 8.7. Newton-KASKADE iterates: *Top:* initial guess u^0 on initial coarse grid, *bottom:* iterate u^3 on automatically refined grid. *Thick lines:* homogeneous Dirichlet boundary conditions and level lines, *thin lines:* Neumann boundary conditions.

In order to solve this problem, we used the linear convergence mode in an adaptive Newton multilevel FEM with KASKADE to solve the arising linear elliptic PDEs. Figure 8.4 compares the starting guess u^0 on its coarse mesh ($j = 1$) with the inexact Newton iterate u^3 on its fine mesh ($j = 14$). The coarse mesh consists of $n_1 = 17$ nodes, the fine mesh of $n_{14} = 2054$ nodes; for comparison, a uniformly refined mesh at level $j = 14$ would have about $\overline{n}_{14} \approx 136000$ nodes. Note that—in the setting of multigrid methods, which require $O(n)$ operations—the total amount of work would be essentially blown up by the factor $\overline{n}_{14}/n_{14} \approx 65$. Apart from this clear computational saving, adaptivity also nicely models the two critical points on the boundary, the re-entrant corner and the discontinuity point.

8.3.2 Global Newton-Galerkin methods

In this section we study the inexact *global* Newton-Galerkin method

$$x^{k+1} = x^k + \lambda_k \delta x^k$$

in terms of iterates $x^k \in W^{1,p}$, inexact Newton corrections δx^k satisfying (8.13), and damping factors λ_k to be chosen appropriately. As the most prominent representatives of such methods we will take adaptive Newton multilevel FEMs, whenever it comes to numerical examples.

In Section 3.4.3 we had already discussed the finite dimensional analogue, the global inexact Newton-PCG method. With the theoretical considerations at the beginning of Section 8.3, we are prepared to modify the global convergence theorems for the Newton-PCG methods in such a way that they apply to the more general Newton-Galerkin case. In what follows, we just combine and modify our previous Theorems 3.23 and 3.26.

Theorem 8.7 *Notation as introduced above. Let $f : D \to \mathbb{R}^1$ be a strictly convex C^2-functional to be minimized over some open convex domain $D \subset W^{1,p}$ and $F'(x) = f''(x)$ be strictly positive. For $x, y \in D$, assume the affine conjugate Lipschitz condition*

$$\|F'(x)^{-1/2}(F'(y) - F'(x))(y - x)\| \leq \omega \|F'(x)^{1/2}(y - x)\|^2$$

with $0 \leq \omega < \infty$. Let Δx^k denote the exact and δx^k the inexact Newton correction. For each well-defined iterate $x^k \in D$, define the quantities

$$\epsilon_k = \|F'(x^k)^{1/2}\Delta x^k\|^2 \,, \qquad \epsilon_k^\delta = \|F'(x^k)^{1/2}\delta x^k\|^2 = \frac{\epsilon_k}{1 + \delta_k^2} \,,$$

$$h_k = \omega \|F'(x^k)^{1/2}\Delta x^k\| \,, \qquad h_k^\delta = \omega \|F'(x^k)^{1/2}\delta x^k\| = \frac{h_k}{\sqrt{1 + \delta_k^2}} \,.$$

Moreover, let $x^k + \lambda \delta x^k \in D$ for $0 \leq \lambda \leq \lambda_{\max}^k$ with

$$\lambda_{\max}^k := \frac{4}{1+\sqrt{1+8h_k^\delta/3}} \le 2\,.$$

Then

$$f(x^k + \lambda \Delta x^k) \le f(x^k) - t_k(\lambda)\epsilon_k^\delta$$

where

$$t_k(\lambda) = \lambda - \tfrac{1}{2}\lambda^2 - \tfrac{1}{6}\lambda^3 h_k^\delta\,.$$

The optimal choice of damping factor is

$$\overline{\lambda}_k = \frac{2}{1+\sqrt{1+2h_k^\delta}} \le 1\,.$$

As in the local convergence case, h_k is the Kantorovich quantity and δ_k the relative Galerkin approximation error.

Adaptive damping and accuracy matching. Following our usual paradigm, the unknown theoretical quantities h_k and δ_k are replaced by computationally available estimates $[h_k]$ and $[\delta_k]$. For $[h_k]$ we just use the terms $E_{1,2,3}(\lambda)$ as given in Section 3.4.2 for the exact Newton method and in Section 3.4.3 for the inexact Newton-PCG method. On this basis we realize the correction strategy (3.88) with h_k replaced by the a-posteriori estimate $[h_k^\delta]$ and the prediction strategy (3.89) with h_{k+1} replaced by the a-priori estimate $[h_{k+1}^\delta]$. Unless stated otherwise, we choose the approximation error bound $\delta_k = 1$ as a default throughout the Newton-Galerkin iteration, thus eventually merging into the linear local convergence mode.

Example 8.8 Good versus bad initial coarse grid. We return to our previous Example 8.7, but this time for the critical value $p = 1$, which characterizes the (parametric) *minimal surface* problem. This value is critical, since then $u \in W^{1,1}$, a *nonreflexive* Banach space, which implies that the existence of a unique solution is no longer guaranteed. For special boundary conditions and inhomogeneities g, however, a unique solution can be shown to exist, even in $C^{0,1}$ (see, e.g., the textbooks by E. Zeidler [205]). Such a situation occurs, e.g., for

$$\Omega = \left[-\frac{\pi}{2}, 0\right] \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\,, \quad u|_{\partial\Omega} = s\cos x \cos y\,, \quad g \equiv 0\ .$$

Taking the Z_2-symmetry along the x-axis into account, we may halve Ω and impose homogeneous Neumann boundary conditions at $y = 0$. The parameter s is set to $s = 3.5$. From a quick rough examination of the problem, we expect a boundary layer at $x = 0$. As initial guess u^0 we take the prescribed values on the Dirichlet boundary part and otherwise just zero.

Again we solve the problem by Newton-KASKADE. As good initial coarse grid we select the grid in Figure 8.5, left, which takes the expected boundary layer into account. As bad initial coarse grid we choose the one in Figure 8.5, right,

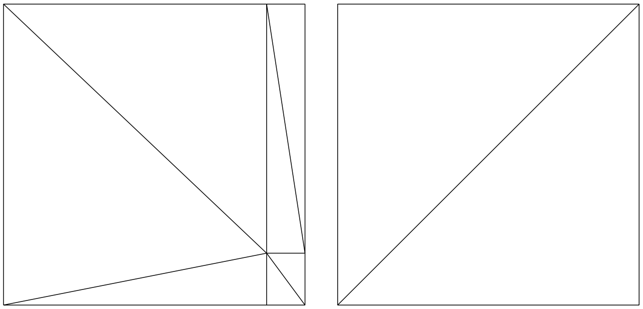

Fig. 8.5. Example 8.8. Good (left) and bad (right) initial coarse grid.

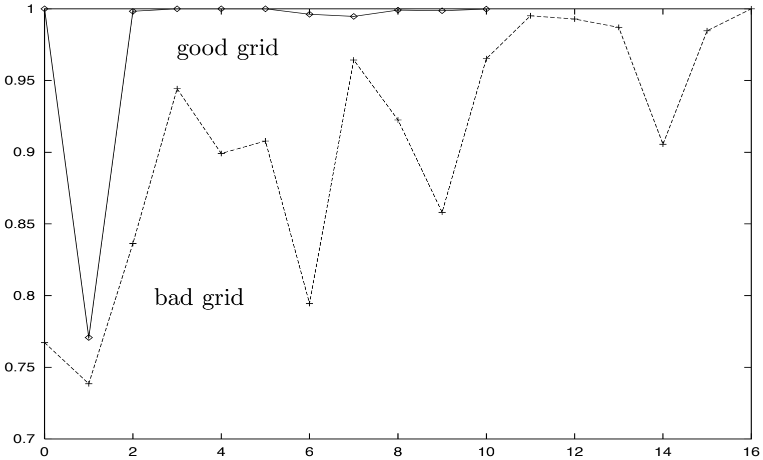

Fig. 8.6. Example 8.8. Comparative damping strategies for good and bad initial coarse grids.

which deliberately ignores any knowledge about the occurrence of a boundary layer.

In Figure 8.6, the comparative performance of our global `Newton-KASKADE` algorithm is documented in terms of the obtained damping factors for both initial grids. As expected from reports in the engineering literature, the bad coarse grid requires many more iterations to eventually capture the nonlinearity.

Example 8.9 Function space versus finite dimensional approach. Once again, we return to Example 8.7, this time for the critical value $p = 1$. In Figure 8.7, we show two settings: On the left (Example 8.9a), a unique solution exists, which has been computed, but is not documented here; this example serves for comparison only, see Table 8.11 below. Our main interest focuses on Example 8.9b, where *no (physical) solution exists.* For the initial guess u^0 we take the prescribed values on the Dirichlet boundary and zero otherwise.

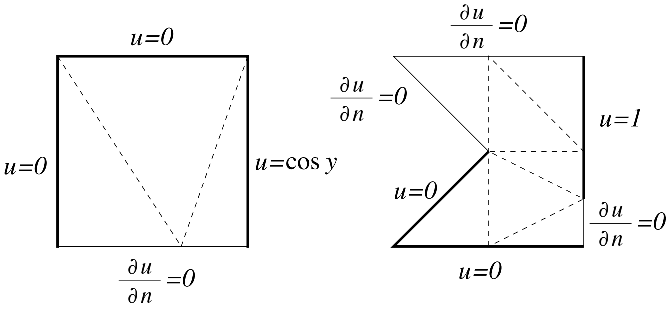
u=0
u=0
u=cos y
$\frac{\partial u}{\partial n}$ =0
$\frac{\partial u}{\partial n}$ =0
$\frac{\partial u}{\partial n}$ =0
u=1
u=0
$\frac{\partial u}{\partial n}$ =0
u=0

Fig. 8.7. Example 8.9. Domains and initial coarse grids. *Black lines:* Dirichlet boundary conditions, *grey lines:* Neumann boundary conditions. *Left:* Example 8.9a, unique solution exists. *Right:* Example 8.9b, no solution exists.

At Example 8.9b, we want to compare the algorithmic behavior of two different Newton-FEM approaches:

- our *function space* oriented approach, as presented in this section, and
- the *finite dimensional* approach, which is typically implemented in classical Newton-multigrid FEMs.

In the finite dimensional approach, the discrete FE problem is solved successively *on each of the mesh levels* so that there the damping factors will repeatedly run up to values $\lambda = 1$. In contrast to that behavior, our function space approach aims at directly solving the continuous problem by exploiting information available from the whole mesh refinement history. Consequently, if a unique solution exists, this approach will reach the local convergence phase in accordance with the mesh refinement process. Such a behavior has already been shown for our preceding Example 8.8 in Figure 8.6.

Figure 8.8 gives an account for Example 8.9b. In the finite dimensional option, damping factors $\lambda = 1$ arise repeatedly on each of the mesh refinement levels. After more than 60 Newton-FEM iterations, this approach gives the impression of a unique solvability of the problem—on the basis of the local convergence of the Newton-FEM algorithm on each of the successive meshes. Our function space option, however, terminates already after 20 Newton iterations for $\lambda < \lambda_{\min} = 0.01$.

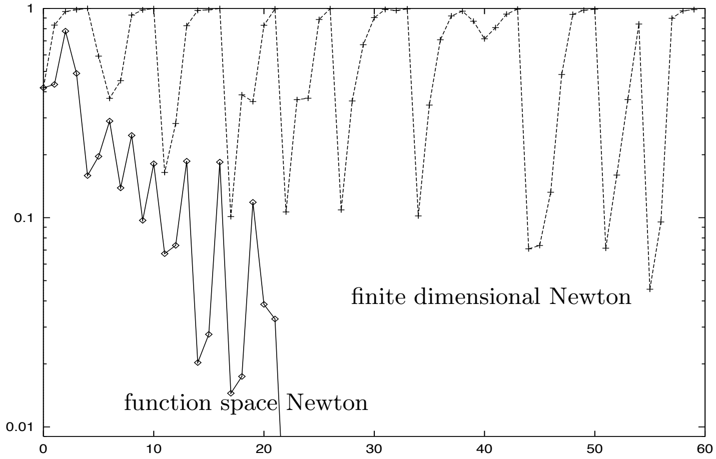

Fig. 8.8. Example 8.9b. Iterative damping factors for two Newton-FEM algorithms. To be compared with Figure 8.9.

To understand this discrepancy, we simultaneously look at the local energy norms ϵ_k, which measure the *exact* Newton corrections Δx^k, see Figure 8.9. The finite dimensional method ends up with 'sufficiently small' Newton corrections on each of the refinement levels, pretending some local *pseudoconvergence*. Our function space Newton method, however, stays with 'moderate size' corrections throughout the iteration.

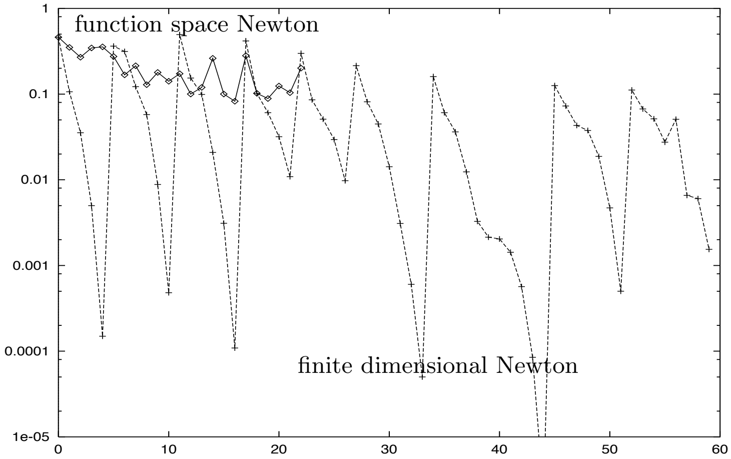

Fig. 8.9. Example 8.9b. Iterative energy norms $\epsilon_k^{1/2}$ for two Newton-FEM algorithms. To be compared with Figure 8.8.

Asymptotic mesh dependence. Table 8.11 (from [198]) compares the actually computed affine conjugate Lipschitz estimates $[\omega_j]$ as obtained from `Newton-KASKADE` in Example 8.9a and in Example 8.9b. Obviously, Example 8.9a, which has a unique solution, exhibits asymptotic mesh independence as studied in Section 8.1. Things are totally different in Example 8.9b, where the Lipschitz estimates clearly increase. Note that the blow-up of the lower bounds $[\omega_j]$ in Table 8.11 implies a blow-up of the Lipschitz constants ω_j—for this purpose the bounds are on the rigorous side.

	Example 8.9a		Example 8.9b	
j	♯ unknowns	$[\omega_j]$	♯ unknowns	$[\omega_j]$
0	4	1.32	5	7.5
1	7	1.17	10	4.2
2	18	4.55	17	7.3
3	50	6.11	26	9.6
4	123	5.25	51	22.5
5	158	20.19	87	50.3
6	278	19.97	105	1486.2
7	356	9.69	139	2715.6
8	487	8.47	196	5178.6
9	632	11.73	241	6837.2
10	787	44.21	421	12040.2
11	981	49.24	523	167636.0
12	1239	20.10	635	1405910.0
13	1610	32.93		
14	2054	37.22		

Table 8.11. Computational Lipschitz estimates $[\omega_j]$ on levels j. *Example 8.9a:* unique solution exists, *Example 8.9b:* no solution exists.

Interpretation. Putting all pieces of available information together, we now understand that on each of the levels j this problem has a finite dimensional solution x_j^*, unique within the finite dimensional Kantorovich ball with radius $\rho_j \sim 1/\omega_j$; however, these balls shrink from radius $\rho_1 \sim 1$ down to $\rho_{22} \sim 10^{-6}$. Frank extrapolation of this effect suggests that

$$\lim_{j\to\infty} \rho_j = 0 \, .$$

Obviously, *the algorithm insinuates that a unique continuous solution of the stated PDE problem does not exist.* This feature would certainly be desirable for any numerical PDE solver.

Exercises

Exercise 8.1 In Section 8.3.1, a rough computational complexity model of adaptive multilevel FEM for nonlinear elliptic PDEs leads to the problem (dropping the index k)

$$I \sim \left(1 + 1/\delta^2\right)^{-d/2} \log\left(1 + 1/\delta^2\right) = \max\,,$$

where d is the spatial dimension.

a) Calculate the maximum point δ and evaluate it for $d = 2$ and $d = 3$.

b) How would the rough model need to be changed, if the situation $h \neq 0$ were to be modeled?

Exercise 8.2 Consider the finite dimensional Newton sequence

$$x^{k+1} = x^k + \Delta x^k\,,$$

where x^0 is given and Δx^k is the solution of a linear system. In sufficiently large scale computations, rounding errors caused by direct elimination or truncation errors from iterative linear solvers will generate a different sequence

$$y^{k+1} = y^k + \Delta y^k + \epsilon_k\,,$$

where $y^0 = x^0$ is given, Δy^k is understood to be the exact Newton correction at y^k, and

$$\|\epsilon_k\| \leq \delta \|\Delta y^k\|\,,$$

Upon using analytical tools of Section 8.1, derive iterative error bounds for $\|y^k - x^k\|$ and $\|y^k - x^*\|$.

Exercise 8.3 Consider the nonlinear ODE boundary value problem

$$\dot{x} = f(x)\,, \quad Ax(a) + Bx(b) = 0$$

with linear separable boundary conditions. We want to study asymptotic mesh independence for Gauss collocation methods of order $s \geq 1$ (compare Section 7.4). For the approximating space we select $X = W^{1,\infty}$ and impose the assumptions from Section 8.1. Let $X_j \subset X$ denote a finite dimensional subspace characterizing the collocation discretization with maximum mesh size τ_j. Assume that f is sufficiently smooth and the BVP is well-conditioned for all required arguments.

a) In view of (8.5), derive upper bounds δ_j such that

$$\|\Delta x_j - \Delta x\|_{W^{1,\infty}} \leq \delta_j$$

with the asymptotic property

$$\delta_j \to 0.$$

Hint: Compare the exact solution w of

$$\dot{w} + f_x(x_j)w = -f_x(x_j)x_j + f(x_j) \,, \quad Aw(a) + Bw(b) = 0$$

and its approximation w_j using the error estimate (as given in [180, 49])

$$\|w - w_j\|_{W^{1,\infty}} \le C\tau_j \|\ddot{w}\|_\infty \,,$$

where C is a bounded generic constant, which is independent of j.

b) Under the assumptions of Theorem 2.2 derive some bound

$$\|w - w_j\| \le \sigma_j \|v_j\|^2$$

with the asymptotic property

$$\sigma_j \to 0 \,.$$

Exercise 8.4 We consider linear finite element approximations on quasi-uniform triangulations for semilinear elliptic boundary value problems

$$F(x) = -\mathrm{div}\nabla x - f(x) = 0, \quad x \in H_0^1(\Omega)$$

on convex polygonal domains $\Omega \subset \mathbb{R}^d$, $d \le 3$. For this setting, we want to study asymptotic mesh independence. The notation is as in Section 8.1.

In view of (8.5) and (8.9), derive upper bounds δ_j such that

$$\|\Delta x_j - \Delta x\| \le \delta_j$$

and σ_j such that

$$\|w - w_j\| \le \sigma_j \|v_j\|^2 \,,$$

Assume that the above right hand term $f : \mathbb{R} \to \mathbb{R}$ is globally Lipschitz continuously differentiable. In particular, show that for the process of successive refinement the asymptotic properties

$$\lim_{j\to\infty} \delta_j = 0 \,, \quad \lim_{j\to\infty} \sigma_j = 0$$

hold.

Hint: Exploit the H^2-regularity of $x_j + \Delta x$ and use the embedding $H^1 \hookrightarrow L_4$.