3 Systems of Equations: Global Newton Methods

As in the preceding chapter, the discussion here is also restricted to systems of n nonlinear equations, say

$$F(x) = 0\,,$$

where $F \in C^1(D)$, $D \subseteq \mathbb{R}^n$, $F : D \longrightarrow \mathbb{R}^n$ with Jacobian (n,n)-matrix $F'(x)$. In contrast to the preceding chapter, however, available initial guesses x^0 of the solution point x^* are no longer assumed to be 'sufficiently close' to x^*.

In order to specify the colloquial term 'sufficiently close', we recur to any of the local convergence conditions of the preceding chapter. Let ω denote an affine covariant Lipschitz constant. Then Theorem 2.3 presents an appropriate local convergence condition of the form

$$\|x^* - x^0\| < 2/\omega\,.$$

In the error oriented framework, Theorem 2.2 yields a characterization in terms of the Kantorovich quantity

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

which restricts the ordinary Newton correction Δx^0. Under any of these conditions local Newton methods are guaranteed to converge. Such problems are sometimes called *mildly* nonlinear. Their computational complexity is *a priori bounded* in terms of the computational complexity of solving linear problems of the same structure—see, for example, the bound (2.71).

In contrast to that, under a condition of the type $h_0 \gg 1$, which is equivalent to

$$\|\Delta x^0\| \gg 2/\omega \tag{3.1}$$

local Newton methods will not exhibit guaranteed convergence. In this situation, the computational complexity cannot be bounded a priori. Such problems are often called *highly* nonlinear. Nevertheless, local Newton methods may actually converge for some of these problems even in the situation of condition (3.1). A *guaranteed* convergence, however, will only occur, if additional *global structure* on F can be exploited: as an example, we treat *convex* nonlinear mappings in Section 3.1.1 below.

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_3, © Springer-Verlag Berlin Heidelberg 2011

For general mapping F, a *globalization* of local Newton methods must be constructed. In Section 3.1 we survey globalization concepts such as

- steepest descent methods,
- trust region methods,
- the Levenberg-Marquardt method, and
- the Newton method with damping strategy.

In Section 3.1.4, a rather general geometric approach is taken: the idea is to derive a globalization concept without pre-occupation to any of the known iterative methods, just starting from the requirement of affine covariance as a 'first principle'. Surprisingly, this general approach leads to the derivation of Newton's method with damping strategy.

Monotonicity tests. Monotonicity tests serve the purpose to accept or reject a new iterate. We study different such tests, according to different affine invariance requirements:

- the most popular *residual* monotonicity test, which is based on affine contravariance (Section 3.2),
- the error oriented so-called *natural* monotonicity test, which is based on affine covariance (Section 3.3), and
- the convex functional test as the natural requirement in convex optimization, which reflects affine conjugacy (Section 3.4).

For each of these three affine invariance classes, *adaptive trust region strategies* are designed in view of an efficient choice of damping factors in Newton's method. They are all based on the *paradigm* already mentioned at the end of Section 1.2. On a theoretical basis, details of algorithmic realization in combination with either *direct* or *iterative* linear solvers are worked out. As it turns out, an efficient determination of the steplength factor in global Newton methods is intimately linked with the accuracy matching for affine invariant combinations of inner and outer iteration within various *inexact* Newton methods.

3.1 Globalization Concepts

Efficient iterative methods should be able to cope with 'bad' guesses x^0. In this section we survey methods that permit rather general initial guesses x^0, not only those sufficiently close to the solution point x^*. Of course, such methods should merge into local Newton techniques as soon as the iterates x^k come 'close to' the solution point x^*—to exploit the local *quadratic* or *superlinear* convergence property.

Parameter continuation methods. The simplest way of globalization of local Newton methods is to embed the given problem $F(x) = 0$ into a one-parameter family of problems, a so-called *homotopy*

$$F(x, \tau) = 0\,, \quad \tau \in [0, 1]\,,$$

such that the starting point x^0 is the solution for $\tau = 0$ and the desired solution point x^* is the solution point for $\tau = 1$. If we choose sufficiently many intermediate problems in the *discrete homotopy*

$$F(x, \tau_\nu) = 0\,, \quad 0 = \tau_0 < \cdots < \tau_\nu < \cdots < \tau_N = 1\,,$$

then the solution point of one problem can serve as initial guess in a local Newton method for the next problem. In this way, global convergence can be assured under the assumption that existence and uniqueness of the solution along the *homotopy path* is guaranteed. In this context, questions like the adaptive choice of the stepsizes $\Delta\tau_\nu$ or the computation of *bifurcation diagrams* are of interest. An efficient choice of the embedding will exploit specific features of the given problem to be solved—with consequences for the local uniqueness of the solution along the homotopy path and for the computational speed of the discrete continuation process. The discussion of these and many related topics is postponed to Chapter 5.

Pseudo-transient continuation methods. Another continuation method uses the embedding of the algebraic equation into an *initial value problem* of the type

$$x' = F(x)\,, \qquad x(0) = x^0\,.$$

Discretization of this problem with respect to a timestep τ by the *explicit Euler method* leads to the fixed point iteration

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

or, by the *linearly implicit Euler method* to the iteration scheme

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

Note that for $\tau \to \infty$ the latter scheme merges into the ordinary Newton method. The scheme reflects *affine similarity* as described in Section 1.2 and will be treated in detail in Section 6.4 in the context of so-called pseudo-transient continuation methods, which are a special realization of stiff integrators for ordinary differential equations.

3.1.1 Componentwise convex mappings

The Newton-Raphson method for scalar equations (see Figure 1.1) may be geometrically interpreted as taking the intersection of the local tangent with

the axis—and repeating this process until sufficient accuracy is achieved. In this interpretation, the simplified Newton method just means to keep the initial tangent throughout the whole iterative process. From this it can be directly seen that both the ordinary and the simplified Newton method converge *globally* for *convex* (or concave) scalar functions. The convergence is *monotone*, i.e., the iterates x^k approach the solution point x^* from one side only. On the basis of this insight we are now interested in a generalization of such a monotonicity property to systems. General convex minimization problems, which lead to gradient mappings F, will be treated in the subsequent Section 3.4. Here we concentrate on some componentwise convexity as discussed in the textbook of J.M. Ortega and W.C. Rheinboldt [163].

Such componentwise convex mappings F may be characterized by one of the following equivalent properties (let $x,\, y \in D \subseteq \mathbb{R}^n$, D convex, $\lambda \in [0,1]$):

$$\begin{aligned} F\big(\lambda x + (1-\lambda)y\big) &\leq \lambda F(x) + (1-\lambda)F(y)\,, \qquad (3.2)\\ F(y) - F(x) &\geq F'(x)(y-x)\,,\\ \big(F'(y) - F'(x)\big)(y-x) &\geq 0\,. \qquad (3.3) \end{aligned}$$

Herein, the inequalities are understood componentwise. Since the objects of interest will be the iterates, we miss *affine covariance* in the above formulation. In fact, Ortega and Rheinboldt show monotone convergence of the ordinary Newton method under the *additional* assumption

$$F'(z)^{-1} \geq 0\,,\ z \in D\,, \qquad (3.4)$$

which is essentially a global *M-matrix property* (cf. R.S. Varga [192]) for the Jacobian. Upon combining the above three equivalent convexity characterizations with (3.4), we obtain the three equivalent affine covariant formulations

$$\begin{aligned} F'(z)^{-1}F\big(\lambda x + (1-\lambda)y\big) &\leq F'(z)^{-1}\big(\lambda F(x) + (1-\lambda)F(y)\big) \qquad (3.5)\\ F'(x)^{-1}\big(F(y) - F(x)\big) &\geq y - x \qquad (3.6)\\ F'(z)^{-1}\big(F'(y) - F'(x)\big)(y-x) &\geq 0\,. \qquad (3.7) \end{aligned}$$

Note that these conditions cover any mapping F such that (3.2) up to (3.3) together with (3.4) hold for AF and AF' with some $A \in \mathrm{GL}(n)$.

Lemma 3.1 *Let $F : D \longrightarrow \mathbb{R}^n$ be a continuously differentiable mapping with $D \subseteq \mathbb{R}^n$ open and convex. Let this mapping satisfy one of the convexity characterizations* (3.5)-(3.7)*. Then the ordinary Newton iteration starting at some $x^0 \in D$ converges monotonically and globally such that componentwise*

$$x^* \leq x^{k+1} \leq x^k\,, \quad k = 1, 2, \ldots\,. \qquad (3.8)$$

Proof. For the *ordinary* Newton iteration, one obtains:

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

$$= -F'(x^k)^{-1} \int_{\delta=0}^{1} \left[F'\big(x^{k-1} + \delta(x^k - x^{k-1})\big) - F'(x^{k-1})\right](x^k - x^{k-1})d\delta\,.$$

Insertion of (3.7) for

$$z = x^k, x = x^{k-1}, y = x^{k-1} + \delta(x^k - x^{k-1}), x^k - x^{k-1} = (y - x)/\delta$$

leads to

$$x^{k+1} - x^k \leq 0 \;\text{ for } k \geq 1\,.$$

In a similar way, one derives

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

which, by application of (3.6), supplies

$$x^{k+1} - x^* \geq 0\,,\; k \geq 0\,.$$

The rest of the proof can be found in [163], p. 453. □

Remark 3.1 An immediate generalization of this lemma is obtained by allowing *different* inequalities for different components in (3.5) to (3.7)—which directly leads to the corresponding inequalities in (3.8).

Note that the above results do *not* apply to the *simplified* Newton iteration, unless $n = 1$: following the lines of the above proof, the application of (3.7) here would lead to

$$x^{k+1} - x^k \leq -F'(x^0)^{-1}\big(F'(x^{k-1}) - F'(x^0)\big)(x^k - x^{k-1})\,.$$

In order to apply (3.7) once more, a relation of the kind

$$x^{k-1} - x^0 = \Theta \cdot (x^k - x^{k-1})$$

for some $\Theta > 0$ would be required—which will only hold in $\mathbb{R}^1$.

In actual computation, the global monotone convergence property does not require any control in terms of some monotonicity test. Only reasonable componentwise termination criteria need to be implemented. It is worth mentioning that this special type of convergence of the ordinary Newton method does *not* mean global *quadratic* convergence: rather this type of convergence may be arbitrarily slow, as can be verified in simple scalar problems—see Exercise 1.3. Not even an a-priori estimation for the number of iterations needed to achieve a prescribed accuracy may be possible.

Bibliographical Note. The above componentwise monotonicity results are discussed in detail in the classical monograph [163] by J.M. Ortega and W.C. Rheinboldt, there in not affine invariant form. In 1987, F.A. Potra and W.C. Rheinboldt proved affine invariant conditions, under which the simplified Newton method and other Newton-like methods converge, see [172, 170].

3.1.2 Steepest descent methods

A *desirable requirement* for any iterative methods would be that the iterates x^k successively *approach* the solution point x^*—which may be written as

$$\|x^{k+1} - x^*\| < \|x^k - x^*\|\,, \quad \text{if } x^k \neq x^*\,.$$

Local Newton techniques implicitly realize such a criterion under affine covariant theoretical assumptions, as has been shown in detail in the preceding chapter. *Global* methods, however, require a *substitute approach criterion*, which may be based on the *residual level function*

$$T(x) := \tfrac{1}{2} \left\| F(x) \right\|_2^2 \equiv \tfrac{1}{2} F(x)^T F(x)\,. \tag{3.9}$$

Such a function has the property

$$\begin{aligned} T(x) &= 0 \Longleftrightarrow x = x^*\,, \\ T(x) &> 0 \Longleftrightarrow x \neq x^*\,. \end{aligned} \tag{3.10}$$

In terms of this level function, the approach criterion may be formulated as a *monotonicity criterion*

$$T(x^{k+1}) < T(x^k)\,, \quad \text{if } T(x^k) \neq 0\,.$$

Associated with the level function are the so-called *level sets*

$$G(z) := \{x \in D \subseteq \mathbb{R}^n |\; T(x) \leq T(z)\}\,.$$

Let $\overset{\circ}{G}$ denote the interior of G. Then property (3.10) implies

$$x^* \in G(x)\,,\ x \in D$$

and the monotonicity criterion may be written in geometric terms as

$$x^{k+1} \in \overset{\circ}{G}(x^k)\,, \quad \text{if } \overset{\circ}{G}(x^k) \neq \emptyset\,.$$

An intuitive approach based on this geometrical insight, which dates back even to A. Cauchy [44] in 1847, is to choose the *steepest descent* direction as the direction of the iterative correction—see Figure 3.1.

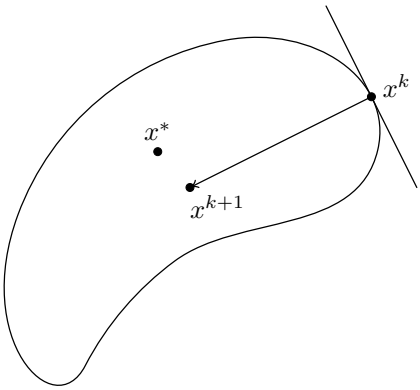

Fig. 3.1. Geometric interpretation: level set and steepest descent direction.

This idea leads to the following iterative method:

$$\begin{aligned} \Delta x^k &:= -\operatorname{grad} T(x^k) = -F'(x^k)^T F(x^k) && (3.11)\\ x^{k+1} &:= x^k + s_k \Delta x^k && (3.12)\\ s_k &> 0: \text{ steplength parameter.} \end{aligned}$$

Figure 3.1 also nicely shows the so-called *downhill property*.

Lemma 3.2 *Let $F : D \longrightarrow \mathbb{R}^n$ be a continuously differentiable mapping with $D \subseteq \mathbb{R}^n$. Dropping the iterative index k in the notation of* (3.11)*, let $\Delta x \neq 0$. Then there exists a $\mu > 0$ such that*

$$T(x + s\Delta x) < T(x)\,,\ 0 < s < \mu\,. \tag{3.13}$$

Proof. Define $\varphi(s) := T(x + s\Delta x)$. As $F \in C^1(D)$, one has $\varphi \in C^1(D_1)$, $D_1 \subseteq \mathbb{R}^1$. Then

$$\varphi'(0) = \left(\operatorname{grad} T(x + s \cdot \Delta x)^T \Delta x\right) \Big|_{s=0} = -\|\Delta x\|_2^2 < 0\,.$$

With $\varphi \in C^1$, the result (3.13) is established. □

Steplength strategy. This result is the theoretical basis for a strategy to select the steplength in method (3.12). It necessarily consists of two parts: a reduction strategy and a prediction strategy. The *reduction strategy* applies whenever

$$T(x^k + s_k^0 \Delta x^k) > T(x^k)$$

for some given parameter s_k^0. In this case, the above monotonicity test is repeated with some

$$s_k^{i+1} := \kappa \cdot s_k^i\,, \quad i = 0, 1, \ldots, \qquad \kappa < 1 \quad (\text{typically } \kappa = 1/2)\,.$$

Lemma 3.2 assures that a *finite* number i^* of reductions will ultimately lead to a feasible steplength factor $s_k^* > 0$. The *prediction strategy* applies, when s_{k+1}^0 must be selected—usually based on an ad-hoc rule that takes the steplength history into account such as

$$s_{k+1}^0 := \begin{cases} \min\left(s_{\max}, s_k^*/\kappa\right), & \text{if } s_{k-1}^* \le s_k^* \\ s_k^* & \text{else} \end{cases} . \tag{3.14}$$

The possible increase from s_k to s_{k+1} helps to avoid inefficiency coming from 'too small' local corrections. One may also aim at implementing an optimal choice of s_k out of a sequence of sample values—a strategy, which is often called *optimal line search*. However, since s_k may range from 0 to ∞, a reasonable set of values to be sampled may be hard to define, if a sufficiently large class of problems is to be considered.

Convergence properties. An elementary convergence analysis shows that the iterative scheme (3.12) with steplength strategies like (3.14) converges *linearly* even for rather bad initial guesses x^0—however, possibly arbitrarily slow. Moreover, so-called 'pseudo-convergence' characterized by

$$\|F'(x)^T F(x)\| \quad \text{'small'}$$

may occur far from the solution point due to local ill-conditioning of the Jacobian matrix.

General level functions. In a large class of problems, the described difficulties are a consequence of the fact that the whole scheme is not affine covariant so that the choice of $T(x)$ as a level function appears to be rather arbitrary. In principle, *any* level function

$$T(x|A) := \tfrac{1}{2} \left\| AF(x) \right\|_2^2 \tag{3.15}$$

with arbitrary nonsingular (n, n)-matrix A could be used in the place of $T(x)$ above. To make things worse, even though the direction of steepest descent Δx is 'downhill' with respect to $T(x)$, there nearly always exists a matrix A such that Δx is 'uphill' with respect to $T(x|A)$, as will be shown in the following lemma.

Lemma 3.3 *Let $\Delta x = -\operatorname{grad} T(x)$ denote the direction of steepest descent with respect to the level function $T(x)$ as defined in* (3.9)*. Then, unless*

$$F'(x)\Delta x = \chi \cdot F(x), \quad \textit{for some } \chi < 0, \tag{3.16}$$

there exists a class of nonsingular matrices A such that

$$T(x + s\Delta x|A) > T(x|A), \; 0 < s < \nu,$$

for some $\nu > 0$.

Proof. Let $F = F(x), J = F'(x), \overline{J} = JJ^T, \overline{A} = A^T A$. Then

$$\Delta x^T \operatorname{grad} T(x|A) = -F^T \overline{J}\, \overline{A} F\,.$$

Now, choose

$$\overline{A} := \overline{J} + \mu y y^T$$

with some $\mu > 0$ to be specified later and $y \in \mathbb{R}^n$ such that

$$F^T(\overline{J} + I)y = 0\,, \quad \text{but} \quad F^T y \neq 0\,.$$

Here the assumption (3.16) enters for any $\chi \in \mathbb{R}^1$. By definition, however, the choice $\chi \geq 0$ is impossible, since (3.16) implies that

$$\chi = -\frac{\|J^T F\|^2}{\|F\|^2} < 0\,.$$

Hence, for the above choice of $\overline{A}$, we obtain

$$\Delta x^T \operatorname{grad} T(x|A) = -\|\overline{J}F\|_2^2 + \mu (F^T y)^2\,.$$

Then the specification

$$\mu > \|\overline{A}F\|_2^2 / (F^T y)^2$$

leads to

$$\Delta x^T \operatorname{grad} T(x|A) > 0\,,$$

which, in turn, implies the statement of the lemma. □

Summarizing, even though the underlying geometrical idea of steepest descent methods is intriguing, the technical details of implementation cannot be handled in a theoretically satisfactory manner, let alone in an affine covariant setting.

3.1.3 Trust region concepts

As already shown in Section 1.1, the ordinary Newton method can be algebraically derived by linearization of the nonlinear equation around the solution point x^*. This kind of derivation supports the interpretation that the Newton correction is useful only in a close neighborhood of x^*. Far away from x^*, such a linearization might still be trusted in some 'trust region' around the current iterate x^k. In what follows we will present several models defining such a region. For a general survey of trust region methods in optimization see, e.g., the book [45] by A.R. Conn, N.I.M. Gould, and P.L. Toint.

Levenberg-Marquardt model. The above type of elementary consideration led K.A. Levenberg [143] and later D.W. Marquardt [147]) to suggest a modification of Newton's method for 'bad' initial guesses that merges into the ordinary Newton method close to the solution point. Following the presentation by J.J. Moré in [152] we define a correction vector Δx (dropping the iteration index k) by the constrained quadratic minimization problem:

$$\|F(x) + F'(x)\Delta x\|_2 = \min$$

subject to the constraint

$$\|\Delta x\|_2 \leq \delta$$

in terms of some prescribed parameter $\delta > 0$, which may be understood to quantify the *trust region* in this approach.

The trust region constraint may be treated by the introduction of a *Lagrange multiplier* $p \geq 0$ subject to

$$p\left(\|\Delta x\|_2^2 - \delta^2\right) = 0,$$

which yields the equivalent unconstrained quadratic optimization problem

$$\|F(x^0) + F'(x^0)\Delta x\|_2^2 + p\|\Delta x\|_2^2 = \min\,.$$

After a short calculation and re-introduction of the iteration index k, we then end up with the *Levenberg-Marquardt method:*

$$\left(F'(x^k)^T F'(x^k) + pI\right)\Delta x^k = -F'(x^k)^T F(x^k)\,,\ x^{k+1} := x^k + \Delta x^k\,. \quad (3.17)$$

The correction vector $\Delta x^k(p)$ has two interesting limiting cases:

$$\begin{aligned} p \to 0^+ \quad &: \quad \Delta x^k(0) = -F'(x^k)^{-1}F(x^k)\,, \quad \text{if } F'(x^k) \text{ nonsingular} \\ p \to \infty \quad &: \quad \Delta x^k(p) \to -\frac{1}{p}\,\mathrm{grad}\,T(x^k)\,. \end{aligned}$$

In other words: Close to the solution point, the method merges into the ordinary Newton method; far from the solution point, it turns into a steepest descent method with steplength parameter $1/p$.

Trust region strategies for the Levenberg-Marquardt method. All strategies to choose the parameter p or, equivalently, the parameter δ are based on the following simple lemma.

Lemma 3.4 *Under the usual assumptions of this section let $\Delta x(p) \neq 0$ denote the Levenberg-Marquardt correction defined in* (3.17)*. Then there exists a $p_{\min} \geq 0$ such that*

$$T\big(x + \Delta x(p)\big) < T(x)\,,\ p > p_{\min}\,.$$

Proof. Substitute $q := 1/p\,,\ 0 \le q \le \infty\,,$ and define

$$\varphi(q) := T\big(x + \Delta x(1/q)\big)\,,\ \varphi(0) = T(x)\,.$$

Then

$$\varphi'(0) = 0\,,\ \varphi''(0) < 0\,.$$

Hence, there exists a $q_{\max} = 1/p_{\min}$ such that

$$\varphi(q) < \varphi(0)\,,\ 0 < q < q_{\max}\,.$$

□

The method looks rather robust, since for any $p > 0$ the matrix $J^T J + pI$ is nonsingular, even when the Jacobian J itself is singular. Nevertheless, similar as the steepest descent method, the above iteration may also terminate at 'small' gradients, since for singular J the right-hand side of (3.17) also degenerates. This latter property is often overlooked both in the literature and by users of the method. Since the Levenberg-Marquardt method lacks affine invariance, special scaling methods are often recommended.

Bibliographical Note. Empirical trust region strategies for the Levenberg-Marquardt method have been worked out, e.g., by M.D. Hebden [118], by J.J. Moré [152], or by J.E. Dennis, D.M. Gay, and R. Welsch [54]. The associated codes are rather popular and included in several mathematical software libraries. However, as already stated above, these algorithms may terminate at a wrong solution with small gradient. When more than one solution exists locally, these algorithms might not indicate that. The latter feature is particularly undesirable in the application of the Levenberg-Marquardt method to nonlinear least squares problems—for details see Chapter 4 below.

Affine covariant trust region model. A straightforward affine covariant reformulation of the Levenberg-Marquardt model would be the following constrained quadratic optimization problem:

$$\|F'(x^0)^{-1}\big(F(x^0) + F'(x^0)\Delta x\big)\|_2 = \min$$

subject to the constraint

$$\|\Delta x\|_2 \le \delta_0\,. \tag{3.18}$$

This problem can easily be solved geometrically—as shown in Figure 3.2, where the constraint (3.18) is represented by a sphere around the current iterate x^0 with radius δ_0. Whenever δ_0 exceeds the length of the ordinary Newton correction, which means that the constraint is *not active*, then Δx is just the ordinary Newton correction—and the quadratic functional vanishes. Whenever the constraint is *active*, then the direction of Δx still is the Newton

direction, but now with reduced steplength. This leads to the Newton method with so-called *damping*

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

wherein the damping factor varies in the range $0 < \lambda_k \leq 1$.

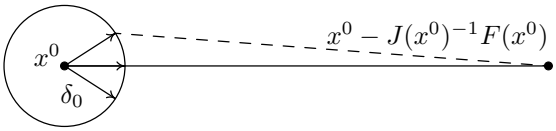

Fig. 3.2. Geometric interpretation: affine covariant trust region model.

Affine contravariant trust region model. An affine contravariant reformulation of the Levenberg-Marquardt model would lead to a constrained quadratic optimization problem of the form:

$$\|F(x^0) + F'(x^0)\Delta x\| = \min$$

subject to the constraint

$$\|F'(x^0)\Delta x\|_2 \leq \delta_0\,.$$

Once again, the problem can be solved geometrically by Figure 3.2: only the terms of the domain space of F must be reinterpreted by the appropriate terms in the image space of F. As a consequence, the Newton method with damping is obtained again.

Damping strategies for Newton method. All strategies for choosing the above damping factors λ_k are based on the following insight.

Lemma 3.5 *Under the usual assumptions of this section let $F'(x)$ be nonsingular and $F \neq 0$. With Δx defined to be the Newton direction there exists some $\mu > 0$ such that*

$$T(x + \lambda\Delta x) < T(x)\,,\ 0 < \lambda < \mu\,.$$

Proof. As before, let $F = F(x), J = F'(x)$ and define $\varphi(\lambda) := T(x + \lambda\Delta x)$, which then yields

$$\varphi(0) = T(x)\,,\ \varphi'(0) = (J^T F)^T \Delta x = -F^T F = -2T(x) < 0\,.$$

□

Among the most popular empirical damping strategies is the

Armijo strategy [7]. Let $\Lambda_k \subset \left\{1, \frac{1}{2}, \frac{1}{4}, \ldots, \lambda_{\min}\right\}$ denote a sequence such that

$$T(x^k + \lambda_k \Delta x^k) \leq \left(1 - \tfrac{1}{2}\lambda_k\right) T(x^k)\,, \quad \lambda \in \Lambda_k \tag{3.19}$$

holds and define an optimal damping factor via

$$T(x^k + \lambda_k \Delta x^k) = \min_{\lambda \in \Lambda_k} T(x^k + \lambda \Delta x^k)\,.$$

In order to avoid overflow in critical examples, the above evaluation of T will be sampled from the side of small values λ. In a neighborhood of x^*, this strategy will produce $\lambda = 1$. If $\lambda < \lambda_{\min}$ would be required, the iteration should be terminated with a warning. Unfortunately, the latter occurrence appears quite frequently in realistic problems of scientific computing, especially when the arising Jacobian matrices are ill-conditioned. This failure is a consequence of the fact that the choice $T(x)$ for the level function destroys the affine covariance of the local Newton methods—a consequence that will be analyzed in detail in Section 3.3 below.

3.1.4 Newton path

All globalization techniques described up to now were based on the requirement of local monotonicity with respect to the standard level function $T(x) = T(x|I)$ as defined in (3.9). In this section we will follow a more general approach, which covers general level functions $T(x|A)$ for *arbitrary* nonsingular matrix A as defined in (3.15). The associated *level sets* are written as

$$G(z|A) := \left\{x \in D \subseteq \mathbb{R}^n |\; T(x|A) \leq T(z|A)\right\}. \tag{3.20}$$

With this notation, iterative monotonicity with respect to $T(x|A)$ can be written in the form

$$x^{k+1} \in \overset{\circ}{G}(x^k|A)\,, \quad \text{if } \overset{\circ}{G}(x^k|A) \neq \emptyset\,.$$

We start from the observation that each choice of the matrix A could equally well serve within an iterative method. With the aim of getting rid of this somewhat arbitrary choice, we now focus on the intersection of all corresponding level sets:

$$\overline{G}(x) := \bigcap_{A \in \mathrm{GL}(n)} G(x|A)\,. \tag{3.21}$$

By definition, the thus defined geometric object is affine covariant. Its nature will be revealed by the following theorem.

Theorem 3.6 *Let $F \in C^1(D)$, $D \subseteq \mathbb{R}^n$, $F'(x)$ nonsingular for all $x \in D$. For some $\widehat{A} \in \mathrm{GL}(n)$, let the path-connected component of $G(x^0|\widehat{A})$ in x^0 be compact and contained in D. Then the path-connected component of $\overline{G}(x^0)$*

as defined in (3.21) *is a topological path* $\overline{x} : [0,2] \to \mathbb{R}^n$*, the so-called Newton path, which satisfies*

$$F\big(\overline{x}(\lambda)\big) = (1-\lambda)F(x^0)\,, \tag{3.22}$$

$$T\big(\overline{x}(\lambda)|A\big) = (1-\lambda)^2 T(x^0|A)\,, \tag{3.23}$$

$$\frac{d\overline{x}}{d\lambda} = -F'(\overline{x})^{-1}F(x^0)\,, \tag{3.24}$$

$$\overline{x}(0) = x^0\,,\ \overline{x}(1) = x^*\,,$$

$$\left.\frac{d\overline{x}}{d\lambda}\right|_{\lambda=0} = -F'(x^0)^{-1}F(x^0) \equiv \Delta x^0\,, \tag{3.25}$$

where Δx^0 *is the ordinary Newton correction.*

Proof. Let $F_0 = F(x^0)$. In a *first stage of the proof*, level sets and their intersection are defined in the *image space* of F using the notation

$$H(x^0|A) := \big\{y \in \mathbb{R}^n |\ \|Ay\|_2^2 \le \|AF_0\|_2^2\big\}\,,$$

$$\overline{H}(x^0) := \bigcap_{A\in \mathrm{GL}(n)} H(x^0|A)\,.$$

Let σ_i denote the singular values of A and q_i the eigenvectors of A^TA such that

$$A^TA = \sum_{i=1}^n \sigma_i^2 q_i q_i^T\,.$$

Select those A with $q_1 := F_0/\|F_0\|_2$, which defines the matrix set:

$$\mathcal{A} := \big\{A \in \mathrm{GL}(n) | A^TA = \sum_{i=1}^n \sigma_i^2 q_i q_i^T\,,\ q_1 = F_0/\|F_0\|_2\big\}\,.$$

Then every $y \in \mathbb{R}^n$ can be represented by

$$y = \sum_{j=1}^n b_j q_j\,,\ b_j \in \mathbb{R}\,.$$

Hence

$$\|Ay\|_2^2 = y^TA^TAy = \sum_{i=1}^n \sigma_i^2 b_i^2\,,$$

$$\|AF_0\|_2^2 = \sigma_1^2\|F_0\|_2^2\,,$$

which, for $A \in \mathcal{A}$, yields

$$H(x^0|A) = \left\{ y \;\middle|\; \sum_{i=1}^{n} \sigma_i^2 b_i^2 \le \sigma_1^2 \|F_0\|_2^2 \right\},$$

or, equivalently, the n-dimensional ellipsoids

$$\frac{1}{\|F_0\|_2^2} b_1^2 + \left(\frac{\sigma_2}{\sigma_1 \|F_0\|_2} \right)^2 b_2^2 + \cdots + \left(\frac{\sigma_n}{\sigma_1 \|F_0\|_2} \right)^2 b_n^2 \le 1 .$$

For $A \in \mathcal{A}$, all corresponding ellipsoids have a common b_1-axis of length $\|F_0\|_2$—see Figure 3.3. The other axes are arbitrary.

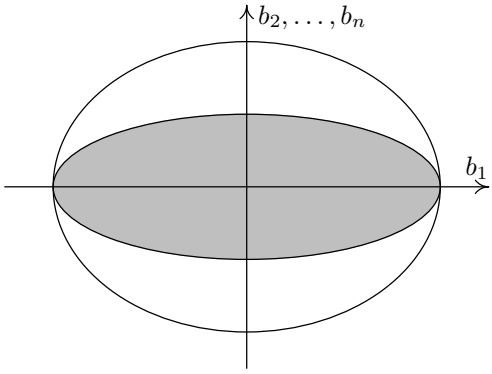

Fig. 3.3. Intersection of ellipsoids $H(x^0|A)$ for $A \in \mathcal{A}$.

Figure 3.3 directly shows that

$$\begin{aligned} \widehat{H}(x^0) &:= \bigcap_{A \in \mathcal{A}} H(x^0|A) = \{ y = b_1 q_1 \mid |b_1| \le \|F_0\|_2 \} \\ &= \{ y \in \mathbb{R}^n | y = (1-\lambda) F_0 \,,\ \lambda \in [0,2] \} \\ &= \{ y \in \mathbb{R}^n | Ay = (1-\lambda) A F_0 \,,\ \lambda \in [0,2] \,,\ A \in \mathrm{GL}(n) \} . \end{aligned}$$

As $\mathcal{A} \subset \mathrm{GL}(n) : \overline{H}(x^0) \subseteq \widehat{H}(x^0)$. On the other hand, for $y \in \widehat{H}(x^0)$, $A \in \mathrm{GL}(n)$, one has

$$\|Ay\|_2^2 = (1-\lambda)^2 \|AF_0\|_2^2 \le \|AF_0\|_2^2 ,$$

which implies $\widehat{H}(x^0) \subseteq \overline{H}(x^0)$ and, in turn, confirms

$$\widehat{H}(x^0) = \overline{H}(x^0) .$$

The *second stage of the proof* now involves 'lifting' of the path $\overline{H}(x^0)$ to $\overline{G}(x^0)$. This is done by means of the *homotopy*

$$\Phi(x, \lambda) := F(x) - (1 - \lambda)F(x^0)\,.$$

Note that

$$\Phi_x = F'(x)\,,\ \Phi_\lambda = F(x^0)\,.$$

Hence, Φ_x is nonsingular for $x \in D$. As $D \supset G(x^0|\widehat{A}\,)$, local continuation starting at $\overline{x}(0) = x^0$ by means of the *implicit function theorem* finally establishes the existence of the path

$$\overline{x} \subset G(x^0|\widehat{A}\,) \subset D\,,$$

which is defined by (3.22) from $\Phi \equiv 0$. The differentiability of $\overline{x}$ follows, since $F \in C^1(D)$, which confirms (3.24) and (3.25). □

The above theorem deserves some contemplation. The constructed Newton path $\bar{x}$ is outstanding in the respect that *all level functions* $T(x|A)$ *decrease along* $\bar{x}$—this is the result (3.22). Therefore, a rather natural approach would be to just follow that path computationally—say, by numerical integration of the initial value problem (3.24). Arguments, why this is *not* a recommended method of choice, will be presented in Section 5 in a more general context. Rather, the local information about the tangent direction

$$\frac{\Delta x^0}{\|\Delta x^0\|}$$

should be used—which is just the Newton direction. In other words:

Even 'far away' from the solution point x^*, the Newton direction is an outstanding direction, only its length may be 'too large' for highly nonlinear problems.

Such an insight could not have been gained from the merely algebraic local linearization approach that had led to the ordinary Newton method.

The assumptions in the above theorem deliberately excluded the case that the Jacobian may be singular at some $\widehat{x}$ close to x^0. This case, however, may and will occur in practice. Application of the implicit function theorem in a more general situation shows that all Newton paths starting at points x^0 will end at one of the following three classes of points—see the schematic Figure 3.4:

- at the 'nearest' solution point x^*, or
- at some sufficiently close *critical point* $\widehat{x}$ with singular Jacobian, or
- at some point on the boundary ∂D of the domain of F.

The situation is also illustrated in the rather simple, but intuitive Example 3.2, see Figure 3.10 below.

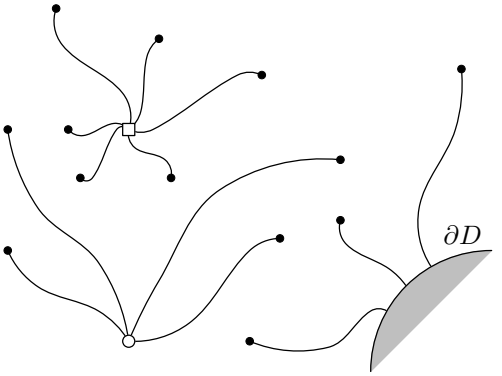

Fig. 3.4. Newton paths starting at initial points x^0 (•) will end at a solution point x^* (○), at a critical point $\widehat{x}$ (□), or on the domain boundary ∂D.

Bibliographical Note. Standard derivations of the Newton path as a mathematical object had started from the so–called continuous analog of Newton's method, which is the ODE initial value problem (3.23)—see, e.g., the 1953 paper [48] by D. Davidenko. The geometric derivation of the Newton path from affine covariance as a 'first principle' dates back to the author's dissertation [59, 60] in 1972.

3.2 Residual Based Descent

In this section we study the *damped Newton iteration*

$$F'(x^k)\Delta x^k = -F(x^k), \qquad x^{k+1} = x^k + \lambda_k \Delta x^k, \qquad \lambda_k \in]0, 1]$$

under the requirement of *residual contraction*

$$\|F(x^{k+1})\| < \|F(x^k)\|\,,$$

which is certainly the most popular and the most widely used global convergence measure.

From Section 3.1.4 we perceive this iterative method as the tangent deviation from the *Newton path*, which connects the given initial guess x^0 to the unknown solution point x^*—under sufficient regularity assumptions on the Jacobian matrix, of course. The deviation is theoretically characterized by means of *affine contravariant* Lipschitz conditions as defined in the convergence theory for residual based *local* Newton methods in Section 2.2.

In what follows, we derive *theoretically optimal* iterative damping factors and prove global convergence within some range around these optimal factors

(Section 3.2.1). On this basis we then develop residual based trust region strategies for the algorithmic choice of the damping factors. This is first done for the *exact* Newton correction Δx as defined above (Section 3.2.2) and second for an *inexact* variant using the iterative solver `GMRES` for the inner iteration (Section 3.2.3).

3.2.1 Affine contravariant convergence analysis

From Lemma 3.5 above we already know that the Newton correction Δx^k points *downhill* with respect to the *residual level function*

$$T(x) := \tfrac{1}{2}\|F(x)\|_2^2$$

and therefore into the interior of the associated *residual level set*

$$G(x) := \{y \in D | T(y) \le T(x)\}\,.$$

At a given iterate x^k, we are certainly interested to determine some steplength (defined by the associated damping factor λ_k) along the Newton direction such that the residual reduction is in some sense *optimal*.

Theorem 3.7 *Let $F \in C^1(D)$ with $D \subset \mathbb{R}^n$ open convex and $F'(x)$ nonsingular for all $x \in D$. Assume the special affine contravariant Lipschitz condition*

$$\|\big(F'(y) - F'(x)\big)(y-x)\| \le \omega \|F'(x)(y-x)\|^2 \text{ for } x, y \in D\,.$$

Then, with the convenient notation

$$h_k := \omega \|F(x^k)\|\,,$$

and $\lambda \in [0, \min(1, 2/h_k)]$ we have:

$$\|F(x^k + \lambda \Delta x^k)\|_2 \le t_k(\lambda) \|F(x^k)\|_2\,,$$

where

$$t_k(\lambda) := 1 - \lambda + \tfrac{1}{2}\lambda^2 h_k\,. \tag{3.26}$$

The optimal choice of damping factor in terms of this local estimate is

$$\overline{\lambda}_k := \min(1\,,\ 1/h_k)\ . \tag{3.27}$$

Proof. Dropping the superscript index k we may derive

$$\begin{aligned}
\|F(x+\lambda\Delta x)\| &= \|F(x+\lambda\Delta x) - F(x) - F'(x)\Delta x\| \\
&= \Big\| \int_{s=0}^{\lambda} \big(F'(x + s\cdot\Delta x) - F'(x)\big)\Delta x\, ds - (1-\lambda)F'(x)\Delta x \ \Big\| \\
&\le (1-\lambda)\|F(x)\| + O(\lambda^2) \quad \text{for } \lambda \in [0,1]\,.
\end{aligned}$$

The arising $O(\lambda^2)$-term obviously characterizes the deviation from the Newton path and can be estimated as:

$$\Big\| \int_{s=0}^{\lambda} \big(F'(x+s\cdot\Delta x)-F'(x)\big)\Delta x ds \Big\| \le \omega\cdot\tfrac{1}{2}\lambda^2\|F'(x)\Delta x\|^2 = \tfrac{1}{2}\lambda^2 h_k\cdot\|F(x)\|\,.$$

Minimization of the above defined parabola t_k then directly yields $\overline{\lambda}_k$ with the a-priori restriction to the unit interval due to the underlying Newton path concept. □

We are now ready to derive a global convergence theorem on the basis of this local descent result.

Theorem 3.8 *Notation and assumptions as in the preceding Theorem* 3.7. *In addition, let* D_0 *denote the path-connected component of* $G(x^0)$ *in* x^0 *and assume that* $D_0 \subseteq D$ *is compact. Let the Jacobian* $F'(x)$ *be nonsingular for all* $x \in D_0$. *Then the damped Newton iteration* $(k = 0, 1, \ldots)$ *with damping factors in the range*

$$\lambda_k \in \big[\varepsilon\,,\ 2\overline{\lambda}_k - \varepsilon\big]$$

and sufficiently small $\varepsilon > 0$, *which depends on* D_0, *converges to some solution point* x^*.

Proof. The proof is by induction using the local results of the preceding theorem. In Figure 3.5, the estimation parabola t_k defined in (3.26) is depicted as a function of the damping factor λ together with the polygonal upper bound

$$t_k(\lambda) \le \begin{cases} 1-\frac{1}{2}\lambda & ,\quad 0 \le \lambda \le \dfrac{1}{h_k}\,, \\[2ex] 1+\frac{1}{2}\lambda - \dfrac{1}{h_k} & ,\quad \dfrac{1}{h_k} \le \lambda \le \dfrac{2}{h_k}\,. \end{cases}$$

Upon restricting λ to the range indicated in the present theorem, we immediately have

$$t_k(\lambda) \le 1-\tfrac{1}{2}\varepsilon\,,\ 0 < \varepsilon \le \frac{1}{h_k}\,, \tag{3.28}$$

which induces *strict* reduction of the residual level function $T(x)$ in each iteration step k. In view of a proof of *global* convergence, a question left to discuss is whether there exists some global $\varepsilon > 0$. This follows from the fact that

$$\max_{x\in D_0}\|F(x)\| < \infty$$

under the *compactness* assumption on D_0. Hence, whenever $G(x^k) \subseteq D_0$, then (3.28) assures that

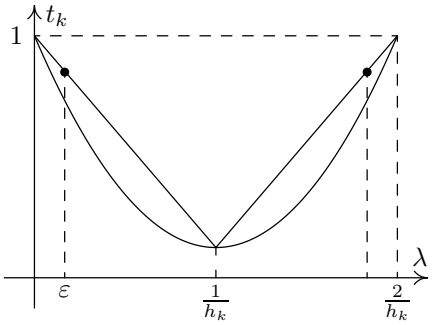

Fig. 3.5. Local reduction parabola t_k together with polygonal upper bounds.

$$G\big(x^{k+1}(\lambda)\big) \subset G(x^k) \subseteq D_0\,.$$

With arguments similar as in the proof of Theorem 2.12, we finally conclude by induction that the defined damped Newton iteration converges towards some limit point x^* with $F(x^*) = 0$, which completes the proof. □

3.2.2 Adaptive trust region strategies

The above derived theoretical damping strategy (3.27) cannot be implemented directly, since the arising Kantorovich quantities h_k are computationally unavailable due to the arising Lipschitz constant ω. The obtained theoretical results can nevertheless be exploited for the construction of computational strategies. Following the paradigm of Section 1.2.3, we may determine damping factors in the course of the iteration *as close to the convergence analysis as possible* replacing the unavailable Lipschitz constants ω by computational estimates $[\omega]$ and the unavailable Kantorovich quantities $h_k = \omega\|F(x^k)\|$ by computational estimates $[h_k] = [\omega]\|F(x^k)\|$. Such estimates can only be obtained by *pointwise sampling* of the domain dependent Lipschitz constants, which immediately implies that

$$[\omega] \leq \omega\,, [h_k] \leq h_k\,. \tag{3.29}$$

By definition, the estimates $[\cdot]$ will inherit the *affine contravariant* structure. As soon as we have iterative estimates $[h_k]$ at hand, associated estimates of the optimal damping factors may be naturally defined:

$$\big[\overline{\lambda}_k\big] := \min\left(1, 1/[h_k]\right)\,. \tag{3.30}$$

The relation (3.29) induces the equivalent relation

$$[\overline{\lambda}_k] \geq \overline{\lambda}_k\,.$$

This means that the estimated damping factors might be 'too large'—obviously *an unavoidable gap between analysis and algorithm.* As a consequence, repeated reductions might be necessary, which implies that any damping strategy to be derived will have to split into a *prediction strategy* and a *correction strategy.*

Bit counting lemma. As for the *required accuracy* of the computational estimates, the following lemma is important.

Lemma 3.9 *Notation as just introduced. Assume that the damped Newton method with damping factors as defined in* (3.30) *is realized. As for the accuracy of the computational estimates let*

$$0 \le h_k - [h_k] < \sigma \max\bigl(1, [h_k]\bigr) \text{ for some } \sigma < 1\,. \tag{3.31}$$

Then the residual monotonicity test will yield

$$\|F(x^{k+1})\| \le \bigl(1 - \tfrac{1}{2}(1-\sigma)\lambda\bigr)\,\|F(x^k)\|\,.$$

Proof. We reformulate the relation (3.31) as

$$[h_k] \le h_k < (1+\sigma)\max\bigl(1, [h_k]\bigr)\,.$$

Then the above notation directly leads to the estimation

$$\begin{aligned}\frac{\|F(x^{k+1})\|}{\|F(x^k)\|} &\le \bigl[1 - \lambda + \tfrac{1}{2}\lambda^2 h_k\bigr]_{\lambda=[\overline{\lambda}_k]} \\ &< \bigl[1 - \lambda + \tfrac{1}{2}(1+\sigma)\lambda^2 [h_k]\bigr]_{\lambda=[\overline{\lambda}_k]} \le 1 - \tfrac{1}{2}(1-\sigma)\overline{\lambda}_k\,.\end{aligned}$$

□

For $\sigma < 1$, any computational estimates $[h_k]$ are just required to catch the *leading binary digit* of h_k, in order to assure residual monotonicity. For $\sigma \le \frac{1}{2}$, we arrive at the *restricted residual monotonicity test*

$$\|F(x^{k+1})\| \le \bigl(1 - \tfrac{1}{4}\lambda\bigr)\,\|F(x^k)\|. \tag{3.32}$$

This test nicely compares with the *Armijo strategy* (3.19), though derived by a different argument.

Computational estimates. After these preliminary considerations, we now proceed to identify *affine contravariant* computational estimates $[\cdot]$—preferably those, which are cheap to evaluate in the course of the damped Newton iteration. In order to derive such estimates, we first recall from Section 3.1.4 that the damped Newton method may be interpreted as a deviation

from the associated Newton path. Measuring the deviation in an affine contravariant setting leads us to the bound

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

which, in turn, leads to the following lower bound for the affine contravariant Kantorovich quantity:

$$[h_k] := \frac{2\|F(x^{k+1}) - (1-\lambda)F(x^k)\|}{\lambda^2\|F(x^k)\|} \le h_k\,.$$

This estimate requires at least one trial value $x^{k+1} = x^k + \lambda_k^0 \Delta x^k$ so that it can only be exploited for the design of a *correction strategy* of the kind $(i = 0, 1, \ldots)$:

$$\lambda_k^{i+1} := \min\bigl(\tfrac{1}{2}\lambda_k^i, 1/[h_k^{i+1}]\bigr)\,. \tag{3.33}$$

In order to construct a theoretically backed initial estimate λ_k^0, we may apply the relation

$$h_{k+1} = \frac{\|F(x^{k+1})\|}{\|F(x^k)\|} h_k,$$

which directly inspires estimates of the kind

$$[h_{k+1}^0] = \frac{\|F(x^{k+1})\|}{\|F(x^k)\|}[h_k^{i_*}] < [h_k^{i_*}]\,,$$

wherein i_* indicates the final computable index within estimate (3.33) for the previous iterative step k. Thus we are led to the following *prediction strategy* for $k \ge 0$:

$$\lambda_{k+1}^0 := \min\bigl(1, 1/[h_{k+1}^0]\bigr)\,.$$

As can be seen, the only empirical choice left to be made is the starting value λ_0^0. It is recommended to set $\lambda_0^0 = 1$ for 'mildly nonlinear' problems and $\lambda_0^0 = \lambda_{\min} \ll 1$ for 'highly nonlinear' problems in a definition to be put in the hands of the users.

Intermediate quasi-Newton steps. Whenever $\lambda_k = 1$ and the residual monotonicity test yields

$$\Theta_k = \frac{\|F(x^{k+1})\|}{\|F(x^k)\|} \le \Theta_{\max} < 1$$

for some default value $\Theta_{\max}$, then the residual based quasi-Newton method of Section 2.2.3 may be applied—compare Theorem 2.14. This means that Jacobian evaluations are replaced by *residual rank-1 updates*. As for a possible switch back from quasi-Newton steps to Newton steps just look into the details of the informal quasi-Newton algorithm `QNRES`, also in Section 2.2.3.

The just described adaptive trust region strategy is realized in

Algorithm `NLEQ-RES`**.** Set a required *residual accuracy* ε sufficiently above the machine precision.
Guess an initial iterate x^0. Evaluate $F(x^0)$.
Set an initial damping factor either $\lambda_0 := 1$ or $\lambda_0 \ll 1$.
Norms are tacitly understood to be scaled smooth norms, such as $\|\bar{D}^{-1} \cdot \|_2$, where $\bar{D}$ is a diagonal scaling matrix, constant throughout the iteration.
For iteration index $k = 0, 1, \ldots$ do:

1. **Step** k:
   Convergence test: If $\|F(x^k)\| \leq \varepsilon$: **stop**. Solution found $x^* := x^k$.
   Else: Evaluate Jacobian matrix $F'(x^k)$. Solve linear system

$$F'(x^k)\Delta x^k = -F(x^k)$$

   For $k > 0$: compute a prediction value for the damping factor

$$\lambda_k := \min(1, \mu_k), \quad \mu_k := \frac{\|F(x^{k-1})\|}{\|F(x^k)\|}\mu'_{k-1}\,.$$

   Regularity test: If $\lambda_k < \lambda_{\min}$: **stop**. Convergence failure.
2. **Else**: compute the trial iterate $x^{k+1} := x^k + \lambda_k \Delta x^k$ and evaluate $F(x^{k+1})$.
3. Compute the monitoring quantities

$$\Theta_k := \frac{\|F(x^{k+1})\|}{\|F(x^k)\|}\,, \qquad \mu'_k := \frac{\frac{1}{2}\|F(x^k)\| \cdot \lambda_k^2}{\|F(x^{k+1}) - (1-\lambda_k)F(x^k)\|}$$

   If $\Theta_k \geq 1$ (or, if **restricted**: $\Theta_k > 1 - \lambda_k/4$):

   then replace λ_k by $\lambda'_k := \min(\mu'_k, \frac{1}{2}\lambda_k)$. **Go to** Regularity test.

   Else: let $\lambda'_k := \min(1, \mu'_k)$.

   If $\lambda'_k = \lambda_k = 1$ **and** $\Theta_k < \Theta_{\max}$: switch to `QNRES`.

   Else: **If** $\lambda'_k \geq 4\lambda_k$: replace λ_k by λ'_k and **goto** 2.

   Else: accept x^{k+1} as new iterate. **Goto** 1 with $k \to k+1$.

3.2.3 Inexact Newton-RES method

In this section we discuss the *inexact* global Newton method

$$x^{k+1} = x^k + \lambda_k \delta x^k\,,\ 0 < \lambda_k \leq 1$$

realized by means of `GMRES` such that (dropping the inner iteration index i)

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

Let $\delta x_0^k = 0$ and thus $r_0^k = F(x^k)$. The notation here follows Section 2.2.4 on local Newton-RES methods.

Convergence analysis. Before going into details of the analysis, we want to point out that the inexact Newton method with damping can be viewed as a tangent step in x^k for the *inexact Newton path*

$$F\left(\widetilde{x}(\lambda)\right) - r^k = (1-\lambda)\left(F(x^k) - r^k\right)$$

or, equivalently,

$$F\left(\widetilde{x}(\lambda)\right) = (1-\lambda)F(x^k) + \lambda r^k\,, \tag{3.34}$$

wherein $\widetilde{x}(0) = x^k$, $\dot{\widetilde{x}}(0) = \delta x^k$, but $\widetilde{x}(1) \neq x^*$. Hence, when approaching x^*, we will have to assure that $r^k \to 0$. With this geometric interpretation in mind, we are now prepared to derive the following convergence statements.

Theorem 3.10 *Under the assumptions of Theorem* 3.7 *for the exact Newton iteration with damping, the inexact Newton-*GMRES *iteration can be shown to satisfy*

$$\Theta_k := \frac{\|F(x^{k+1})\|}{\|F(x^k)\|} \leq t_k(\lambda_k, \eta_k) \tag{3.35}$$

with

$$t_k(\lambda,\eta) = 1 - (1-\eta)\lambda + \tfrac{1}{2}\lambda^2(1-\eta^2)h_k\,,\ \eta_k = \frac{\|r^k\|_2}{\|F(x^k)\|_2} < 1\,.$$

The optimal choice of damping factor is

$$\overline{\lambda}_k := \min\left(1, \frac{1}{(1+\eta_k)h_k}\right)\,. \tag{3.36}$$

Proof. Recall from Section 2.1.5 that for GMRES

$$\|F(x^k) - r^k\|_2^2 = \|F(x^k)\|_2^2 - \|r^k\|_2^2 = (1-\eta_k^2)\|F(x^k)\|_2^2$$

for well-defined $\eta_k < 1$. Along the line $x^k + \lambda\delta x^k$ the descent behavior can be estimated using

$$F(x^k + \lambda\delta x^k) = (1-\lambda)F(x^k) + \lambda r^k + \lambda \int_{s=0}^{1} \left(F'(x^k + s\lambda\delta x^k) - F'(x^k)\right)\delta x^k ds\,.$$

The last right hand term is directly comparable to the exact case: in the application of the affine contravariant Lipschitz condition we merely have to replace Δx^k by δx^k and, accordingly,

$$\|F'(x^k)\Delta x^k\|^2 = \|F(x^k)\|^2$$

by

$$\|F'(x^k)\delta x^k\|^2 = (1-\eta_k^2)\|F(x^k)\|^2\,.$$

With this modification, the result (3.35) is readily verified. The optimal damping factor follows from setting $t_k'(\lambda) = 0$. With $\lambda_k \leq 1$ as restriction, we have (3.36). □

Adaptive trust region method. In order to exploit the above convergence analysis for the construction of an inexact Newton-GMRES algorithm, we will follow the usual *paradigm* and certainly aim at defining certain damping factors

$$\big[\overline{\lambda}_k\big] := \min\left(1, \frac{1}{(1+\eta_k)[h_k]}\right)$$

in terms of affine contravariant computationally available estimates. First, if we once again apply the 'bit counting' Lemma 3.9, we arrive at the *inexact variant* of the *restricted residual monotonicity test*

$$\|F(x^{k+1})\|_2 \le \left(1 - \frac{1-\eta_k}{4}\lambda_k\right)\|F(x^k)\|_2\,,$$

which here replaces (3.32). Next, upon returning to the above proof of Theorem 3.10, we readily observe that

$$\|F(x^{k+1}) - (1-\lambda_k)F(x^k) - \lambda_k r^k\|_2 \le \frac{\lambda_k^2}{2}(1-\eta_k^2)h_k\|F(x^k)\|_2\,.$$

On this basis, we may simply define the *a-posteriori* estimates

$$[h_k](\lambda) := \frac{2\|F\left(x^{k+1}(\lambda)\right) - (1-\lambda)F(x^k) - \lambda r^k\|_2}{\lambda^2(1-\eta_k^2)\|F(x^k)\|_2} \le h_k\,,$$

which give rise to the *correction strategy* ($i = 0, 1, \ldots, i_k^*$)

$$\lambda_k^{i+1} = \min\left(\tfrac{1}{2}\lambda_k^i\,,\ \frac{1}{(1+\eta_k)[h_k^i]}\right)\,,$$

and the associated *a-priori* estimates

$$[h_{k+1}^0] := \Theta_k[h_k^{i_*}] \le h_{k+1}\,,$$

which induce the *prediction strategy* ($k = 0, 1, \ldots$)

$$\lambda_{k+1}^0 := \min\left(1, \frac{1}{(1+\eta_{k+1})[h_{k+1}^0]}\right)\,.$$

As for the choice of η_k, we already have a strategy for $\lambda_k = \lambda_{k-1} = 1$—see Section 2.1.5. For $\lambda_k < 1$, some constant value $\eta_k \le \eta$ with some sufficiently small threshold value η can be selected (and handed over to the local Newton method, see Section 2.2.4). Then only λ_0^0 remains to be set externally.

The just described residual based adaptive trust region strategy in combination with the strategy to match inner and outer iteration is realized in the code GIANT-GMRES.

Bibliographical Note. Residual based inexact Newton methods date back to R.S. Dembo, S.C. Eisenstat, and T. Steihaug [51]. Quite popular algorithmic heuristics have been worked out by R.E. Bank and D.J. Rose in [19] and are applied in a number of published algorithms. A different global convergence analysis has been given in [90, 91] by S.C. Eisenstat and H.F. Walker. Their strategies are implemented in the code `NITSOL` due to M. Pernice and H.F. Walker [166]. They differ from the ones presented here.

3.3 Error Oriented Descent

In this section we study the *damped Newton iteration*

$$F'(x^k)\Delta x^k = -F(x^k), \qquad x^{k+1} = x^k + \lambda_k \Delta x^k, \qquad \lambda_k \in]0,1]$$

in an error oriented framework, which aims at overcoming certain difficulties that are known to arise in the residual based framework, especially in situations where the Jacobian matrices are ill-conditioned—such as in discretized nonlinear partial differential equations. Once again, we treat the damped Newton method as a deviation from the Newton path, but this time we characterize the deviation by means of *affine covariant* Lipschitz conditions such as those used in the convergence theory for error oriented *local* Newton methods in Section 2.1.

The construction of an error oriented globalization of local Newton methods is slightly more complicated than in the residual based approach. For this reason, we first recur to the concept of *general level functions* $T(x|A)$ as already introduced in (3.15) for *arbitrary* nonsingular matrix A and study the descent behavior of the damped Newton method for the whole class of such functions in an affine covariant theoretical framework (Section 3.3.1). As it turns out, the obtained theoretically optimal damping factors actually reflect the observed difficulties of the residual based variants. Moreover, the analysis directly leads to the specific choice $A = F'(x^k)^{-1}$, which defines the so-called *natural* level function (Section 3.3.2). As a consequence for actual computation, the iterates are required to satisfy the so-called *natural monotonicity test*

$$\|\overline{\Delta x}^{k+1}\| < \|\Delta x^k\|,$$

wherein the *simplified* Newton correction $\overline{\Delta x}^{k+1}$ defined by

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

is only computed to evaluate this test (and later also for an adaptive trust region method). As for a proof of global convergence, only a theorem covering a slightly different situation is available up to now—despite the convincing global convergence properties of the thus derived algorithm! From the associated theoretically optimal damping factors we develop computational trust

region strategies—first for the *exact* damped Newton method as defined above (Section 3.3.3) and second for *inexact* variants using error reducing linear iterative solvers for the inner iteration (Section 3.3.4).

3.3.1 General level functions

In order to derive an affine covariant or error oriented variant of the damped Newton method we first recur to general level functions, which have been already defined in (3.15) as

$$T(x|A) := \tfrac{1}{2}\|AF(x)\|_2^2 \,.$$

Local descent. It is an easy task to verify that the Newton direction points 'downhill' with respect to *all* such level functions.

Lemma 3.11 *Let $F \in C^1(D)$ and let Δx denote the ordinary Newton correction (dropping the iteration index k). Then, for all $A \in \mathrm{GL}(n)$,*

$$\Delta x^T \operatorname{grad} T(x|A) = -2T(x|A) < 0 \,.$$

This is certainly a distinguishing feature to any other descent directions—compare Lemma 3.3. Hence, on the basis of *first* order information only, all monotonicity criteria look equally well-suited for the damped Newton method. The selection of a specific level function will therefore require *second* order information.

Theorem 3.12 *Let $F \in C^1(D)$ with $D \subset \mathbb{R}^n$ convex and $F'(x) = F'(x)$ nonsingular for all $x \in D$. For a given current iterate $x^k \in D$ let $G(x^k|A) \subset D$ for some $A \in \mathrm{GL}(n)$. For $x, y \in D$ assume that*

$$\|F'(x)^{-1}\big(F'(y) - F'(x)\big)(y - x)\| \le \omega \|y - x\|^2 \,.$$

Then, with the convenient notation

$$h_k := \|\Delta x^k\|\omega \,,\ \overline{h}_k := h_k \operatorname{cond}\big(AF'(x^k)\big)$$

one obtains for $\lambda \in \big[0, \min\big(1, 2/\overline{h}_k\big)\big]$:

$$\|AF(x^k + \lambda \Delta x^k)\| \le t_k(\lambda|A)\|AF(x^k)\| \,, \tag{3.37}$$

where

$$t_k(\lambda|A) := 1 - \lambda + \tfrac{1}{2}\lambda^2 \overline{h}_k \,. \tag{3.38}$$

The optimal choice of damping factor in terms of this local estimate is

$$\overline{\lambda}_k(A) := \min\big(1\,,\ 1/\overline{h}_k\big) \,.$$

Proof. Dropping the superscript index k we may derive

$$\begin{aligned}
&\|AF(x+\lambda\Delta x)\| = \|A\big(F(x+\lambda\Delta x) - F(x) - F'(x)\Delta x\big)\| \\
&= \Big\|A\left(\int_{\delta=0}^{\lambda}\big(F'(x+\delta\Delta x) - F'(x)\big)\Delta x d\delta - (1-\lambda)F'(x)\Delta x\right)\Big\| \\
&\le (1-\lambda)\|AF(x)\| + O(\lambda^2) \quad \text{for } \lambda\in[0,1]\,.
\end{aligned}$$

The arising $O(\lambda^2)$-term obviously characterizes the deviation from the Newton path and can be estimated as:

$$\begin{aligned}
&\Big\|\, AF'(x)\int_{\delta=0}^{\lambda} F'(x)^{-1}\big(F'(x+\delta\Delta x) - F'(x)\big)\Delta x d\delta\,\Big\| \\
&\le \|AF'(x)\|\omega\tfrac{1}{2}\lambda^2\|\Delta x\|^2 = \tfrac{1}{2}\lambda^2\|AF'(x)\|h_k\|\big(AF'(x)\big)^{-1}AF(x)\| \\
&\le \tfrac{1}{2}\lambda^2 h_k\|AF'(x)\|\|\big(AF'(x)\big)^{-1}\|\|AF(x)\| = \tfrac{1}{2}\lambda^2\overline{h}_k\|AF(x)\|\,.
\end{aligned}$$

Minimization of the parabola t_k then directly yields $\overline{\lambda}_k(A)$ with the a-priori restriction to the unit interval due to the underlying Newton path concept. □

Global convergence. The above *local descent* result may now serve as a basis for the following global convergence theorem.

Theorem 3.13 *Notation and assumptions as in the preceding Theorem* 3.12. *In addition, let D_0 denote the path-connected component of $G(x^0|A)$ in x^0 and assume that $D_0 \subseteq D$ is compact. Let the Jacobian $F'(x)$ be nonsingular for all $x \in D_0$. Then the damped Newton iteration $(k = 0, 1, \ldots)$ with damping factors in the range*

$$\lambda_k \in \big[\varepsilon\,,\ 2\overline{\lambda}_k(A) - \varepsilon\big]$$

and sufficiently small $\varepsilon > 0$, which depends on D_0, converges to some solution point x^*.

Proof. The proof is by induction using the local results of the preceding theorem. Moreover, it is just a slight modification of the proof of Theorem 3.8 for the residual level function. In particular, Figure 3.5 shows the same type of estimation parabola t_k as defined here in (3.38): once again, the proper polygonal upper bound supplies the global upper bound

$$t_k(\lambda|A) \le 1 - \tfrac{1}{2}\varepsilon\,,\ 0 < \varepsilon \le \frac{1}{\overline{h}_k}\,, \tag{3.39}$$

which induces *strict* reduction of the general level function $T(x|A)$ in each iteration step k. We are now just left to discuss whether there exists some global $\varepsilon > 0$. This follows from the fact that

$$\max_{x \in D_0} \|F'(x)^{-1} F(x)\| \cdot \mathrm{cond}_2\big(AF'(x)\big) < \infty$$

under the *compactness* assumption on D_0. Hence, whenever $G(x^k|A) \subseteq D_0$, then (3.39) assures that

$$G\big(x^{k+1}(\lambda)|A\big) \subset G(x^k|A) \subseteq D_0\,.$$

We therefore conclude by induction that

$$\lim_{k \to \infty} x^k = x^*\,,$$

which completes the proof. □

Algorithmic limitation of residual monotonicity. The above theorem offers an intriguing explanation, why the damped Newton method endowed with the traditional *residual monotonicity criterion*

$$T(x^{k+1}|I) \leq T(x^k|I)$$

may fail in practical computation despite its proven global convergence property (compare Theorem 3.8): in fact, whenever the *Jacobian* is *ill-conditioned*, then the 'optimal' damping factors are bound to satisfy

$$\overline{\lambda}_k(I) = \Big(h_k \mathrm{cond}_2\big(F'(x^k)\big)\Big)^{-1} < \lambda_{\min} \ll 1\,. \tag{3.40}$$

Therefore, in worst cases, also any computational damping strategies (no matter, how sophisticated they might be) will lead to a *practical termination of the iteration*, since then $x^{k+1} \approx x^k$, which means that the iteration 'stands still'. For illustration of this effect see Example 3.1 below, especially Fig. 3.9.

Another side of the same medal is the quite often reported observation that for 'well-chosen' initial guesses x^0 residual monotonicity may be violated over several initial iterative steps even though the ordinary Newton iteration converges when allowed to do so by skipping the residual monotonicity test. In fact, from the error oriented local convergence analysis of Section 2.1, one would expect to obtain the optimal value $\overline{\lambda}_k = 1$ roughly as soon as the iterates are contained in the 'neighborhood' of the solution x^*—say, as soon as for some iterate $h_k < 1$. A comparison with the above theorem, however, shows that a condition of the kind $\overline{h}_k = h_k \mathrm{cond}_2\big(F'(x^k)\big) < 1$ would be required in the residual framework. The effect is illustrated by Example 3.1 at the end of Section 3.3.2.

Summarizing, we have the following situation:

Combining any damping strategy with the residual monotonicity criterion may have the consequence that mildly nonlinear problems actually 'look like' highly nonlinear problems, especially in the situation (3.40)*; as a consequence, especially in the presence of ill-conditioned Jacobians, the Newton iteration with damping tends to terminate without the desired result—despite an underlying global convergence theorem like Theorem* 3.13 *for* $A = I$.

3.3.2 Natural level function

The preceding section seemed to indicate that 'all level functions are equal'; here we want to point out that 'some animals are more equal than others' (compare George Orwell, *Animal Farm*).

As has been shown, the condition number $\mathrm{cond}_2\big(AF'(x^k)\big)$ plays a central role in the preceding analysis, at least in the worst case situation. Therefore, due to the well-known property

$$\mathrm{cond}_2\big(AF'(x^k)\big) \geq 1 = \mathrm{cond}_2\big(I\big)\,,$$

the special choice

$$A_k := F'(x^k)^{-1}$$

seems to be *locally optimal* as a specification of the matrix in the general level function. The associated level function $T\big(x|F'(x^k)^{-1}\big)$ is called *natural level function* and the associated *natural monotonicity test* requires that

$$\|\overline{\Delta x}^{k+1}\|_2 \leq \|\Delta x^k\|_2 \tag{3.41}$$

in terms of the *ordinary* Newton correction Δx^k and the *simplified* Newton correction defined by

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

This specification gives rise to several outstanding properties.

Extremal properties. For $A \in \mathrm{GL}(n)$ the reduction factors $t_k(\lambda|A)$ and the theoretical optimal damping factors $\overline{\lambda}_k(A)$ satisfy:

$$\begin{aligned} t_k(\lambda|A_k) &= 1 - \lambda + \tfrac{1}{2}\lambda^2 h_k \;\leq\; t_k(\lambda|A) \\ \overline{\lambda}_k(A_k) &= \min\left(1\,,\; 1/h_k\right) \geq \overline{\lambda}_k(A)\,. \end{aligned}$$

An associated graphical representation is given in Figure 3.6.

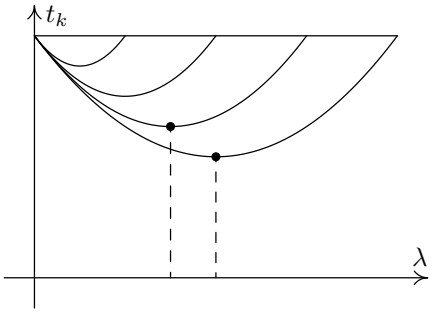

Fig. 3.6. Extremal properties of natural level function: reduction factors $t_k(\lambda|A)$ and optimal damping factors $\overline{\lambda}_k(A)$.

Steepest descent property. The steepest descent direction for $T(x|A)$ in x^k is

$$-\operatorname{grad} T(x^k|A) = -\big(AF'(x^k)\big)^T AF(x^k)\,.$$

With the specification $A = A_k$ this leads to

$$\Delta x^k = -\operatorname{grad} T(x^k|A_k)\,,$$

which means that *the damped Newton method in x^k is a method of steepest descent for the natural level function $T(x|A_k)$.*

Merging property. The locally optimal damping factors nicely reflect the expected behavior in the contraction domain of the ordinary Newton method: in fact, we have

$$h_k \le 1 \implies \overline{\lambda}_k(A_k) = 1\,.$$

Hence, quadratic convergence is asymptotically achieved by the damping strategy based on $\overline{\lambda}_k$.

Asymptotic distance function. For $F \in C^2(D)$, we easily verify that

$$T\big(x|F'(x^*)^{-1}\big) = \tfrac{1}{2}\|x - x^*\|_2^2 + O(\|x - x^*\|^3)\,.$$

Hence, for $x^k \to x^*$, the natural monotonicity criterion asymptotically merges into a desirable distance criterion of the form

$$\|x^{k+1} - x^*\|_2 \mathrel{\dot{\le}} \|x^k - x^*\|_2\,,$$

which is exact for *linear* problems. The situation is represented graphically in Figure 3.7. Far away from the solution point, this nice geometrical property survives in the form that the osculating ellipsoid to the level surface at the current iterate turns out to be an *osculating sphere*.

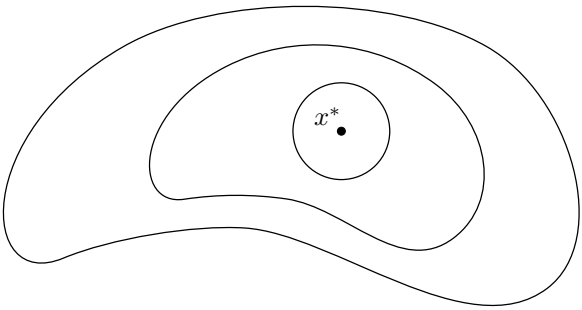

Fig. 3.7. Natural level sets: asymptotic distance spheres.

In the linear case, the Jacobian condition number represents the quotient of the largest over the smallest half-axis of the level ellipsoid. In the nonlinear case, too, Jacobian ill-conditioning gives rise to cigar-shaped residual level sets, which, in general, are distorted ellipsoids. Therefore, geometrically speaking, the natural level function realizes some *nonlinear preconditioning*.

Local descent. Any damping strategy based on the natural monotonicity test is sufficiently characterized by Theorem 3.12: just insert $A = A_k$ into (3.37) and (3.38), which then yields

$$\|\overline{\Delta x}^{k+1}\| \leq \left(1 - \lambda + \tfrac{1}{2}\lambda^2 h_k\right) \|\Delta x^k\| \,.$$

Global convergence. In the present situation, the above global convergence theorem for general level functions, Theorem 3.13, does *not* apply, since the choice A_k varies from step to step. In order to obtain an *affine covariant* global convergence theorem, the locally optimal choice $A = F'(x^k)^{-1}$ will now be *modelled* by the *fixed* choice $A = F'(x^*)^{-1}$—in view of the asymptotic distance function property.

Theorem 3.14 *Let $F : D \longrightarrow \mathbb{R}^n$ be a continuously differentiable mapping with $D \subseteq \mathbb{R}^n$ open convex. Assume that $x^0, x^* \in D$ with x^* unique solution of F in D and the Jacobian $F'(x^*)$ nonsingular. Furthermore, assume that*

(I) *$F'(x)$ is nonsingular for all $x \in D$,*

(II) *the path-connected component D_0 of $G\big(x^0|F'(x^*)^{-1}\big)$ in x^0 is compact and contained in D,*

(III) *the following affine covariant Lipschitz condition holds*

$$\left\| F'(x^*)^{-1}\big(F'(y) - F'(x)\big)(y - x) \right\| \leq \omega_* \|y - x\|^2 \text{ for } y, x \in D \,,$$

(IV) *for any iterate $x^k \in D$ let $h_k^* := \omega_* \|\Delta x^k\| \cdot \|F'(x^k)^{-1} F'(x^*)\| < \infty$.*

As the locally optimal damping strategy we obtain

$$\lambda_k^* := \min\left(1\,,\ 1/h_k^*\right)\,.$$

Then any damped Newton iteration with iterative damping factors in the range

$$\lambda_k \in \left[\varepsilon\,,\ 2\lambda_k^* - \varepsilon\right)\ \ \textit{for}\ 0 < \varepsilon < 1/h_k^*$$

converges globally to x^*.

Proof. In the proof of Theorem 3.12 we had for $x = x^k$:

$$\|AF(x + \lambda\Delta x)\| \le (1-\lambda)\|AF(x)\| + O(\lambda^2)\,.$$

For $A = F'(x^*)^{-1}$ the $O(\lambda^2)$-term may now be treated differently as follows:

$$\begin{aligned}
&\Big\|\ F'(x^*)^{-1} \int_{\delta=0}^{\lambda} \big(F'(x+\delta\Delta x) - F'(x)\big)\Delta x d\delta\ \Big\| \le \int_{\delta=0}^{\lambda} \omega_* \cdot \delta \|\Delta x\|^2 d\delta \\
&\le \omega_* \cdot \tfrac{1}{2}\lambda^2 \|\Delta x\| \cdot \|F'(x)^{-1}F'(x^*)\| \cdot \|F'(x^*)^{-1}F(x)\| \\
&= \tfrac{1}{2}\lambda^2 h_k^* \|F'(x^*)^{-1}F(x)\|\,.
\end{aligned}$$

On the basis of the thus modified local reduction property, global convergence in terms of the above specified level function can be shown along the same lines of argumentation as in Theorem 3.13. The above statements are just the proper copies of the statements of that theorem. □

Corollary 3.15 *Under the assumptions of the preceding theorem with the replacement of* x^* *by an arbitrary* $z \in D_0$ *in the Jacobian inverse and the associated affine covariant Lipschitz condition*

$$\big\|\ F'(z)^{-1}\big(F'(y) - F'(x)\big)(y-x)\ \big\| \le \omega(z)\ \|y-x\|^2 \textit{for}\ x, y, z \in D_0\,,$$

a local level function reduction of the form

$$T\big(x^k + \lambda\Delta x^k | F'(z)^{-1}\big) \le \big(1 - \lambda + \tfrac{1}{2}\lambda^2 h_k(z)\big)^2 T\big(x^k | F'(z)^{-1}\big)$$

in terms of

$$h_k(z) := \|\Delta x^k\|\ \omega(z)\ \|F'(x^k)^{-1}F'(z)\|$$

and a locally optimal damping factor

$$\overline{\lambda}_k(z) := \min\left(1, 1/h_k(z)\right)$$

can be shown to hold. Assuming further that the used matrix norm is submultiplicative, then we obtain for best possible estimates $\omega(z)$ *the extremal properties*

$$\|F'(x^k)^{-1}F'(z)\|\,\omega(z) \geq \omega(x^k)$$

and

$$h_k(x^k) \leq h_k(z)\,,\ \overline{\lambda}_k \geq \overline{\lambda}_k(z)\,,\ z \in D_0\,.$$

The corollary states that the locally optimal damping factors in terms of the locally defined natural level function are outstanding among all possible globally optimal damping factors in terms of any globally defined affine covariant level function. Our theoretical convergence analysis shows that we may substitute the global affine covariant Lipschitz constant ω by its more local counterpart $\omega_k = \omega(x^k)$ defined via

$$\left\| F'(x^k)^{-1}\big(F'(x) - F'(x^k)\big)(x - x^k) \right\| \leq \omega_k \|x - x^k\|^2 \text{ for } x, x^k \in D_0\,. \quad (3.42)$$

We have thus arrived at the following *theoretically optimal damping strategy* for the *exact* Newton method

$$x^{k+1} = x^k + \overline{\lambda}_k \Delta x^k, \quad \overline{\lambda}_k := \min(1, 1/h_k), \quad h_k = \omega_k \|\Delta x^k\|\,. \quad (3.43)$$

We must state again that this Newton method with damping based on the natural monotonicity test does not have the comfort of an accompanying global convergence theorem. In fact, U.M. Ascher and M.R. Osborne in [10] constructed a simple example, which exhibits a 2–cycle in the Newton method when monitored by the natural level function. Details are left as Exercise 3.3. However, as shown in [33] by H.G. Bock, E.A. Kostina, and J.P. Schlöder, such 2–cycles can be generally avoided, if the theoretical optimal steplength $\overline{\lambda}_k$ is restricted such that $\lambda h_k \leq \eta < 1$. Details are left as Exercise 3.4. This restriction does not avoid m–cycles for $m > 2$—which still makes the derivation of a global convergence theorem *solely based on natural monotonicity* impossible. Numerical experience advises not to implement this kind of restriction—generically it would just increase the number of Newton iterations required.

Geometrical interpretation. This strategy has a nice geometrical interpretation, which is useful for a deeper understanding of the computational strategies to be developed in the sequel. Recalling the derivation in Section 3.3.1, the damped Newton method at some iterate x^k continues along the tangent of the Newton path $\overline{G}(x^k)$ with effective correction length

$$\|x^{k+1} - x^k\| = \overline{\lambda}_k \|\Delta x^k\| \leq \rho_k := 1/\omega_k\,.$$

Obviously, the radius ρ_k characterizes the *local trust region* of the linear Newton model around x^k. The situation is represented schematically in Figure 3.8.

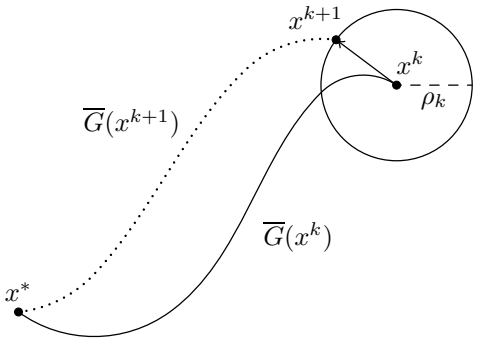

Fig. 3.8. Geometrical interpretation: Newton path $\overline{G}(x^k)$, trust region around x^k, and Newton step with locally optimal damping factor $\overline{\lambda}_k$.

Interpretation via Jacobian information. In terms of a relative change of the Jacobian matrix we may write

$$\left\| F'(x^k)^{-1}\big(F'(x^{k+1}) - F'(x^k)\big)\Delta x^k \right\| \Big/ \|\Delta x^k\| \le \overline{\lambda}_k h_k \le 1\,.$$

This relation suggests the interpretation that Jacobian information at the center x^k of the trust region ball is valid along the Newton direction up to the surface of the ball, which is x^{k+1}. Of course, such an interpretation implicitly assumes that the maximum change actually occurs at the most distant point on the surface—this property certainly holds for the derived upper bounds. Beyond the trust region the Jacobian information from the center x^k is no longer useful, which then implies the construction of a *new* Newton path $\overline{G}(x^{k+1})$ and the subsequent continuation along the new tangent—see once again Figure 3.8.

Behavior near critical points. Finally, we want to discuss the expected behavior of the Newton method with damping in the presence of some close-by *critical point*, say $\widehat{x}$ with *singular Jacobian* $F'(\widehat{x})$. In this situation, *the Newton path and, accordingly, the Newton iteration with optimal damping will be attracted by* $\widehat{x}$. Examples of such a behavior have been observed fairly often—in particular, when multiple solutions are separated by manifolds with singular Jacobian, compare, e.g., Figure 3.10. Nevertheless, even in such a situation, a structural advantage of the natural level function approach may play a role: whereas points $\widehat{x}$ represent *local minima* of $T(x|I)$, which will attract iterative methods based on the residual monotonicity test, they show up as *local maxima* of the natural level functions since $T\big(\widehat{x}|F'(\widehat{x})^{-1}\big)$ is unbounded. For this reason, the above *theoretical* damped Newton method tends to avoid local minima of $T(x)$ whenever they correspond to locally isolated critical points.

Deliberate rank reduction. In rare emergency cases only, a deliberate reduction of the Jacobian rank (the so-called *rank-strategy*) turns out to be helpful—which means the application of *intermediate damped Gauss-Newton steps*. For details, see Section 4.3.5 below.

At the end of Section 3.3.1, we described the limitations of residual monotonicity in connection with Newton's method for systems of equations with ill–conditioned Jacobian. This effect can be neutralized by requiring natural monotonicity instead, as can be seen from the following illustrative example.

Example 3.1 *Optimal orbit plane change of a satellite around Mars.* This optimal control problem has been modeled by the space engineer E.D. Dickmanns [86] at NASA. The obtained ODE boundary value problem has been treated by multiple shooting techniques (see [71, Sect. 8.6.2.] and Section 7.1 below). This led to a system of $n = 72$ nonlinear equations with a Jacobian known to be ill–conditioned. The results given here are taken from the author's dissertation [59, 60]. The problem is a typical representative out of a large class of problems that 'look highly nonlinear', but are indeed essentially 'mildly nonlinear' as discussed at the end of Section 3.3.1.

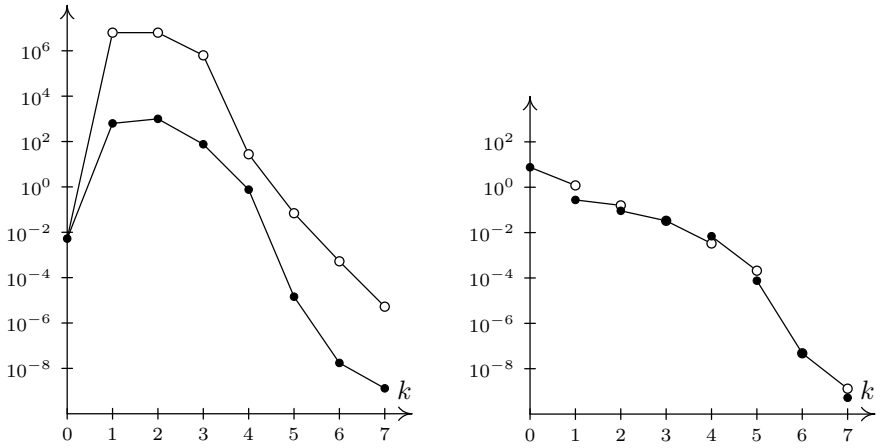

Fig. 3.9. Mars satellite orbit problem. *Left:* no convergence in residual norms $\|F(x^k)\|_2^2$ (∘) or scaled residual norms $\|D_k^{-1}F(x^k)\|_2^2$ (•). *Right:* convergence in natural level function, ordinary Newton corrections $\|D_k^{-1}\Delta x^k\|_2^2$ (•) versus simplified Newton corrections $\|D_k^{-1}\overline{\Delta x}^{k+1}\|_2^2$ (∘).

Figure 3.9 documents the comparative behavior of residual level functions (with and without diagonal scaling) and natural level functions. The Newton iteration has been controlled by scaled natural monotonicity tests

$$\|D_k^{-1}\overline{\Delta x}^{k+1}\|_2^2 \leq \|D_k^{-1}\Delta x^k\|_2^2\,, \quad k = 0, 1, \ldots\,,$$

as shown on the right; in passing, we note that the second Newton step has been performed using 'deliberate rank reduction' as described just above. On the left, the iterative values of the traditional residual level functions, both unscaled and scaled, are seen to increase drastically for the accepted Newton steps. Obviously, only the natural level functions reflect the 'approach' of the iterates x^k toward the solution x^*.

Bibliographical Note. The algorithmic concept of natural level functions has been suggested in 1972 by P. Deuflhard [59] for highly nonlinear problems with ill-conditioned Jacobian. In the same year, *linear preconditioning* has been suggested by O. Axelsson (see [11]) on a comparable geometrical basis, but for the purpose of speeding up the convergence of iterative solvers.

3.3.3 Adaptive trust region strategies

The above derived theoretical damping strategy (3.43) cannot be implemented directly, since the arising Kantorovich quantities h_k contain the computationally unavailable Lipschitz constants ω_k, which are defined over some domain D_0—in view of Figure 3.8, even a definition over some local trust region would be enough. The obtained theoretical results can nevertheless be exploited for the construction of computational strategies. Following our paradigm in Section 1.2.3 again, we determine damping factors in the course of the iteration *as close to the convergence analysis as possible* by introducing *computationally available estimates* $[\omega_k]$ and $[h_k] = [\omega_k]\|\Delta x^k\|$ for the unavailable theoretical quantities ω_k and $h_k = \omega_k\|\Delta x^k\|$.

Such estimates can only be obtained by *pointwise sampling* of the domain dependent Lipschitz constants, which immediately implies that

$$[\omega_k] \le \omega_k \le \omega_k(z) \;, \quad [h_k] \le h_k \le h_k(z) \tag{3.44}$$

compare Corollary 3.15. By definition, the estimates $[\cdot]$ will inherit the *affine covariant* structure. Suppose now that we have certain estimates $[h_k]$ at hand. Then associated estimates of the optimal damping factors may be naturally defined as

$$\left[\overline{\lambda}_k\right] := \min\left(1, 1/[h_k]\right) \;. \tag{3.45}$$

The above relation (3.44) gives rise to the equivalent relation

$$[\overline{\lambda}_k] \ge \overline{\lambda}_k \;.$$

Clearly, any computed estimated damping factors may be 'too large'—which, in turn, means that repeated reductions might be necessary. Therefore, any damping strategy to be derived will have to split into a *prediction strategy* and a *correction strategy*.

Bit counting lemma. The efficiency of such damping strategies will depend on the *required accuracy* of the computational estimates—a question, which is studied in the following lemma.

Lemma 3.16 *Notation as just introduced. Assume that the damped Newton method with damping factors as defined in* (3.45) *is realized. As for the accuracy of the computational estimates, let*

$$O \le h_k - [h_k] < \sigma \max\bigl(1, [h_k]\bigr) \text{ for some } \sigma < 1. \tag{3.46}$$

Then the natural monotonicity test will yield

$$\|\overline{\Delta x}^{k+1}\| \le \bigl(1 - \tfrac{1}{2}(1-\sigma)\lambda\bigr)\, \|\Delta x^k\|\,.$$

Proof. We reformulate the relation (3.46) as

$$[h_k] \le h_k < (1+\sigma) \max\bigl(1, [h_k]\bigr)\,.$$

For $[h_k] \ge 1$, the above notation directly leads to the estimation

$$\begin{aligned} \frac{\|\overline{\Delta x}^{k+1}\|}{\|\Delta x^k\|} &\le \bigl[1 - \lambda + \tfrac{1}{2}\lambda^2 h_k\bigr]_{\lambda=[\overline{\lambda}_k]} \\ &< \bigl[1 - \lambda + \tfrac{1}{2}(1+\sigma)\lambda^2 [h_k]\bigr]_{\lambda=[\overline{\lambda}_k]} \le 1 - \tfrac{1}{2}(1-\sigma)\overline{\lambda}_k\,. \end{aligned}$$

The case $[h_k] < 1$ follows similarly. □

The above lemma states that, for $\sigma < 1$, the computational estimates $[h_k]$ are just required to catch the *leading binary digit* of h_k, in order to assure natural monotonicity. For $\sigma \le \frac{1}{2}$, we arrive at the following *restricted natural monotonicity test*

$$\|\overline{\Delta x}^{k+1}\|_2 \le \bigl(1 - \tfrac{1}{4}\lambda\bigr)\, \|\Delta x^k\|_2, \tag{3.47}$$

which might be useful in actual computation to control the whole iterative process more closely—compare also the residual based restricted monotonicity test (3.32) and the Armijo strategy (3.19).

Correction strategy. After these abstract considerations, we now proceed to derive specific affine covariant computational estimates $[\cdot]$—preferably those, which are cheap to evaluate in the course of the damped Newton iteration. For this purpose, we first recall the interpretation of the damped Newton method as the tangent continuation along the Newton path as given in Section 3.1.4. Upon measuring the deviation in an affine covariant setting, we are led to the upper bound

$$\|\overline{\Delta x}^{k+1}(\lambda) - (1-\lambda)\Delta x^k\| \le \tfrac{1}{2}\lambda^2 \omega_k \|\Delta x^k\|^2\,,$$

which leads to estimates for the Kantorovich quantities

$$[h_k] = [\omega_k]\|\Delta x^k\| := \frac{2\|\overline{\Delta x}^{k+1}(\lambda) - (1-\lambda)\Delta x^k\|}{\lambda^2\|\Delta x^k\|} \le h_k\,.$$

The evaluation of such an estimate requires at least one trial value λ_k^0 (or x^{k+1}, respectively). As a consequence, it can only be helpful in the design of a *correction strategy* for the damping factor :

$$\lambda_k^{j+1} := \min\big(\tfrac{1}{2}\lambda, 1/[h_k^j]\big)\,\big|_{\lambda=\lambda_k^j} \tag{3.48}$$

for repetition index $j = 0, 1, \ldots$.

Prediction strategy. We are therefore still left with the task of constructing an efficient initial estimate λ_k^0. As it turns out, such an estimate can only be gained by switching from the above defined Lipschitz constant ω_k to some slightly different definition:

$$\|F'(x^k)^{-1}\big(F'(x) - F'(x^k)\big)v\| \le \overline{\omega}_k\|x - x^k\|\|v\|\,,$$

wherein the direction v is understood to be 'not too far away from' the direction $x - x^k$ in order to mimic the above definition (3.42). With this modified Lipschitz condition we may proceed to derive the following bounds:

$$\|\overline{\Delta x}^k - \Delta x^k\| = \big\|\ \big(F'(x^{k-1})^{-1} - F'(x^k)^{-1}\big)F(x^k)\ \big\| =$$

$$= \big\|\ F'(x^k)^{-1}\big(F'(x^k) - F'(x^{k-1})\big)\overline{\Delta x}^k\ \big\| \le\ \overline{\omega}_k\lambda_{k-1}\|\Delta x^{k-1}\| \cdot \|\overline{\Delta x}^k\|.$$

This bound inspires the local estimate

$$[\overline{\omega}_k] := \frac{\|\overline{\Delta x}^k - \Delta x^k\|}{\lambda_{k-1}\|\Delta x^{k-1}\| \cdot \|\overline{\Delta x}^k\|} \le \overline{\omega}_k\,,$$

wherein, as required in the definition above, the direction $\overline{\Delta x}^k$ is 'not too far away from' the direction Δx^{k-1}. In any case, the above computational estimate exploits the 'newest' information that is available in the course of the algorithm just before deciding about the initial damping factor. We have thus constructed a *prediction strategy* for $k > 0$:

$$\lambda_k^0 := \min(1, \mu_k)\,, \qquad \mu_k := \frac{\|\Delta x^{k-1}\|}{\|\overline{\Delta x}^k - \Delta x^k\|} \cdot \frac{\|\overline{\Delta x}^k\|}{\|\Delta x^k\|} \cdot \lambda_{k-1}\,. \tag{3.49}$$

The only empirical choice left to be made is the starting value λ_0^0. In the public domain code `NLEQ1` (see [161]) this value is made an input parameter: if the user classifies the problem as 'mildly nonlinear', then $\lambda_0^0 = 1$ is set internally; otherwise the problem is regarded as 'highly nonlinear' and $\lambda_0^0 = \lambda_{\min} \ll 1$ is set internally.

Intermediate quasi-Newton steps. Whenever $\lambda_k = 1$ and the natural monotonicity test yields

$$\Theta_k = \frac{\|\overline{\Delta x}^{k+1}\|}{\|\Delta x^k\|} < \tfrac{1}{2},$$

then the error oriented quasi-Newton method of Section 2.1.4 may be applied in the present context—compare Theorem 2.9. In this case, Jacobian evaluations are replaced by *Broyden rank-1 updates*. As for a possible switch back from quasi-Newton steps to Newton steps just look into the details of the informal quasi-Newton algorithm `QNERR`.

Termination criterion. Instead of the termination criterion (2.14) we may here use its cheaper substitute

$$\|\overline{\Delta x}^{k+1}\| \le \mathrm{XTOL}\ .$$

Recall that then $x^{k+2} = x^{k+1} + \overline{\Delta x}^{k+1}$ is cheaply available with an accuracy of $O(\mathrm{XTOL}^2)$.

The here described adaptive trust region strategy leads to the global Newton algorithm `NLEQ-ERR`, which is a slight modification of the quite popular code `NLEQ1` [161].

Algorithm `NLEQ-ERR`**.** Set a required *error accuracy* ε sufficiently above the machine precision.
Guess an initial iterate x^0. Evaluate $F(x^0)$.
Set a damping factor either $\lambda_0 := 1$ or $\lambda_0 \ll 1$.
All norms of corrections below are understood to be scaled smooth norms such as $\|D^{-1} \cdot\|_2$, where D is a diagonal scaling matrix, which can be iteratively adapted together with the Jacobian matrix.

For iteration index $k = 0, 1, \ldots$ do:

1. **Step** k: Evaluate Jacobian matrix $F'(x^k)$. Solve linear system

$$F'(x^k)\Delta x^k = -F(x^k)\,.$$

Convergence test: If $\|\Delta x^k\| \le \varepsilon$: **stop**. Solution found $x^* := x^k + \Delta x^k$.

For $k > 0$: compute a prediction value for the damping factor

$$\lambda_k := \min(1, \mu_k), \quad \mu_k := \frac{\|\Delta x^{k-1}\| \cdot \|\overline{\Delta x}^{k}\|}{\|\overline{\Delta x}^{k} - \Delta x^k\| \cdot \|\Delta x^k\|} \cdot \lambda_{k-1}\,.$$

Regularity test: If $\lambda_k < \lambda_{\min}$: **stop**. Convergence failure.

2. **Else**: compute the trial iterate $x^{k+1} := x^k + \lambda_k \Delta x^k$ and evaluate $F(x^{k+1})$. Solve linear system ('old' Jacobian, 'new' right hand side):

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

3. Compute the monitoring quantities

$$\Theta_k := \frac{\|\overline{\Delta x}^{k+1}\|}{\|\Delta x^k\|}\,, \qquad \mu'_k := \frac{\frac{1}{2}\|\Delta x^k\| \cdot \lambda_k^2}{\|\overline{\Delta x}^{k+1} - (1-\lambda_k)\Delta x^k\|}\,.$$

If $\Theta_k \geq 1$ (or, if **restricted**: $\Theta_k > 1 - \lambda_k/4$):

then replace λ_k by $\lambda'_k := \min(\mu'_k, \frac{1}{2}\lambda_k)$. **Go to** Regularity test.

Else: let $\lambda'_k := \min(1, \mu'_k)$.

If $\lambda'_k = \lambda_k = 1$:

If $\|\overline{\Delta x}^{k+1}\| \leq \varepsilon$: **stop.**

Solution found $x^* := x^{k+1} + \overline{\Delta x}^{k+1}$.

If $\Theta_k < \frac{1}{2}$: switch to `QNERR`

Else: **If** $\lambda'_k \geq 4\lambda_k$: replace λ_k by λ'_k and **goto** 2.

Else: accept x^{k+1} as new iterate.

Goto 1 with $k \to k+1$.

In what follows we want to demonstrate the main feature of this algorithm at a rather simple, but very illustrative example for $n = 2$.

Example 3.2 [161]. The equations to be solved are

$$\begin{aligned} \exp(x^2 + y^2) - 3 &= 0\,, \\ x + y - \sin\bigl(3(x+y)\bigr) &= 0\,. \end{aligned}$$

For this simple problem, *critical interfaces* with singular Jacobian can be calculated to be the straight line

$$y = x$$

and the family of parallels

$$y = -x \pm \tfrac{1}{3} \arccos\left(\tfrac{1}{3}\right) \pm \tfrac{2}{3}\pi \cdot j\,,\ j = 0, 1, 2 \ldots\,.$$

For illustration, the quadratic domain

$$-1.5 \leq x\,,\ y \leq 1.5$$

is picked out. This domain contains the six different solution points and five critical interfaces.

Computation of Newton paths. As derived in Section 3.1.4, the Newton path $\overline{x}(\lambda)$, $\lambda \in [0,1]$ may be defined either by the homotopy

$$F\big(\overline{x}(\lambda)\big) - (1-\lambda)F(x^k) = 0$$

or by the *initial value problem*

$$F'\left(\overline{x}\right)\frac{d\overline{x}}{d\lambda} = -F(x^k)\,,\ \overline{x}(0) = x^k\,.$$

This implicit ordinary differential equation can be solved numerically, say, by implicit BDF codes [98] like `DASSL` [167] due to L. Petzold or by linearly implicit extrapolation codes like `LIMEX` [75, 79]. In any such discretization, linear subsystems of the kind

$$F'(\overline{x})\Delta\overline{x} - \beta\Delta\lambda F''(\overline{x})[F'(x^k)^{-1}F(x^k), \Delta\overline{x}] = -\Delta\lambda F(x^k)$$

must be solved. Apparently, this algorithmic approach involves *second order derivative* information in *tensor* form—to be compared with the above described global Newton methods, which involve second order derivative information only in *scalar* form (Lipschitz constant estimates $[\omega]$ entering into the adaptive trust region strategies). Note, however, that the Newton path should be understood as an underlying geometric concept rather than an object to be actually computed.

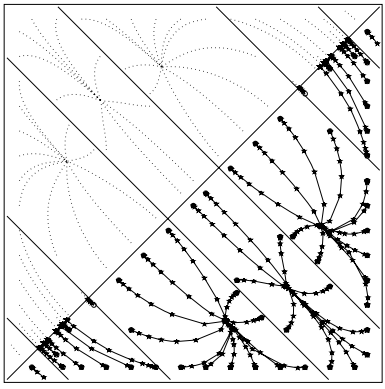

Fig. 3.10. Example 3.2: Newton paths ($\cdots$) versus Newton sequences (—).

Newton paths versus Newton sequences. Figure 3.10 shows the various Newton paths (left upper part) and sequences of Newton iterates (right lower part) as obtained by systematic variation of the initial guesses x_0—separated by the symmetry line $y = x$. The Newton paths have been integrated by

LIMEX, whereas the Newton sequences have been computed by NLEQ1 [161]. As predicted by theory—compare Section 3.1.4 and Figure 3.4 therein—each Newton path either ends at a solution point or at a critical point with singular Jacobian. The figure clearly documents that this same structure is mimicked by the sequence of Newton iterates as selected by the error oriented trust region strategy.

Attraction basins. An adjacent question of interest is the connectivity structure of the different *attraction basins* for the global Newton iteration around the different solution points. In order to visualize these structures, a rectangular grid of starting points (with grid size $\Delta = 0.06$) has been defined and the associated global Newton iteration performed.

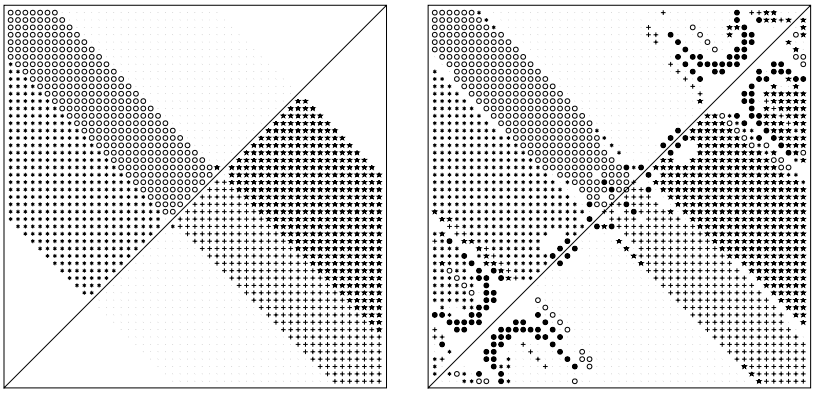

Fig. 3.11. Example 3.2: attraction basins. *Left:* Global Newton method, code NLEQ1 [161]. *Right:* hybrid method, code HYBRJ1 [153]. Outliers are indicated as bullets (•).

The results are represented in Figure 3.11: apart from very few exceptional 'corner points', the attraction basins nicely model the theoretical connectivity structure, which is essentially defined by the critical interfaces—a highly satisfactory performance of the herein advocated global Newton algorithm (code NLEQ1 due to [161]). For comparison, the attraction basins for a hybrid method (code HYBRJ1 in MINPACK due to [153]) are given as well. There are still some people who prefer the rather chaotic convergence pattern of such algorithms. However, in most scientific and engineering problems, a *crossing beyond critical interfaces* is undesirable, because this means an unnoticed switching between different solutions—an important aspect especially in the parameter dependent case.

3.3.4 Inexact Newton-ERR methods

Suppose that, instead of the *exact* Newton corrections Δx^k, we are only able to compute *inexact* Newton corrections δx^k from (dropping the inner iteration index i)

$$F'(x^k)\left(\delta x^k - \Delta x^k\right) = r^k\,,\ x^{k+1} = x^k + \lambda_k \delta x^k\,,\ 0 < \lambda_k \leq 1\,, \quad k = 0, 1, \ldots\,.$$

As for local inexact Newton-ERR methods (Section 2.1.5), we characterize the inner iteration errors by the quantity

$$\delta_k = \frac{\|\delta x^k - \Delta x^k\|}{\|\delta x^k\|}\,. \tag{3.50}$$

Inner iterative solvers treated here are either `CGNE` or `GBIT`. As a guiding principle for global convergence, we will focus on *natural monotonicity* (3.41) subject to the perturbation coming from the truncation of the inner iteration.

Accuracy matching: inexact Newton corrections. First, we study the contraction factors

$$\Theta_k(\lambda) = \frac{\|\overline{\Delta x}^{k+1}\|}{\|\Delta x^k\|}$$

in terms of the exact Newton corrections Δx^k and the exact simplified Newton corrections $\overline{\Delta x}^{k+1}$ defined via

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

Note that the *inexact* Newton correction arises in the argument on the right. Of course, none of the above exact Newton corrections will be actually computed.

Lemma 3.17 *We consider the inexact Newton iteration with* `CGNE` *or* `GBIT` *as inner iteration. Assume* $\delta_k < \frac{1}{2}$. *Then, with* $h_k^\delta = \omega\|\delta x^k\|$, *we obtain the estimate*

$$\Theta_k(\lambda) \leq 1 - \left(1 - \frac{\delta_k}{1-\delta_k}\right)\lambda + \tfrac{1}{2}\lambda^2 \frac{h_k^\delta}{1-\delta_k}\,. \tag{3.51}$$

The optimal damping factor is

$$\overline{\lambda}_k = \min\left(1, \frac{1-2\delta_k}{h_k^\delta}\right)\,. \tag{3.52}$$

If we impose

$$\delta_k = \frac{\rho}{2}\lambda h_k^\delta\,, \quad \rho \leq 1\,, \tag{3.53}$$

we are led to the optimal damping factor

$$\overline{\lambda}_k = \min\left(1, \frac{1}{(1+\rho)h_k^\delta}\right)\,. \tag{3.54}$$

Proof. First, we derive the identity

$$\overline{\Delta x}^{k+1}(\lambda) = \Delta x^k - \lambda \delta x^k - F'(x^k)^{-1} \int_{s=0}^{\lambda} \left(F'(x^k + s\delta x^k) - F'(x^k) \right) \delta x^k \, ds.$$

Upon inserting definition (3.50) and using the triangle inequality

$$\frac{\|\delta x^k - \Delta x^k\|}{\|\Delta x^k\|} \leq \frac{\delta_k}{1 - \delta_k}, \quad \|\delta x^k\| \leq \frac{\|\Delta x^k\|}{1 - \delta_k},$$

the above estimate (3.51) follows directly. The optimal damping factors $\overline{\lambda}_k$ in the two different forms arise by minimization of the upper bound parabola, as usual. □

Condition (3.53) is motivated by the idea that the $O(\lambda)$ perturbation due to the inner iteration should not dominate the $O(\lambda^2)$ term, which characterizes the nonlinearity of the problem.

Accuracy matching strategy. Upon inserting $\lambda = \overline{\lambda}_k$ into (3.53) and selecting some $\rho \leq 1$, we are led to

$$\delta_k \leq \bar{\delta} = \frac{\rho}{2(1+\rho)} \leq \tfrac{1}{4} \quad \text{for} \quad \overline{\lambda}_k < 1 \tag{3.55}$$

and

$$\delta_k \leq \frac{\rho}{2} h_k^\delta \quad \text{for} \quad \overline{\lambda}_k = 1 .$$

Of course, the realization of the latter rule will be done via computational Kantorovich estimates $[h_k^\delta] \leq h_k^\delta$ such that

$$\delta_k \leq \frac{\rho}{2} [h_k^\delta] \quad \text{for} \quad \overline{\lambda}_k = 1 . \tag{3.56}$$

Obviously, the relation (3.55) reflects the 'fight for the first binary digit' as discussed in the preceding section; under this condition the optimal damping factors (3.52) and (3.54) are identical. In passing we note that requirement (3.56) nicely agrees with the 'quadratic convergence mode' (2.62) in the *local* Newton-ERR methods. (The slight difference reflects the different contraction factors in the local and the global case.) The condition (3.56) is a simple nonlinear scalar equation for an upper bound of δ_k.

As already mentioned at the beginning of this section, the exact natural monotonicity test cannot be directly implemented within our present algorithmic setting. We will, however, use this test and the corresponding optimal damping factor as a guideline.

Accuracy matching: inexact simplified Newton corrections. In order to construct an appropriate substitute for the nonrealizable Θ_k, we recur to the *inexact Newton path* $\widetilde{x}(\lambda), \lambda \in [0,1]$, from (3.34), which satisfies

$$F\left(\widetilde{x}(\lambda)\right) = (1-\lambda)F(x^k) + \lambda r^k\,.$$

Recall that the local inexact Newton correction δx^k can be interpreted as the tangent direction $\dot{\widetilde{x}}(0)$ in x^k. On this background, we are led to define a perturbed (exact) simplified Newton correction via

$$F'(x^k)\widetilde{\Delta x}^{k+1} = -F(x^{k+1}) + r^k\,. \tag{3.57}$$

Lemma 3.18 *With the notation and definitions of this section the following estimate holds:*

$$\|\widetilde{\Delta x}^{k+1} - (1-\lambda)\delta x^k\| \le \tfrac{1}{2}\lambda^2 h_k^\delta \|\delta x^k\|\,. \tag{3.58}$$

Proof. It is easy to verify the identity

$$\widetilde{\Delta x}^{k+1}(\lambda) - (1-\lambda)\delta x^k = -F'(x^k)^{-1}\int\limits_{s=0}^{\lambda}\left(F'(x^k + s\delta x^k) - F'(x^k)\right)\delta x^k ds\,.$$

From this identity, the above estimate can be immediately derived in the usual manner. □

Of course, the linear equation (3.57) can only be solved iteratively. This means the computation of an *inexact simplified Newton correction* satisfying

$$F'(x^k)\widetilde{\delta x_i}^{k+1} = \left(-F(x^{k+1}) + r^k\right) + \widetilde{r}_i^{k+1}$$

for inner iteration index $i = 0, 1, \ldots$. (In what follows, we will drop this index wherever convenient.)

Initial values for inner iterations. In view of (3.58) we set the initial value

$$\widetilde{\delta x_0}^{k+1} = (1-\lambda)\delta x^k\,. \tag{3.59}$$

This means that the inner iteration has to recover only second order information. The same idea also supplies an initial guess for the inner iteration of the inexact ordinary Newton corrections:

$$\delta x_0^k = \widetilde{\delta x}^k\,. \tag{3.60}$$

This cross-over of initial values has proven to be really important in the realization of any Newton-ERR method.

In what follows, we replace the nonrealizable contraction factor $\Theta_k(\lambda)$ by its realizable inexact counterpart

$$\widetilde{\Theta}_k(\lambda) = \frac{\|\widetilde{\delta x}^{k+1}\|}{\|\delta x^k\|} \tag{3.61}$$

and study its dependence on the damping factor λ.

We start with CGNE as inner iterative solver.

Lemma 3.19 *Notation as just introduced. Assume that the inner* CGNE *iteration with initial guess* (3.59) *has been continued up to some iteration index* $i > 0$ *such that*

$$\widetilde{\rho}_i = \frac{\|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_i^{k+1}\|}{\|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_0^{k+1}\|} < 1 \,. \tag{3.62}$$

Then we obtain the estimate

$$\|\widetilde{\delta x}_i^{k+1} - (1-\lambda)\delta x^k\| \le \tfrac{1}{2}\sqrt{1-\widetilde{\rho}_i^2}\,\lambda^2 h_k^\delta \|\delta x^k\| \,. \tag{3.63}$$

Proof. In our context, the orthogonal decomposition (1.28) reads

$$\|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_0^{k+1}\|^2 = \|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_i^{k+1}\|^2 + \|\widetilde{\delta x}_i^{k+1} - \widetilde{\delta x}_0^{k+1}\|^2 \,. \tag{3.64}$$

Insertion of (3.62) then leads to

$$(1-\widetilde{\rho}_i^2)\|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_0^{k+1}\|^2 = \|\widetilde{\delta x}_i^{k+1} - \widetilde{\delta x}_0^{k+1}\|^2 \,. \tag{3.65}$$

With the insertion of (3.65) into (3.58) the proof is complete. □

Observe that in CGNE the condition $\widetilde{\rho}_i < 1$ arises by construction. The parameter $\widetilde{\rho}_i$, however, is not directly computable from (3.62): the denominator cannot be evaluated, since we do not have $\widetilde{\Delta x}^{k+1}$, but for the numerator a rough estimate

$$\widetilde{\epsilon}_i \approx \|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_i^{k+1}\|$$

is available (see Section 1.4.3). Therefore we may define the computable parameter

$$\overline{\rho}_i = \frac{\|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_i^{k+1}\|}{\|\widetilde{\delta x}_i^{k+1} - \widetilde{\delta x}_0^{k+1}\|} \approx \frac{\widetilde{\epsilon}_i}{\|\widetilde{\delta x}_i^{k+1} - \widetilde{\delta x}_0^{k+1}\|} \,. \tag{3.66}$$

By means of (3.65), we then get

$$\|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_i^{k+1}\| = \overline{\rho}_i \|\widetilde{\delta x}_i^{k+1} - \widetilde{\delta x}_0^{k+1}\| = \overline{\rho}_i \sqrt{1-\widetilde{\rho}_i^2}\,\|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_0^{k+1}\| \,.$$

This result can be compared with (3.62) to supply the identification

$$\overline{\rho}_i = \widetilde{\rho}_i/\sqrt{1-\widetilde{\rho}_i^2} \quad \text{or} \quad \widetilde{\rho}_i = \overline{\rho}_i/\sqrt{1+\overline{\rho}_i^2}\,. \tag{3.67}$$

For GBIT as inner iterative solver, we also use $\overline{\rho}_i$ from (3.66), but in combination with a slightly different estimate.

Lemma 3.20 *Notation as in the preceding lemma. Let* $\widetilde{\rho}_i < 1$ *according to* (3.62)*. Then, for* GBIT *as inner iteration, we obtain*

$$\|\widetilde{\delta x}_i^{k+1} - (1-\lambda)\delta x^k\| \le \frac{1+\widetilde{\rho}_i}{2}\lambda^2 h_k^\delta \|\delta x^k\|\,. \tag{3.68}$$

Proof. We drop the iteration index i. For GBIT, we cannot do better than apply the triangle inequality

$$\|\widetilde{\delta x}^{k+1} - (1-\lambda)\delta x^k\| \le \|\widetilde{\Delta x}^{k+1} - (1-\lambda)\delta x^k\| + \|\widetilde{\delta x}^{k+1} - \widetilde{\Delta x}^{k+1}\|\,.$$

With the requirement (3.62) we get

$$\|\widetilde{\delta x}^{k+1} - (1-\lambda)\delta x^k\| \le (1+\widetilde{\rho})\|\widetilde{\Delta x}^{k+1} - (1-\lambda)\delta x^k\|\,. \tag{3.69}$$

Application of Lemma 3.18 then directly verifies the estimate (3.68). □

Note that in GBIT the condition $\widetilde{\rho}_i < 1$ is not automatically fulfilled, but must be assured by the implementation. In order to actually estimate $\widetilde{\rho}_i$, we again recur to $\overline{\rho}_i$ from (3.66). If we combine (3.66) with (3.69), we now arrive at

$$\|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_i^{k+1}\| \le \overline{\rho}_i(1+\widetilde{\rho}_i)\|\widetilde{\Delta x}^{k+1} - \widetilde{\delta x}_0^{k+1}\|\,.$$

Comparison with (3.62) then supplies the identification

$$\overline{\rho}_i = \widetilde{\rho}_i/(1+\widetilde{\rho}_i) \quad \text{or} \quad \widetilde{\rho}_i = \overline{\rho}_i/(1-\overline{\rho}_i) \text{ for } \overline{\rho}_i < 1\,. \tag{3.70}$$

Accuracy matching strategy. On the basis of the presented convergence analysis, we might suggest to run the inner iteration until

$$\widetilde{\rho}_i \le \widetilde{\rho}_{\max} \quad \text{with} \quad \widetilde{\rho}_{\max} \le \tfrac{1}{4}$$

for both CGNE and GBIT. By means of the transformations (3.67) or (3.70), respectively, this idea can be transferred to the realizable strategy

$$\overline{\rho}_i \le \overline{\rho}_{\max}\,. \tag{3.71}$$

Affine covariant Kantorovich estimates. Upon applying our algorithmic paradigm from Section 1.2.3, we will replace the optimal damping factor $\overline{\lambda}_k$ by computational estimates

$$[\overline{\lambda}_k] = \min\left(1, \frac{1-2\delta_k}{[h_k^\delta]}\right) = \min\left(1, \frac{1}{(1+\rho)[h_k^\delta]}\right) , \tag{3.72}$$

where $[h_k^\delta] = [\omega]\|\delta x^k\| \le h_k^\delta$ are Kantorovich estimates to be carefully selected.

For CGNE, we will exploit (3.63) thus obtaining the *a-posteriori* estimates

$$[h_k^\delta]_i = \frac{2\|\widetilde{\delta x}_i^{k+1} - \widetilde{\delta x}_0^{k+1}\|_2}{\sqrt{1-\widetilde{\rho}_i^2}\lambda^2\|\delta x^k\|_2} \le h_k^\delta .$$

Note that (3.64) assures the saturation property

$$[h_k^\delta]_i \le [h_k^\delta]_{i+1} \le h_k^\delta .$$

Replacing $\widetilde{\rho}_i$ by $\overline{\rho}_i$ then gives rise to the computable a-posteriori estimate

$$[h_k^\delta]_i \approx \frac{2\sqrt{1+\overline{\rho}_i^2}\|\widetilde{\delta x}_i^{k+1} - \widetilde{\delta x}_0^{k+1}\|_2}{\lambda^2\|\delta x^k\|_2} . \tag{3.73}$$

For GBIT, we will exploit (3.68) and obtain the a-posteriori estimates

$$[h_k^\delta]_i = \frac{2\|\widetilde{\delta x}_i^{k+1} - \widetilde{\delta x}_0^{k+1}\|}{(1+\widetilde{\rho}_i)\lambda^2\|\delta x^k\|} \le h_k^\delta .$$

Here a saturation property does not hold. Replacement of $\widetilde{\rho}_i$ by $\overline{\rho}_i$ leads to the computable a-posteriori estimates

$$[h_k^\delta]_i \approx \frac{2(1-\overline{\rho}_i)\|\widetilde{\delta x}_i^{k+1} - \widetilde{\delta x}_0^{k+1}\|}{\lambda^2\|\delta x^k\|} . \tag{3.74}$$

As for the construction of computational *a-priori* Kantorovich estimates, we suggest to simply go back to the definitions and realize the estimate

$$[h_{k+1}^\delta] = \frac{\|\delta x^{k+1}\|}{\|\delta x^k\|}[h_k^\delta]_* , \tag{3.75}$$

where $[h_k^\delta]_*$ denotes the final estimate $[h_k^\delta]_i$ from either (3.73) for CGNE or (3.74) for GBIT, i.e. the estimate obtained at the final inner iteration step $i = \widetilde{i}_k$ of the previous outer iteration step k.

Bit counting lemma. Once computational estimates $[h_k^\delta]$ are available, we may realize the damping strategy (3.72). In analogy to Lemma 3.16, we now study the influence of the accuracy of the Kantorovich estimates.

Lemma 3.21 *Notation as just introduced. Let an inexact Newton-ERR method with damping factors* $\lambda = [\overline{\lambda}_k]$ *due to* (3.72) *be realized. Assume that*

$$0 \le h_k^\delta - [h_k^\delta] < \sigma \max\Big(\frac{1}{1+\rho}, [h_k^\delta]\Big) \text{ for some } \sigma < 1. \tag{3.76}$$

Then the exact natural contraction factor satisfies

$$\Theta_k(\lambda) = \frac{\|\overline{\Delta x}^{k+1}\|}{\|\Delta x^k\|} < 1 - \frac{1}{2+\rho}(1-\sigma)\lambda\,.$$

For `CGNE`*, the inexact natural contraction factor is bounded by*

$$\widetilde{\Theta}_k = \frac{\|\widetilde{\delta x}^{k+1}\|}{\|\delta x^k\|} < 1 - \left(1 - \tfrac{1}{2}\frac{1+\sigma}{1+\rho}\right)\lambda\,.$$

For `GBIT`*, the inexact natural contraction factor is bounded by*

$$\widetilde{\Theta}_k = \frac{\|\widetilde{\delta x}^{k+1}\|}{\|\delta x^k\|} < 1 - \left(1 - \tfrac{1}{2}\frac{(1+\widetilde{\rho})(1+\sigma)}{1+\rho}\right)\lambda\,.$$

Proof. Throughout this proof, we will omit any results for $\lambda = 1$, since these can be directly verified by mere insertion. This means that we assume $\lambda = [\overline{\lambda}_k] < 1$ in the following.

For the exact natural monotonicity test we return to the inequality (3.51), which reads

$$\widetilde{\Theta}_k \le 1 - \frac{1-2\delta_k}{2(1-\delta_k)}\lambda + \tfrac{1}{2}\lambda^2\frac{h_k^\delta}{2(1-\delta_k)}\,.$$

Insertion of $\lambda = [\overline{\lambda}_k] < 1$ then yields

$$\lambda h_k^\delta = (1-2\delta_k)\frac{h_k^\delta}{[h_k^\delta]} < \frac{1+\sigma}{1+\rho}\,.$$

Inserting this into the above upper bound and switching from the parameter δ_k to ρ via (3.55) then verifies the first statement of the lemma.

For the inexact natural monotonicity test with `CGNE` as inner iterative solver, we go back to (3.63), which yields

$$\widetilde{\Theta}_k(\lambda) \le 1 - \lambda + \tfrac{1}{2}\sqrt{1-\widetilde{\rho}_i^2}\lambda^2 h_k^\delta < 1 - \lambda + \tfrac{1}{2}\lambda^2 h_k^\delta\,.$$

If we again insert the above upper bound λh_k^δ, we arrive at

$$\widetilde{\Theta}_k < 1 - \lambda + \tfrac{1}{2}\lambda\frac{1+\sigma}{1+\rho}\,,$$

which is equivalent to the second statement of the lemma.

For the inexact natural monotonicity test with GBIT as inner iterative solver, we recur to (3.68), which yields

$$\widetilde{\Theta}_k(\lambda) \le 1 - \lambda + \frac{1+\widetilde{\rho}}{2}\lambda^2 h_k^\delta \,.$$

Following the same lines as for CGNE now supplies the upper bound

$$\widetilde{\Theta}_k < 1 - \lambda + \frac{1+\widetilde{\rho}}{2}\lambda\frac{1+\sigma}{1+\rho} \,,$$

which finally confirms the third statement of the lemma. □

Inexact natural monotonicity tests. Suppose now that we require at least one binary digit in the Kantorovich estimate, i.e., $\sigma < 1$ in Lemma 3.21. In this case, exact natural monotonicity

$$\Theta_k(\lambda) = \frac{\|\overline{\Delta x}^{k+1}\|}{\|\Delta x^k\|} < 1$$

would hold—which, however, is not realizable in the present algorithmic setting.

For CGNE, a computable substitute is the inexact natural monotonicity test

$$\widetilde{\Theta}_k = \frac{\|\widetilde{\delta x}^{k+1}\|}{\|\delta x^k\|} < 1 - \frac{\rho}{1+\rho}\lambda \,. \tag{3.77}$$

For GBIT, we similarly get

$$\widetilde{\Theta}_k = \frac{\|\widetilde{\delta x}^{k+1}\|}{\|\delta x^k\|} < 1 - \frac{\rho-\widetilde{\rho}}{1+\rho}\lambda \,. \tag{3.78}$$

The latter result seems to suggest the setting $\widetilde{\rho} \le \rho$ to assure $\widetilde{\Theta}_k < 1$; otherwise inexact natural monotonicity need not hold.

Correction strategy. This part of the adaptive trust region strategy applies, if inexact natural monotonicity, (3.77) for CGNE or (3.78) for GBIT, is violated. Then the damping strategy can be based on the *a-posteriori* Kantorovich estimates (3.73) for CGNE or (3.74) for GBIT, respectively, again written as $[h_k^\delta]_*$. In view of the exact correction strategy (3.48) with repetition index $j = 0, 1, \ldots$, we set

$$\lambda_k^{j+1} := \min\left(\tfrac{1}{2}\lambda, \frac{1}{(1+\rho)[h_k^\delta]_*}\right)\Big|_{\lambda=\lambda_k^j} \,.$$

Prediction strategy. This part of the trust region strategy is based on the *a-priori* Kantorovich estimates (3.75). On the basis of information from outer iteration step $k-1$, we obtain

$$\lambda_k^0 = \min\left(1, \frac{1}{(1+\rho)[h_k^\delta]}\right), \quad k > 0\,.$$

For $k = 0$, we can only start with some prescribed initial value λ_0^0 to be chosen by the user.

If $\lambda < \lambda_{\min}$ for some threshold value $\lambda_{\min} \ll 1$ arises in the prediction or the correction strategy, then the outer iteration must be terminated—indicating some critical point with ill-conditioned Jacobian.

The described inexact Newton-ERR methods are realized in the programs `GIANT-CGNE` and `GIANT-GBIT` with error oriented adaptive trust region strategy and corresponding matching between inner and outer iteration. The realization of the local Newton part here is slightly different from the one suggested in the previous Section 2.1.5, since here we have the additional information of the inexact simplified Newton correction $\widetilde{\delta x}$ available.

Algorithms `GIANT-CGNE` and `GIANT-GBIT`.

1. **Step k** Evaluate $F'(x^k)$.
   Solve linear system

   $$F'(x^k)\delta x_i^k = -F(x^k) + r_i^k \quad \text{for} \quad i = 0, 1, \ldots, i_k$$

   iteratively by `CGNE` or `GBIT`. Control of i_k via accuracy matching strategy (3.55) or (3.56).

   If $\|\delta x^k\| \le$ XTOL: **Solution:** $x^* = x^k + \delta x^k$.

   Else: For $k = 0$: select λ_0 ad hoc.
   For $k > 0$: determine $\lambda_k = \min\left(1, \frac{1}{(1+\rho)[h_k^\delta]}\right)$ from the prediction strategy.

   Regularity test. If $\lambda_k < \lambda_{\min}$: **stop**

2. **Else:** compute trial iterate $x^{k+1} := x^k + \lambda_k \delta x^k$ and evaluate $F(x^{k+1})$.
   Solve linear system

   $$F'(x^k)\widetilde{\delta x}_i^{k+1} = \left(-F(x^{k+1}) + r^k\right) + \widetilde{r}_i^{k+1} \quad \text{for} \quad i = 0, 1, \ldots, \widetilde{i}_k$$

   iteratively by `CGNE` or `GBIT`. Control of $\widetilde{i}_k$ via accuracy matching strategy (3.71).

Computation of Kantorovich estimates $[h_k^\delta]$.

3. Evaluate the monitor

$$\widetilde{\Theta}_k := \frac{\|\widetilde{\delta x}^{k+1}\|_2}{\|\delta x^k\|_2} .$$

If monotonicity test (3.77) for `CGNE` or (3.78) for `GBIT` violated, **then** refine λ_k according to correction strategy and go to **regularity test.**

Else go to 1.

As soon as the global Newton-ERR method approaches the solution point, one may either directly switch to the local Newton-ERR methods presented in Section 2.1.5 or merge the Kantorovich estimates from here with the 'standard', 'linear', or 'quadratic' convergence mode as described there.

Remark 3.2 The *residual based* algorithm `GIANT-GMRES` as presented in Section 3.2.3 above seems to represent an easier implementable alternative to the here elaborated error oriented algorithms `GIANT-CGNE` and `GIANT-GBIT`. This is only true, if a 'good' *left* preconditioner C_L is available. Indeed, if spectral equivalence $C_L A \sim I$ holds, then the *preconditioned initial residual* satisfies $C_L r_0 \sim x^0 - x^*$. Otherwise, the here presented error oriented algorithms realize some *nonlinear preconditioning*.

A numerical comparison of `GIANT-CGNE`, `GIANT-GBIT`, and `NLEQ-ERR` in the context of discretized nonlinear PDEs is given in Section 8.2.1 below.

Bibliographical Note. The first affine covariant convergence proof for local inexact Newton methods has been given by T.J. Ypma [203]. A first error oriented global inexact Newton algorithm has been suggested by P. Deuflhard [67] on the basis of some slightly differing affine covariant convergence analysis. These suggestions led to the code `GIANT` by U. Nowak and coworkers [160], wherein the inner iteration has been realized by an earlier version of `GBIT`.

3.4 Convex Functional Descent

In the present section we want to minimize a general convex function f or, equivalently, solve the nonlinear system $F(x) = f'^T(x) = 0$ with $F'(x) = f''(x)$ symmetric positive definite. It is not at all clear whether for general functional the damped Newton method still is an efficient globalization. As the damped Newton method can be interpreted as a tangent continuation along the Newton path, we first study the behavior of an arbitrary convex functional f along the Newton path $\overline{x}(\lambda)$ as a function of λ.

Lemma 3.22 *Let $f \in C^2(D)$ denote some strictly convex functional to be minimized over some convex domain $D \in \mathbb{R}^n$. Let $F'(x) = f''(x)$ be symmetric positive definite in D and let $\overline{x} : [0,1] \to D$ denote the Newton path starting at some iterate $\overline{x}(0) = x^k$ and ending at the solution point $\overline{x}(1) = x^*$ with $F(x^*) = f'^T(x^*) = 0$. Then $f(\overline{x}(\lambda))$ is a strictly monotone decreasing function of λ.*

Proof. In the usual way we just verify that

$$f(\overline{x}(\lambda)) - f(x^k) = \int_{\sigma=0}^{\lambda} \langle F(\overline{x}(\sigma)), \dot{\overline{x}}(\sigma) \rangle \, d\sigma \,.$$

Insertion of (3.22) and (3.24) then leads to

$$f(\overline{x}(\lambda)) - f(x^k) = - \int_{\sigma=0}^{\lambda} (1-\sigma) \| F'(\overline{x}(\sigma))^{-1/2} F(x^k) \|_2^2 d\sigma$$

with a strictly positive definite integrand. Therefore, for $0 \le \lambda_2 < \lambda_1 \le 1$:

$$f(\overline{x}(\lambda_1)) - f(\overline{x}(\lambda_2)) = - \int_{\sigma=\lambda_2}^{\lambda_1} (1-\sigma) \| F'(\overline{x}(\sigma))^{-1/2} F(x^k) \|_2^2 d\sigma < 0 \,.$$

□

Obviously, this result is the desired generalization of the monotone level function decrease (3.23). We are now ready to analyze the *damped Newton iteration* ($k = 0, 1, \ldots$)

$$F'(x^k) \Delta x^k = -F(x^k), \qquad x^{k+1} = x^k + \lambda_k \Delta x^k, \qquad \lambda_k \in]0, 1]$$

under the requirement of iterative *functional decrease* $f(x^{k+1}) < f(x^k)$.

3.4.1 Affine conjugate convergence analysis

As in the preceding sections, we first study the local reduction properties of the damped Newton method within one iterative step from iterate x^k to iterate x^{k+1}.

Theorem 3.23 *Let $f : D \to \mathbb{R}^1$ be a strictly convex C^2-functional to be minimized over some open convex domain $D \subset \mathbb{R}^n$. Let $F(x) = f'(x)^T$ and $F'(x) = f''(x)$ symmetric and strictly positive definite. For $x, y \in D$, assume the special affine conjugate Lipschitz condition*

$$\|F'(x)^{-1/2}(F'(y)-F'(x))(y-x)\| \le \omega\|F'(x)^{1/2}(y-x)\|^2 \tag{3.79}$$

with $0 \le \omega < \infty$. *For some iterate* $x^k \in D$, *define the quantities*

$$\epsilon_k := \|F'(x^k)^{1/2}\Delta x^k\|_2^2\,,\ h_k := \omega\|F'(x^k)^{1/2}\Delta x^k\|_2\,.$$

Moreover, let $x^k + \lambda\Delta x^k \in D$ *for* $0 \le \lambda \le \lambda^k_{\max}$ *with*

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

Then

$$f(x^k+\lambda\Delta x^k) \le f(x^k) - t_k(\lambda)\epsilon_k\,, \tag{3.80}$$

where

$$t_k(\lambda) = \lambda - \tfrac{1}{2}\lambda^2 - \tfrac{1}{6}\lambda^3 h_k\,. \tag{3.81}$$

The optimal choice of damping factor is

$$\overline{\lambda}_k = \frac{2}{1+\sqrt{1+2h_k}} \le 1\,. \tag{3.82}$$

Proof. Dropping the iteration index k, we apply the usual mean value theorem to obtain

$$f(x+\lambda\Delta x) - f(x) = \\ -\lambda\epsilon + \tfrac{1}{2}\lambda^2\epsilon + \lambda^2 \int\limits_{s=0}^{1}\int\limits_{t=0}^{1} s\left\langle \Delta x, \left(F'(x+st\lambda\Delta x) - F'(x)\right)\Delta x\right\rangle dtds\,.$$

Upon recalling the Lipschitz condition (3.79), the Cauchy-Schwarz inequality yields

$$\begin{aligned} &f(x+\lambda\Delta x) - f(x) + \left(\lambda - \tfrac{1}{2}\lambda^2\right)\epsilon \\ &\quad\le \lambda^3 \int\limits_{s=0}^{1}\int\limits_{t=0}^{1} s^2 t\|F'(x)^{1/2}\Delta x\|^3 dtds = \tfrac{1}{6}\lambda^3 h\cdot\epsilon\,, \end{aligned} \tag{3.83}$$

which confirms (3.80) and the cubic parabola (3.81). Maximization of t_k by $t_k' = 0$ and solving the arising quadratic equation then yields $\overline{\lambda}_k$ as in (3.84). Moreover, by observing that

$$t_k = \lambda\left(1 - \tfrac{1}{2}\lambda - \tfrac{1}{6}\lambda^2 h_k\right) = 0$$

has only one *positive* root $\lambda^k_{\max}$, the remaining statements are readily verified. □

From these *local* results, we may easily proceed to obtain the following *global* convergence theorem.

Theorem 3.24 *General assumptions as before. Let the path-connected component of the level set* $\mathcal{L}_0 := \{x \in D \mid f(x) \leq f(x^0)\}$ *be compact. Let* $F'(x) = f''(x)$ *be symmetric positive definite for all* $x \in \mathcal{L}_0$*. Then the damped Newton iteration* $(k = 0, 1, \ldots)$ *with damping factors in the range*

$$\lambda_k \in [\varepsilon\,,\ \min(1, \lambda_{\max}^k - \varepsilon)]$$

and sufficiently small $\varepsilon > 0$*, which depends on* $\mathcal{L}_0$*, converges to the solution point* x^**.*

Proof. The proof just applies the local reduction results of the preceding Theorem 3.23. The essential remaining task to show is that there is a common minimal reduction factor for all possible arguments $x^k \in \mathcal{L}_0$. For this purpose, just construct a polygonal upper bound for $t_k(\lambda)$ comparable to the polygon in Figure 3.5. We then merely have to select ε such that

$$\varepsilon < \min(\overline{\lambda}_k, \lambda_{\max}^k - \overline{\lambda}_k)$$

for all possible iterates x^k. Omitting the technical details, Figure 3.5 then directly helps to verify that

$$f(x^k + \lambda \Delta x^k) \leq (1 - \gamma\varepsilon) f(x^k)$$

for λ in the above indicated range with some global $\gamma > 0$, which yields the desired strict global reduction of the functional. □

Summarizing, we have thus established the *theoretical optimal damping strategy* (3.82) in terms of the computationally unavailable Kantorovich quantities h_k.

Remark 3.3 It may be worth noting that the above analysis is nicely connected with the *local* Newton methods (i.e., with $\lambda = 1$) as discussed in Section 2.3.1. If we require that

$$\lambda_{\max}^k = \frac{4}{1 + \sqrt{1 + 8h_k/3}} \geq 1,$$

then we arrive at the local contraction condition

$$h_k \leq 3.$$

This is exactly the condition that would have been obtained in the proof of Theorem 2.18, if the requirement $f(x^{k+1}) \leq f(x^k)$ had been made for the ordinary Newton method—just compare (2.94). However, just as in the framework of that section, the condition $h_{k+1} \leq h_k$ also cannot be guaranteed here, so that $\lambda_{\max}^{k+1} \geq 1$ is not assured. In order to assure such a condition, the more stringent assumption $h_k < 2$ as in (2.92) would be required.

3.4.2 Adaptive trust region strategies

Following our algorithmic paradigm from Section 1.2.3, we construct *computational* damping strategies on the basis of the above derived *theoretically optimal* damping strategy. This strategy contains the unavailable Kantorovich quantity h_k, which we want to replace by some computational estimate $[h_k] \leq h_k$ and, consequently, the theoretical damping factor $\overline{\lambda}_k$ defined in (3.82) by some computationally available value

$$[\overline{\lambda}_k] := \frac{2}{1+\sqrt{1+2[h_k]}} \leq 1\,. \tag{3.84}$$

Since $[h_k] \leq h_k$, we have

$$[\overline{\lambda}_k] \geq \overline{\lambda}_k$$

so that both a *prediction strategy* and a *correction strategy* need to be developed.

Bit counting lemma. As already observed in the comparable earlier cases, the efficiency of such strategies depends on the required accuracy of the computational estimate, which we now analyze.

Lemma 3.25 *Standard assumptions and notation of this section. Let*

$$0 \leq h_k - [h_k] \leq \sigma[h_k] \quad \text{for some} \quad \sigma < 1\,. \tag{3.85}$$

Then, for $\lambda = [\overline{\lambda}_k]$, the following functional decrease is guaranteed:

$$f(x^k + \lambda \Delta x^k) \leq f(x^k) - \tfrac{1}{6}\lambda(\lambda+2)\epsilon_k\,. \tag{3.86}$$

Proof. With $h_k \leq (1+\sigma)[h_k]$ and (3.83) we have (dropping the index k)

$$f(x+\lambda\Delta x) - f(x) \leq -t_k(\lambda)\epsilon_k = \left(-\lambda + \tfrac{1}{2}\lambda^2 + \tfrac{1}{6}\lambda^3 h_k\right)\epsilon_k$$

$$\leq \left(-\lambda + \tfrac{1}{2}\lambda^2 + \tfrac{1}{6}\lambda^3(1+\sigma)[h_k]\right)\epsilon_k\,.$$

At this point, recall that $\overline{\lambda}_k$ is a root of $t_k' = 0$ so that $[\overline{\lambda}_k]$ is a root of

$$1 - \lambda - \tfrac{1}{2}\lambda^2[h_k] = 0\,.$$

Insertion of the above quadratic term into the estimate then yields

$$f(x+\lambda\Delta x) - f(x) \leq \left(-\lambda + \tfrac{1}{2}\lambda^2 + \tfrac{1}{3}\lambda(1+\sigma)(1-\lambda)\right)\epsilon_k\,. \tag{3.87}$$

Upon using $\sigma < 1$ (3.86) is confirmed. □

The above *functional monotonicity test* (3.86) is suggested for use in actual computation. If we further impose $\sigma = 1/2$ in (3.85), i.e., if we require at least

one exact binary digit in the Kantorovich quantity estimate, then (3.87) leads to the *restricted functional monotonicity test*

$$f(x^k + \lambda \Delta x^k) - f(x^k) \leq -\tfrac{1}{2}\lambda \epsilon_k \,.$$

We are now ready to discuss specific computational estimates $[h_k]$ of the Kantorovich quantities h_k. Careful examination shows that we have three basic cheap options. From (3.83) we have the *third* order bound

$$E_3(\lambda) := f(x^k + \lambda \Delta x^k) - f(x^k) + \lambda \left(1 - \tfrac{1}{2}\lambda\right) \epsilon_k \leq \tfrac{1}{6}\lambda^3 h_k \epsilon_k \,,$$

which, in turn, naturally inspires the computational estimate

$$[h_k] := \frac{6|E_3(\lambda)|}{\lambda^3 \epsilon_k} \leq h_k \,.$$

If $E_3(\lambda) < 0$, this means that the Newton method performs locally better than for the mere quadratic model of f (equivalent to $h_k = 0$). Therefore, we decide to set

$$[\overline{\lambda}_k] = 1 \,, \quad \text{if } E_3(\lambda) < 0 \,.$$

On the level of the first derivative we have the *second* order bound

$$E_2(\lambda) := \left\langle \Delta x^k \,,\, F(x^k + \lambda \Delta x^k) - (1 - \lambda) F(x^k) \right\rangle \leq \tfrac{1}{2}\lambda^2 h_k \epsilon_k \,,$$

which inspires the associated estimate

$$[h_k] := \frac{2|E_2(\lambda)|}{\lambda^2 \epsilon_k} \leq h_k \,.$$

On the second derivative level we may derive the *first* order bound

$$E_1(\lambda) := \left\langle \Delta x^k \,,\, \left(F'(x^k + \lambda \Delta x^k) - F'(x^k)\right) \Delta x^k \right\rangle \leq \lambda h_k \epsilon_k \,,$$

which leads to the associated estimate

$$[h_k] := \frac{|E_1(\lambda)|}{\lambda \epsilon_k} \leq h_k \,.$$

Cancellation of leading digits in the terms E_i, $i = 1, 2, 3$ should be carefully monitored—see Figure 3.12, where a snapshot at some iterate in a not further specified illustrative example is taken. Even though the third order expression is the most sensitive, it is also the most attractive one from the point of view of simplicity. Hence, one should first try E_3 and monitor rounding errors carefully.

In principle, any of the above three estimates can be inserted into (3.84) for $[\overline{\lambda}_k]$ requiring at least one trial value of λ (or, respectively, x^{k+1}). We have therefore only designed a possible *correction strategy*

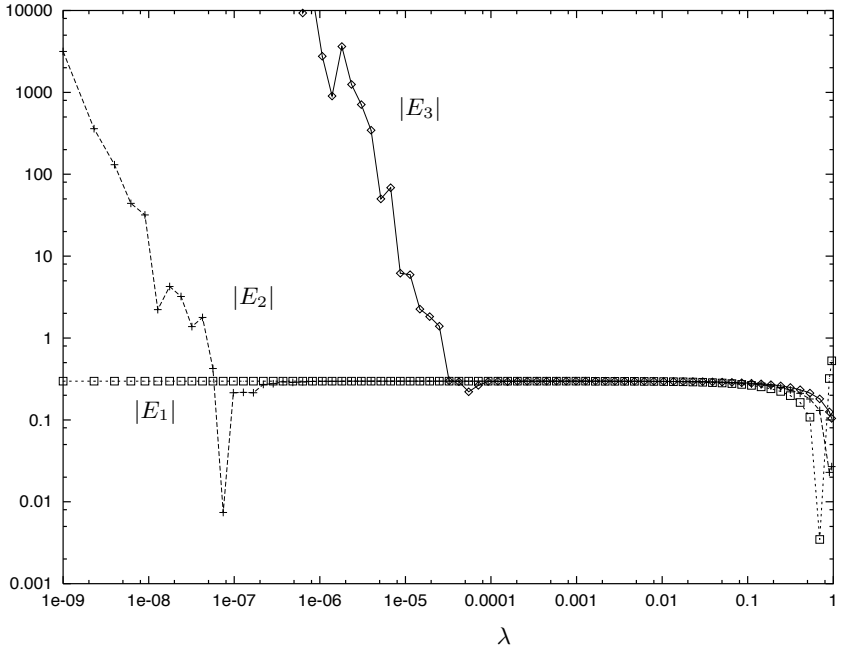

Fig. 3.12. Computational Kantorovich estimates $[h_k]$**:** cancellation of leading digits in $|E_3|$, $|E_2|$, $|E_1|$, respectively.

$$\lambda_k^{i+1} := \frac{2}{1+\sqrt{1+2[h_k(\lambda)]}}\Big|_{\lambda=\lambda_k^i} . \tag{3.88}$$

In order to construct a theoretically backed initial estimate λ_k^0, we may recall that $h_{k+1} = \Theta_k h_k$, where

$$\Theta_k := \frac{\|F'(x^{k+1})^{-\frac{1}{2}} F(x^{k+1})\|_2}{\|F'(x^k)^{-\frac{1}{2}} F(x^k)\|_2} .$$

This relation directly inspires the estimate

$$[h_{k+1}^0] := \Theta_k [h_k^{i*},]$$

wherein i^* indicates the final computable index within estimate (3.88) for the previous iterative step k. Thus we are led to the following *prediction strategy* for $k \geq 0$:

$$\lambda_{k+1}^0 := \frac{2}{1+\sqrt{1+2[h_{k+1}^0]}} \leq 1 . \tag{3.89}$$

As in the earlier discussed approaches, the starting value λ_0^0 needs to be set ad hoc—say, as $\lambda_0^0 = 1$ for 'mildly nonlinear' problems and as $\lambda_0^0 = \lambda_{\min} \ll 1$ for 'highly nonlinear' problems.

3.4.3 Inexact Newton-PCG method

On the basis of the above results for the exact Newton iteration, we may directly proceed to obtain comparable results for the *inexact Newton iteration with damping* ($k = 0, 1, \ldots$, dropping the inner iteration index i)

$$F'(x^k)\left(\delta x^k - \Delta x^k\right) = r^k, \qquad x^{k+1} = x^k + \lambda_k \delta x^k, \qquad \lambda_k \in]0, 1].$$

The inner PCG iteration is formally represented by the introduction of the inner residuals r^k, which are known to satisfy the *Galerkin condition* (compare Section 1.4).

$$\langle \delta x^k, r^k \rangle = 0\,. \tag{3.90}$$

The relative PCG error is denoted by

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

We start the inner iteration with $\delta x_0^k = 0$ so that (1.26) can be applied.

Convergence analysis. With this specification, we immediately verify the following result.

Theorem 3.26 *The statements of Theorem* 3.23 *hold for the inexact Newton-*PCG *method as well, if only the exact Newton corrections Δx^k are replaced by the inexact Newton corrections δx^k and the quantities ϵ_k, h_k are replaced by*

$$\begin{aligned} \epsilon_k^\delta &:= \|F'(x^k)^{1/2}\delta x^k\|^2 = \frac{\epsilon_k}{1+\delta_k^2}\,, \\ h_k^\delta &:= \omega\|F'(x^k)^{1/2}\delta x^k\| = \frac{h_k}{\sqrt{1+\delta_k^2}}\,. \end{aligned}$$

Proof. Dropping the iteration index k, the first line of the proof of Theorem 3.23 may be rewritten as

$$\begin{aligned} &f(x+\lambda\delta x) - f(x) = \\ &-\lambda\epsilon^\delta + \tfrac{1}{2}\lambda^2\epsilon^\delta + \lambda^2 \int_{s=0}^{1} s \int_{t=0}^{1} \left\langle \delta x, \left(F'(x+st\lambda\delta x) - F'(x)\right)\delta x \right\rangle dtds + \langle \delta x, r\rangle\,, \end{aligned}$$

wherein the last right hand term vanishes due to the Galerkin condition (3.90) so that merely the replacement of Δx by δx needs to be performed. □

With these local results established, we are now ready to formulate the associated global convergence theorem.

Theorem 3.27 *General assumptions as Theorem* 3.23 *or Theorem* 3.26*, respectively (in the latter case* δ_k *is formally assumed to be bounded). Let the level set* $\mathcal{L}_0 := \{x \in D \mid f(x) \leq f(x^0)\}$ *be closed and bounded. Let* $F'(x) = f''(x)$ *be symmetric strongly positive for all* $x \in \mathcal{L}_0$*. Then the damped (inexact) Newton iteration (for* $k = 0, 1, \ldots$*) with damping factors in the range*

$$\lambda_k \in [\varepsilon\,,\ \min(1, \lambda_{\max}^k - \varepsilon)]$$

and sufficiently small $\varepsilon > 0$ *depending on* $\mathcal{L}_0$ *converges to the solution point* x^**.*

Proof. The proof just applies the local reduction results of the preceding Theorem 3.23 or Theorem 3.26. The essential remaining task to show is that there is a common minimal reduction factor for all possible arguments $x^k \in \mathcal{L}_0$. For this purpose, we simply construct a polygonal upper bound for $t_k(\lambda)$ such that (omitting technical details)

$$f(x^k + \lambda \Delta x^k) \leq f(x^k) - \tfrac{1}{2}\varepsilon\epsilon_k$$

for λ in the above indicated range and all possible iterates x^k with some

$$\varepsilon < \min(\overline{\lambda}_k, \lambda_{\max}^k - \overline{\lambda}_k)\,.$$

This implies a strict reduction of the functional in each iterative step as long as $\epsilon_k > 0$ and therefore global convergence in the compact level set $\mathcal{L}_0$ towards the minimum point x^* with $\epsilon_* = 0$. □

Adaptive trust region strategy. The strategy as worked out in Section 3.4.2 can be directly copied, just replacing Δx^k by δx^k, ϵ_k by ϵ_k^δ, and h_k by h_k^δ; details are left to Exercise 3.5. In actual computation the orthogonality condition (3.90) may be perturbed by rounding errors from scalar products in `PCG`. Therefore the terms E_3 and E_2 should be evaluated in the special form

$$E_3(\lambda) := f(x^k + \lambda\delta x^k) - f(x^k) - \lambda\langle F(x^k), \delta x^k\rangle - \tfrac{1}{2}\lambda^2\epsilon_k^\delta$$

and

$$E_2(\lambda) := \langle \delta x^k, F(x^k + \lambda\delta x^k) - F(x^k)\rangle - \lambda\epsilon_k^\delta$$

with the local energy computed as $\epsilon_k^\delta = \langle \delta x^k, F'(x^k)\delta x^k\rangle$.

As for the choice of accuracies δ_k arising from the inner `PCG` iteration, we once again require

$$\delta_k \leq \tfrac{1}{4}$$

in the damping phase ($\lambda < 1$) and the appropriate settings as worked out in Section 2.3.3 for the local Newton-`PCG` ($\lambda = 1$)—merging either into the *linear* or the *quadratic* convergence mode.

The affine conjugate inexact Newton method with adaptive trust region method and corresponding matching of the inner `PCG` iteration and outer iteration is realized in the code `GIANT-PCG`.

Bibliographical Note. The first affine conjugate global Newton method has been derived and implemented by P. Deuflhard and M. Weiser [85], there even in the more complicated context of an adaptive multilevel finite element method for nonlinear elliptic PDEs—compare Section 8.3. The strategy presented here is just a finite dimensional analog of the strategy worked out there.

Exercises

Exercise 3.1 *Multipoint homotopy.* Let $F(x) = 0$ denote a system of nonlinear equations to be solved and x^* its solution. Let $x^0, \ldots, x^p$ be a sequence of iterates produced by some iterative process. Consider the homotopy ($\lambda \in \mathbb{R}^1$)

$$H_p(x, \lambda) := F(x) - \sum_{k=0}^{p-1} L_k(\lambda) F(x^k)\,, \quad p \geq 1$$

with L_k being the fundamental Lagrangian polynomials defined over a set of nodes $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_p = 1$.

a) Show that, under the standard assumptions of the implicit function theorem, there exists a homotopy path $\overline{x}(\lambda)$ such that

$$\overline{x}(\lambda_k) = x^k\,, \quad k = 0, \ldots, p-1\,, \quad \overline{x}(1) = x^*\,.$$

Derive the associated Davidenko differential equation.

b) Construct an iterative process for successively increasing $p = 1, 2, \ldots$ by appropriate discretization. What would be a reasonable assignment of the nodes $\lambda_1, \ldots, \lambda_{p-1}$? Consider the local convergence properties of such a process.

c) Write a program for $p = 1, 2$ and experiment with λ_1 over a set of test problems.

Exercise 3.2 An obstacle on the way toward a proof of global convergence for error oriented global Newton methods, controlled only by the natural monotonicity test

$$\|F'(x^k)^{-1} F(x^{k+1})\| \leq \|F'(x^k)^{-1} F(x^k)\|\,,$$

is the fact that a desirable property like

$$\|F'(x^{k+1})^{-1} F(x^{k+1})\| \leq \|F'(x^k)^{-1} F(x^k)\| \tag{3.91}$$

does *not* hold.

a) For $x^{k+1} = x^k + \lambda \Delta x^k$, upon applying Theorem 3.12, verify that

$$\|F'(x^{k+1})^{-1} F(x^{k+1})\| \leq \frac{1 - \lambda + \frac{1}{2}\lambda^2 h_k}{1 - \lambda h_k} \|F'(x^k)^{-1} F(x^k)\| .$$

b) Show that only under the Kantorovich-type assumption $h_k < 1$ the reduction (3.91) can be guaranteed for certain $\lambda > 0$.

Exercise 3.3 *2–cycle example* [10]. Consider a system $F(x) = 0$ of two equations in two unknowns. Let

$$F(0) = -\tfrac{1}{10} \begin{pmatrix} 4\sqrt{3} - 3 \\ -4\sqrt{3} - 3 \end{pmatrix} =: -a , \quad F'(0) = I ,$$

$$F(a) = \tfrac{1}{5} \begin{pmatrix} 4 \\ -3 \end{pmatrix} , \quad F'(a) = \begin{pmatrix} 17\sqrt{3} & -1/\sqrt{3} \\ 1 & 1 \end{pmatrix} .$$

Starting a Newton method with $x^0 := 0$, we want to verify the occurrence of a 2–cycle, if only natural monotonicity is required.

a) Show that in the first Newton step $\lambda_0 = 1$ is acceptable, since the natural monotonicity criterion is passed, which leads to $x^1 = a$.

b) Show that in the second Newton step $\lambda_1 = 1$ is acceptable yielding $x^2 = x^0$.

Exercise 3.4 *Avoidance of 2–cycles* [33]. We study the possible occurrence of 2–cycles for a damped Newton iteration with *natural* monotonicity test and damping factor λ. A special example is given in Exercise 3.3. By definition, such a 2–cycle is characterized by the inequalities

$$\begin{aligned} \|F'(x^k)^{-1} F(x^{k+1})\| &\leq \|F'(x^k)^{-1} F(x^k)\| , \\ \|F'(x^{k+1})^{-1} F(x^{k+2})\| &\leq \|F'(x^{k+1})^{-1} F(x^{k+1})\| \end{aligned}$$

with $x^{k+2} = x^k$.

a) Upon applying Theorem 3.12 verify that

$$\|F'(x^{k+1})^{-1} F(x^{k+1})\| \leq \left(1 - \lambda + \tfrac{1}{2}\lambda^2 h_k \frac{1 + \lambda h_k}{1 - \lambda h_k}\right) \|F'(x^{k+1})^{-1} F(x^k)\| .$$

b) Show that under the restriction

$$\lambda h_k \leq \eta < \tfrac{1}{2}(\sqrt{17} - 3)$$

the occurrence of 2–cycles is impossible.

c) By a proper adaptation of the bit counting Lemma 3.16, modify the damping strategy (3.45) and the restricted monotonicity test (3.47) such that 2–cycles are also algorithmically excluded.

Exercise 3.5 Consider an inexact Newton method for convex optimization, where the inner iteration does *not* satisfy the Galerkin condition (3.90). The aim here is to prove an affine conjugate global convergence theorem as a substitute of Theorem 3.26. Define

$$\sigma_k = -\frac{\langle F(x^k), \delta x^k \rangle}{\epsilon_k^\delta} \, .$$

(a) Show that one obtains the upper bound

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

so that $\sigma_k > 0$ is required to assert functional decrease.

(b) On the basis of the optimal damping factor

$$\overline{\lambda}_k = \frac{2\sigma_k}{1 + \sqrt{1 + 2\sigma_k h_k^\delta}} \leq 1$$

prove a global convergence theorem.

How can this theorem also be exploited for the design of an adaptive inexact Newton method?

Exercise 3.6 Usual `GMRES` codes require the user to formulate the linear equation $Ay = b$ as $A\Delta y = r(y_0)$ with $\Delta y = y - y^0$ and $r(y_0) = b - Ay_0$ so that $\Delta y^0 = 0$. Reformulate the linear system

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

to be solved by the `GMRES` iteration for an initial guess

$$\widetilde{\delta x_0}^{k+1} = (1 - \lambda)\delta x^k \, .$$