

Chapter 2

Equations and Unconstrained Optimization

In this chapter, we start our discussion of Newton-type methods, which are based on the fundamental principle of linear/quadratic approximation of the problem data (or of some part of the problem data). The underlying idea of Newtonian methods is extremely important, as it serves as a foundation for numerous computationally efficient algorithms for optimization and variational problems.

We start with discussing the basic Newton method for nonlinear equations and unconstrained optimization. High rate of convergence of this scheme is due to using information about the derivatives of the problem data (first derivatives of the operator in the case of nonlinear equations, second derivatives of the objective function in the case of optimization). Thus, each iteration of this basic process should be regarded as relatively expensive. However, one of the main messages of this chapter is that various kinds of inexactness, introduced *intentionally* into the basic Newton scheme, can serve to reduce the cost of the iteration while keeping the convergence rate still high enough. Combined with globalization techniques, such modifications lead to truly practical Newtonian methods for unconstrained optimization problems, the most important of which belong to the quasi-Newton class.

As much of the material covered in this chapter can be considered nowadays quite standard (e.g., linesearch quasi-Newton methods, trust-region methods, etc.), we sometimes mention only the main principles behind certain techniques without going into full details. On the other hand, the general perturbed Newton framework is analyzed very thoroughly, as its natural generalization for optimization and variational problems would be one of the main tools for treating various algorithms throughout the book.

2.1 Newton Method

For historical comments regarding the Newton method, we address the reader to [62].

2.1.1 Newton Method for Equations

The classical Newton method is introduced for the equation

$$\Phi(x) = 0, \quad (2.1)$$

where $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is a smooth mapping. Let $x^k \in \mathbf{R}^n$ be the current approximation to a solution of (2.1). Then it is natural to approximate the equation (2.1) near the point x^k by its linearization:

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) = 0. \quad (2.2)$$

The linearized equation (2.2) gives the iteration system of the classical *Newton method*. The idea is transparent — the nonlinear equation (2.1) is replaced by the (computationally much simpler) linear equation (2.2). Iterations of the Newton method for the case when $n = 1$ are illustrated in Fig. 2.1.

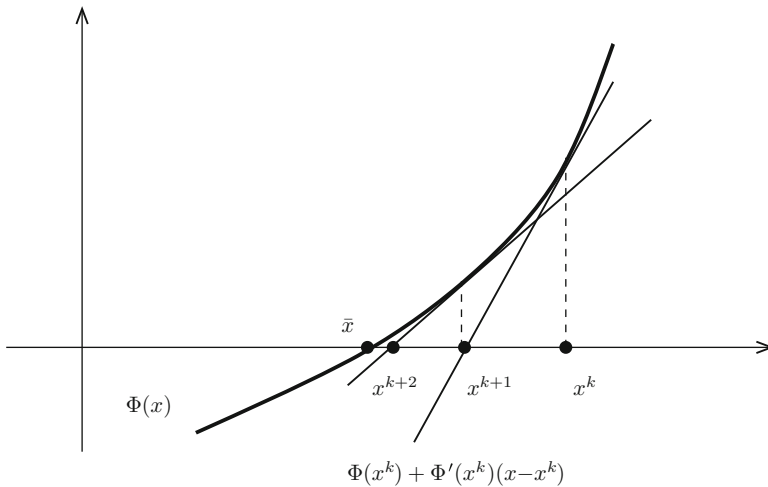


Fig. 2.1 Iterations of the Newton method

Formally, the algorithm is stated as follows.

Algorithm 2.1 Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Phi(x^k) = 0$, stop.
2. Compute $x^{k+1} \in \mathbf{R}^n$ as a solution of (2.2).
3. Increase k by 1 and go to step 1.

Assuming that the Jacobian $\Phi'(x^k)$ is nonsingular, the Newton method is often presented in the form of the explicit iterative scheme

$$x^{k+1} = x^k - (\Phi'(x^k))^{-1}\Phi(x^k), \quad k = 0, 1, \dots, \quad (2.3)$$

with the understanding that an actual implementation of the method need not require computing the complete inverse of the matrix $\Phi'(x^k)$; of interest is only the product $(\Phi'(x^k))^{-1}\Phi(x^k)$.

Under appropriate assumptions, the Newton method is very efficient, which is reflected in the following convergence statements. At the same time, it is clear that in its pure form the method may not converge from points that are not close enough to a solution, even if the latter satisfies all the needed assumptions; see Fig. 2.2 and also Example 2.16 below.

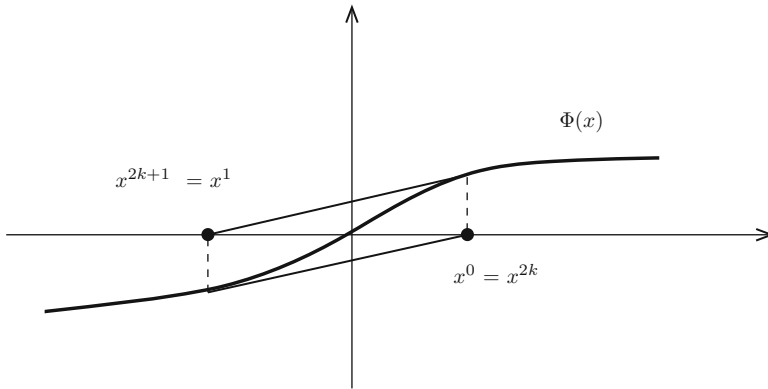


Fig. 2.2 Non-convergence of the Newton method from points far from a solution

The following describes the essential convergence properties of the Newton method.

Theorem 2.2. Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at \bar{x} . Let \bar{x} be a solution of the equation (2.1), and assume that $\Phi'(\bar{x})$ is a nonsingular matrix.

Then the following assertions are valid:

- (a) There exists a neighborhood U of \bar{x} and a function $q(\cdot) : U \rightarrow \mathbf{R}$ such that $\Phi'(x)$ is nonsingular for all $x \in U$,

$$\|x - (\Phi'(x))^{-1}\Phi(x) - \bar{x}\| \leq q(x)\|x - \bar{x}\| \quad \forall x \in U, \quad (2.4)$$

and

$$q(x) \rightarrow 0 \text{ as } x \rightarrow \bar{x}. \quad (2.5)$$

- (b) Any starting point $x^0 \in \mathbf{R}^n$ close enough to \bar{x} uniquely defines a particular iterative sequence of Algorithm 2.1; this sequence converges to \bar{x} , and the rate of convergence is superlinear.
- (c) If the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} , then $q(\cdot)$ can be chosen in such a way that

$$q(x) = O(\|x - \bar{x}\|) \quad (2.6)$$

as $x \rightarrow \bar{x}$, and the rate of convergence is quadratic.

Assertion (a) means that the Newton step from a point close enough to \bar{x} provides a “superlinear decrease” of the distance to \bar{x} , while assertion (c) gives conditions guaranteeing “quadratic decrease” of this distance.

Regarding formal definitions of convergence rates (in particular, superlinear and quadratic), see Sect. A.2.

Proof. According to Lemma A.6, there exist a neighborhood U of \bar{x} and $M > 0$ such that

$$\Phi'(x) \text{ is nonsingular, } \|(\Phi'(x))^{-1}\| \leq M \quad \forall x \in U. \quad (2.7)$$

Employing the mean-value theorem (see Theorem A.10, (a)), we can choose U in such a way that the inclusion $x \in U$ implies

$$\begin{aligned} \|x - (\Phi'(x))^{-1}\Phi(x) - \bar{x}\| &\leq \|(\Phi'(x^k))^{-1}\| \|\Phi(x) - \Phi(\bar{x}) - \Phi'(x)(x - \bar{x})\| \\ &\leq q(x)\|x - \bar{x}\|, \end{aligned} \quad (2.8)$$

where

$$q(x) = M \sup\{\|\Phi'(tx + (1-t)\bar{x}) - \Phi'(x)\| \mid t \in [0, 1]\}. \quad (2.9)$$

It is clear that this $q(\cdot)$ satisfies (2.5), while (2.8) gives (2.4). This completes the proof of assertion (a).

In particular, for $x^k \in U$, the equation (2.2) has the unique solution x^{k+1} given by (2.3). Moreover, from (2.4) and (2.5) it follows that for any $q \in (0, 1)$ there exists $\delta > 0$ such that $B(\bar{x}, \delta) \subset U$, and the inclusion $x^k \in B(\bar{x}, \delta)$ implies

$$\|x^{k+1} - \bar{x}\| \leq q\|x^k - \bar{x}\|.$$

In particular, $x^{k+1} \in B(\bar{x}, \delta)$. It follows that any starting point $x^0 \in B(\bar{x}, \delta)$ uniquely defines a specific iterative sequence $\{x^k\}$ of Algorithm 2.1; this sequence is contained in $B(\bar{x}, \delta)$ and converges to \bar{x} . Moreover, again employing (2.4), we obtain the estimate

$$\|x^{k+1} - \bar{x}\| \leq q(x^k)\|x^k - \bar{x}\| \quad \forall k = 0, 1, \dots, \quad (2.10)$$

which, according to (2.5), implies the superlinear rate of convergence. This completes the proof of assertion (b).

Finally, if the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} with a constant $L > 0$, then, after reducing U if necessary, from (2.9) it follows that the inclusion $x \in U$ implies

$$\begin{aligned} q(x) &\leq M(\sup\{\|\Phi'(\bar{x} + t(x - \bar{x})) - \Phi'(\bar{x})\| \mid t \in [0, 1]\} + \|\Phi'(x) - \Phi'(\bar{x})\|) \\ &\leq 2ML\|x - \bar{x}\|, \end{aligned}$$

which proves (2.6). The quadratic convergence rate now follows from (2.6) and (2.10). This proves (c). \square

The main message of the subsequent discussion in this section is that various kinds of inexactness introduced *intentionally* in the basic Newton scheme may lead to more practical Newton-type methods, with lower computational costs per iteration but convergence rate still high enough. To that end, we consider the following general scheme, which we refer to as the *perturbed Newton method*. For a given $x^k \in \mathbf{R}^n$, the next iterate $x^{k+1} \in \mathbf{R}^n$ satisfies the perturbed version of the iteration system (2.2):

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) + \omega^k = 0. \quad (2.11)$$

Here, $\omega^k \in \mathbf{R}^n$ is a perturbation term, which may have various forms and meanings, may play various roles, and may conform to different sets of assumptions depending on the particular algorithms at hand and on the particular purposes of the analysis. At the moment, we are interested in the following general but simple question: under which assumptions regarding ω^k the local convergence and/or the superlinear rate of convergence of the pure Newton method (2.2) is preserved?

We start with some basic (essentially technical) statements, which do not impose any restrictions on the structure of ω^k . Note that this is an a posteriori kind of analysis: the iterative sequence $\{x^k\}$ is given, and the corresponding sequence $\{\omega^k\}$ is then explicitly defined by (2.11). Thus, in this setting the role of $\{\omega^k\}$ is secondary with respect to $\{x^k\}$. Those technical results would be useful later on for analyzing iterative sequences generated by specific Newton-type schemes.

Lemma 2.3. *Under the assumptions of Theorem 2.2, there exist a neighborhood U of \bar{x} and $M > 0$ such that for any $x^k \in U$ and any $x^{k+1} \in \mathbf{R}^n$ and $\omega^k \in \mathbf{R}^n$ satisfying*

$$\omega^k = -\Phi(x^k) - \Phi'(x^k)(x^{k+1} - x^k), \quad (2.12)$$

it holds that

$$\|x^{k+1} - \bar{x}\| \leq M\omega^k + o(\|x^k - \bar{x}\|) \quad (2.13)$$

as $x^k \rightarrow \bar{x}$. Moreover, if the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} , then the estimate (2.13) can be sharpened as follows:

$$\|x^{k+1} - \bar{x}\| \leq M\omega^k + O(\|x^k - \bar{x}\|^2). \quad (2.14)$$

Proof. By assertion (a) of Theorem 2.2 and by Lemma A.6, there exist a neighborhood U of \bar{x} and $M > 0$ such that (2.7) holds, and

$$x^k - (\Phi'(x^k))^{-1}\Phi(x^k) - \bar{x} = o(\|x^k - \bar{x}\|) \quad (2.15)$$

as $x^k \in U$ tends to \bar{x} . Furthermore, by (2.12),

$$x^{k+1} = x^k - (\Phi'(x^k))^{-1}(\Phi(x^k) + \omega^k).$$

Hence, employing (2.7) and (2.15), we obtain that

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &= \|x^k - (\Phi'(x^k))^{-1}(\Phi(x^k) + \omega^k) - \bar{x}\| \\ &\leq \|(\Phi'(x^k))^{-1}\| \|\omega^k\| + \|x^k - (\Phi'(x^k))^{-1}\Phi(x^k) - \bar{x}\| \\ &\leq M\omega^k + o(\|x^k - \bar{x}\|), \end{aligned}$$

which establishes (2.13).

Finally, if the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} , estimate (2.14) follows by the same argument, but invoking assertion (c) of Theorem 2.2. \square

The next result states a necessary and sufficient condition on the perturbation sequence $\{\omega^k\}$ under which superlinear convergence of $\{x^k\}$ is preserved. Note that convergence itself is not established but assumed here.

Proposition 2.4. *Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at \bar{x} . Let \bar{x} be a solution of the equation (2.1). Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to \bar{x} , and define ω^k according to (2.12) for each $k = 0, 1, \dots$*

If the rate of convergence of $\{x^k\}$ is superlinear, then

$$\omega^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \quad (2.16)$$

as $k \rightarrow \infty$.

Conversely, if $\Phi'(\bar{x})$ is a nonsingular matrix, and (2.16) holds, then the rate of convergence of $\{x^k\}$ is superlinear. Moreover, the rate of convergence is quadratic, provided the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} and

$$\omega^k = O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2) \quad (2.17)$$

as $k \rightarrow \infty$.

Proof. By (2.12) and the mean-value theorem (see Theorem A.10), we obtain that for all k large enough

$$\begin{aligned}
\|\omega^k\| &= \|\Phi(x^k) + \Phi'(x^k)(x^{k+1} - x^k)\| \\
&\leq \|\Phi(x^k) - \Phi(\bar{x}) - \Phi'(x^k)(x^k - \bar{x})\| + \|\Phi'(x^k)\| \|x^{k+1} - \bar{x}\| \\
&\leq \sup\{\|\Phi'(tx^k + (1-t)\bar{x}) - \Phi'(x^k)\| \mid t \in [0, 1]\} \|x^k - \bar{x}\| \\
&\quad + O(\|x^{k+1} - \bar{x}\|) \\
&= o(\|x^k - \bar{x}\|) + O(\|x^{k+1} - \bar{x}\|)
\end{aligned}$$

as $k \rightarrow \infty$. If the sequence $\{x^k\}$ converges to \bar{x} superlinearly, the above implies that $\omega^k = o(\|x^k - \bar{x}\|)$, which in turn implies (2.16).

Suppose now that (2.16) holds. From Lemma 2.3 it then follows that

$$x^{k+1} - \bar{x} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) = o(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|),$$

i.e., there exists a sequence $\{t_k\} \subset \mathbf{R}$ such that $t_k \rightarrow 0$ and

$$\|x^{k+1} - \bar{x}\| \leq t_k(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|).$$

for all k large enough. This implies that

$$(1 - t_k)\|x^{k+1} - \bar{x}\| \leq t_k\|x^k - \bar{x}\|.$$

Hence, for all k large enough

$$\|x^{k+1} - \bar{x}\| \leq \frac{t_k}{1 - t_k} \|x^k - \bar{x}\|,$$

i.e.,

$$x^{k+1} - \bar{x} = o(\|x^k - \bar{x}\|)$$

as $k \rightarrow \infty$, which gives the superlinear convergence rate.

Finally, if the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} , from Lemma 2.3 it follows that (2.17) implies the estimate

$$x^{k+1} - \bar{x} = O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2) = O(\|x^{k+1} - \bar{x}\|^2 + \|x^k - \bar{x}\|^2)$$

as $k \rightarrow \infty$, which means that there exists $M > 0$ such that

$$\|x^{k+1} - \bar{x}\| \leq M(\|x^{k+1} - \bar{x}\|^2 + \|x^k - \bar{x}\|^2) \quad (2.18)$$

for all k large enough. Since $\{x^k\}$ converges to \bar{x} , for any fixed $\varepsilon \in (0, 1)$ it holds that $M\|x^{k+1} - \bar{x}\| \leq 1 - \varepsilon$ for all k large enough. Then from (2.18) we derive

$$(1 - M\|x^{k+1} - \bar{x}\|)\|x^{k+1} - \bar{x}\| \leq M\|x^k - \bar{x}\|^2,$$

and hence, for all k large enough

$$\|x^{k+1} - \bar{x}\| \leq \frac{M}{1 - M\|x^{k+1} - \bar{x}\|} \|x^k - \bar{x}\|^2 \leq \frac{M}{\varepsilon} \|x^k - \bar{x}\|^2,$$

which gives the quadratic convergence rate. \square

Remark 2.5. If $\{x^k\}$ converges to \bar{x} superlinearly, the estimate (2.16) is, in fact, equivalent to either of the following two (generally stronger) estimates:

$$\omega^k = o(\|x^{k+1} - x^k\|), \quad (2.19)$$

or

$$\omega^k = o(\|x^k - \bar{x}\|). \quad (2.20)$$

Indeed, by (2.16) and the superlinear convergence rate of $\{x^k\}$ to \bar{x} , there exist sequences $\{t_k\} \subset \mathbf{R}$ and $\{\tau_k\} \subset \mathbf{R}$ such that $t_k \rightarrow 0$, $\tau_k \rightarrow 0$, and

$$\|\omega^k\| \leq t_k(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \quad (2.21)$$

$$\|x^{k+1} - \bar{x}\| \leq \tau_k \|x^k - \bar{x}\| \quad (2.22)$$

for all k . Then

$$\|x^k - \bar{x}\| \leq \|x^{k+1} - x^k\| + \|x^{k+1} - \bar{x}\| \leq \|x^{k+1} - x^k\| + \tau_k \|x^k - \bar{x}\|,$$

implying that

$$\|x^k - \bar{x}\| \leq \frac{1}{1 - \tau_k} \|x^{k+1} - x^k\|$$

for all k large enough. Combining this with (2.21) we then obtain that

$$\|\omega^k\| \leq t_k \left(1 + \frac{1}{1 - \tau_k}\right) \|x^{k+1} - x^k\| = t_k \frac{2 - \tau_k}{1 - \tau_k} \|x^{k+1} - x^k\|$$

for all k large enough, which gives (2.19). Furthermore, from (2.21) and (2.22) we directly derive that

$$\begin{aligned} \|\omega^k\| &\leq t_k(2\|x^k - \bar{x}\| + \|x^{k+1} - \bar{x}\|) \\ &\leq t_k(2\|x^k - \bar{x}\| + \tau_k\|x^k - \bar{x}\|) \\ &\leq t_k(2 + \tau_k)\|x^k - \bar{x}\| \end{aligned}$$

for all k , which gives (2.20).

The next result provides a sufficient condition on the perturbation sequence $\{\omega^k\}$ for preserving local convergence of $\{x^k\}$.

Proposition 2.6. *Under the assumptions of Theorem 2.2, fix any norm $\|\cdot\|_*$ in \mathbf{R}^n , any $q_1, q_2 \geq 0$ such that $2q_1 + q_2 < 1$, and any $\varepsilon \in (0, 1 - 2q_1 - q_2)$.*

Then there exists $\delta > 0$ such that for any sequence $\{x^k\} \subset \mathbf{R}^n$ the following assertions are valid:

(a) *If for some $k = 0, 1, \dots$, it holds that $x^k \in B(\bar{x}, \delta)$, and ω^k defined according to (2.12) satisfies the condition*

$$\|(\Phi'(\bar{x}))^{-1}\omega^k\|_* \leq q_1\|x^{k+1} - x^k\|_* + q_2\|x^k - \bar{x}\|_* \quad \forall k = 0, 1, \dots, \quad (2.23)$$

then

$$\|x^{k+1} - \bar{x}\|_* \leq \frac{q_1 + q_2 + \varepsilon}{1 - q_1} \|x^k - \bar{x}\|_* \quad (2.24)$$

and, in particular, $x^{k+1} \in B(\bar{x}, \delta)$.

(b) If $x^0 \in B(\bar{x}, \delta)$ and (2.23) is satisfied for all $k = 0, 1, \dots$, then $\{x^k\}$ converges to \bar{x} , and the rate of convergence is (at least) linear. More precisely, either $x^k = \bar{x}$ for all k large enough, or

$$\limsup_{k \rightarrow \infty} \frac{\|x^{k+1} - \bar{x}\|_*}{\|x^k - \bar{x}\|_*} \leq \frac{q_1 + q_2}{1 - q_1}. \quad (2.25)$$

Proof. By (2.12) and (2.23), employing assertion (a) of Theorem 2.2 and the equivalence of norms in \mathbf{R}^n , we obtain that for any $\varepsilon \in (0, 1 - 2q_1 - q_2)$ there exists $\delta > 0$ such that for any $x^k \in B(\bar{x}, \delta)$ it holds that

$$\begin{aligned} \|x^{k+1} - \bar{x}\|_* &= \|x^k - (\Phi'(x^k))^{-1}(\Phi(x^k) + \omega^k) - \bar{x}\|_* \\ &\leq \|(\Phi'(x^k))^{-1}\omega^k\|_* + \|x^k - (\Phi'(x^k))^{-1}\Phi(x^k) - \bar{x}\|_* \\ &\leq q_1 \|x^{k+1} - x^k\|_* + q_2 \|x^k - \bar{x}\|_* + o(\|x^k - \bar{x}\|_*) \\ &\leq q_1 (\|x^{k+1} - \bar{x}\|_* + \|x^k - \bar{x}\|_*) + q_2 \|x^k - \bar{x}\|_* + \varepsilon \|x^k - \bar{x}\|_* \\ &\leq q_1 \|x^{k+1} - \bar{x}\|_* + (q_1 + q_2 + \varepsilon) \|x^k - \bar{x}\|_*. \end{aligned}$$

This implies (2.24). Since $(q_1 + q_2 + \varepsilon)/(1 - q_1) \in (0, 1)$, (2.24) implies that $x^{k+1} \in B(\bar{x}, \delta)$. This proves assertion (a).

Furthermore, the inclusion $x^0 \in B(\bar{x}, \delta)$ and assertion (a) imply that the entire sequence $\{x^k\}$ is contained in $B(\bar{x}, \delta)$, and (2.24) shows convergence of this sequence to \bar{x} at a linear rate. Moreover, since ε can be taken arbitrarily small at a price of reducing δ , and since $\{x^k\}$ converges to \bar{x} (hence, the tail of $\{x^k\}$ is contained in $B(\bar{x}, \delta)$ no matter how small δ is), relation (2.24) implies that either $x^k = \bar{x}$ for all k large enough, or (2.25) holds. This proves assertion (b). \square

Conditions (2.16) and/or (2.23) are not “practical,” because they involve the unknown solution \bar{x} and/or the next iterate x^{k+1} , which is usually computed *after* the perturbation term is settled. Propositions 2.4 and 2.6 are merely technical tools intended for the analysis of some specific algorithms fitting the perturbed Newton method framework.

We start with the class of the so-called *truncated Newton methods*, which were first systematically studied in [55], and which are particular instances of perturbed Newton methods with the perturbation terms satisfying the condition

$$\|\omega^k\| \leq \theta_k \|\Phi(x^k)\|, \quad k = 0, 1, \dots \quad (2.26)$$

Here, $\{\theta_k\}$ is a sequence of nonnegative numbers, called *forcing sequence*, which can be either pre-fixed or computed in the course of iterations. The idea of truncated Newton methods consists of solving the iteration system (2.2)

not exactly, but up to the accuracy defined by the right-hand side of the inequality in (2.26). Note that (2.26) is totally practical as an approximation criterion for solving the Newton method iteration system (2.2), as it does not involve any unknown objects (such as the solution \bar{x} and/or x^{k+1} , as in the technical conditions (2.16) and (2.23)). Thus, (2.26) can be easily checked in the course of solving (2.2). The most popular strategy is to apply to the linear equation (2.2) some iterative method (e.g., the conjugate gradient method for minimizing its squared residual), and to stop this inner process when (2.26) will be satisfied for ω^k defined in (2.11) with x being the current iterate of the inner process. Once (2.26) is satisfied, x in (2.11) is declared to be the next iterate x^{k+1} . Supplied with a rule for computing the forcing sequence $\{\theta_k\}$ and a choice of an inner iterative scheme, this algorithmic framework results in a specific truncated Newton method.

Employing Propositions 2.4 and 2.6, we obtain the following properties.

Theorem 2.7. *Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at \bar{x} . Let \bar{x} be a solution the equation (2.1). Let $\{x^k\} \subset \mathbf{R}^n$ be a sequence convergent to \bar{x} , and let ω^k be defined according to (2.12) for each $k = 0, 1, \dots$*

If the rate of convergence of $\{x^k\}$ is superlinear, then there exists a sequence $\{\theta_k\} \subset \mathbf{R}$ satisfying condition (2.26), and such that $\theta_k \rightarrow 0$.

Conversely, if $\Phi'(\bar{x})$ is a nonsingular matrix and there exists a sequence $\{\theta_k\} \subset \mathbf{R}$ satisfying condition (2.26) and such that $\theta_k \rightarrow 0$, then the rate of convergence of $\{x^k\}$ is superlinear. Moreover, the rate of convergence is quadratic, provided the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} and

$$\theta_k = O(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \quad (2.27)$$

as $k \rightarrow \infty$.

Proof. To prove the first assertion, observe that by the error bound presented in Proposition 1.32, it holds that

$$x^k - \bar{x} = O(\|\Phi(x^k)\|) \quad (2.28)$$

as $k \rightarrow \infty$. By Proposition 2.4 and Remark 2.5, superlinear convergence rate of $\{x^k\}$ implies (2.20). Thus, by (2.28), we have that

$$\omega^k = o(\|x^k - \bar{x}\|) = o(\|\Phi(x^k)\|),$$

which means precisely the existence of a sequence $\{\theta_k\}$ with the needed properties.

The second assertion follows from Proposition 2.4. Indeed,

$$\Phi(x^k) = \Phi(\bar{x}) + \Phi'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) = O(\|x^k - \bar{x}\|) \quad (2.29)$$

as $k \rightarrow \infty$, and therefore, (2.26) with $\theta_k \rightarrow 0$ evidently implies (2.20) (and, hence, (2.16)).

Finally, if the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} , and (2.27) holds, quadratic convergence follows by the last assertion of Proposition 2.4, because in this case, taking into account (2.29), we derive that

$$\begin{aligned}\omega^k &= O((\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\|\Phi(x^k)\|) \\ &= O((\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\|x^k - \bar{x}\|) \\ &= O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2)\end{aligned}$$

as $k \rightarrow \infty$. \square

In the previous result, convergence of $\{x^k\}$ was assumed. But to pass to a constructive result, also establishing convergence, is now easy.

Theorem 2.8. *Suppose that the assumptions of Theorem 2.2 hold, and let $\theta \in (0, 1)$ be arbitrary.*

Then for any $x^0 \in \mathbf{R}^n$ close enough to \bar{x} and any sequences $\{x^k\} \subset \mathbf{R}^n$, $\{\omega^k\} \subset \mathbf{R}^n$ and $\{\theta^k\} \subset [0, \theta]$ satisfying (2.12) and (2.26) for all $k = 0, 1, \dots$, it holds that $\{x^k\}$ converges to \bar{x} and the rate of convergence is (at least) linear. Moreover, the rate of convergence is superlinear provided $\theta_k \rightarrow 0$. The rate of convergence is quadratic, provided the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} and

$$\theta_k = O(\|\Phi(x^k)\|) \quad (2.30)$$

as $k \rightarrow \infty$.

Proof. Define the following norm in \mathbf{R}^n : $\|x\|_* = \|\Phi'(\bar{x})x\|$, $x \in \mathbf{R}^n$ (as $\Phi'(\bar{x})$ is nonsingular, this is indeed a norm). Then, employing (2.26) and the equivalence of norms in \mathbf{R}^n , we obtain that for any $\varepsilon \in (0, 1 - \theta)$ there exists $\delta > 0$ such that for any $x^k \in B(\bar{x}, \delta)$ it holds that

$$\begin{aligned}\|(\Phi'(\bar{x}))^{-1}\omega^k\|_* &= \|\omega^k\| \\ &\leq \theta\|\Phi(x^k)\| \\ &= \theta\|\Phi(\bar{x}) + \Phi'(\bar{x})(x^k - \bar{x})\| + o(\|x^k - \bar{x}\|) \\ &= \theta\|x^k - \bar{x}\|_* + o(\|x^k - \bar{x}\|_*) \\ &\leq (\theta + \varepsilon)\|x^k - \bar{x}\|_*,\end{aligned}$$

which is (2.23) with $q = \theta + \varepsilon$. By assertion (a) of Proposition 2.6, this implies the inclusion $x^{k+1} \in B(\bar{x}, \delta)$, provided δ is chosen small enough. Thus, the inclusion $x^0 \in B(\bar{x}, \delta)$ implies that the entire sequence $\{x^k\}$ is contained in $B(\bar{x}, \delta)$, and that (2.23) holds for all $k = 0, 1, \dots$. It remains to apply assertion (b) of Proposition 2.6. Superlinear rate of convergence when $\theta_k \rightarrow 0$, and quadratic rate when the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} and (2.30) holds, follow from Theorem 2.7, taking into account (2.29). \square

Some specific implementations of truncated Newton methods and related results can be found, e.g., in [208, Chap. 11].

It is interesting to note that the class of *quasi-Newton methods*, which is of great practical importance, can be (theoretically) related to truncated Newton methods, even though the principles behind the two approaches are completely different. Close enough to a solution, a step of any quasi-Newton method is supposed to take the form

$$x^{k+1} = x^k - J_k^{-1} \Phi(x^k), \quad (2.31)$$

where $\{J_k\} \subset \mathbf{R}^{n \times n}$ is a sequence of nonsingular matrices satisfying the so-called *Dennis–Moré condition* (see [57, 58]):

$$(J_k - \Phi'(x^k))(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \quad (2.32)$$

as $k \rightarrow \infty$. (As usual, (2.31) does not mean that a matrix is inverted in actual computation.)

Evidently, x^{k+1} is a solution of (2.11) with

$$\omega^k = (J_k - \Phi'(x^k))(x^{k+1} - x^k).$$

Note that (2.32) is merely an asymptotic condition of an a posteriori kind, not relating the properties of two subsequent iterates in any constructive way. Thus, one should certainly not expect any complete convergence results, and even less so, any a priori results (i.e., proving convergence itself) under an assumption so weak. What can be expected, at best, is the superlinear rate of convergence *assuming* convergence of $\{x^k\}$ to a solution \bar{x} of (2.1) with nonsingular $\Phi'(\bar{x})$. And this is indeed valid, according to Proposition 2.4 and Remark 2.5.

Theorem 2.9. *Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at \bar{x} . Let \bar{x} be a solution of the equation (2.1). Let $\{J_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of nonsingular matrices, and let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to \bar{x} , with (2.31) holding for all k large enough.*

If the rate of convergence of $\{x^k\}$ is superlinear, then condition (2.32) holds.

Conversely, if $\Phi'(\bar{x})$ is a nonsingular matrix and condition (2.32) holds, then the rate of convergence of $\{x^k\}$ is superlinear.

For the basic Newton method (2.3), the Dennis–Moré condition (2.32) is, of course, automatic. The idea of practical quasi-Newton methods is to avoid computation of the exact Jacobian $\Phi'(x^k)$ altogether (since this is often too costly and sometimes simply impossible). The task is to approximate $\Phi'(x^k)$ in some sense, employing information about the values of Φ only. It is important to emphasize that this approximation does not subsume that $\|J_k - \Phi'(x^k)\| \rightarrow 0$ as $k \rightarrow \infty$ and, in fact, this relation indeed does not hold

for specific quasi-Newton methods (in general). The needed approximations must be computed according to some recursive formulas, and without using any information about the derivatives of Φ .

For each k , define

$$s^k = x^{k+1} - x^k, \quad r^k = \Phi(x^{k+1}) - \Phi(x^k).$$

Note that these two vectors are already known by the time when J_{k+1} has to be computed. The goal to satisfy (2.32) can be modeled as the equality

$$r^k = J_{k+1}s^k, \quad (2.33)$$

which is usually referred to as the *quasi-Newton* (or *secant*) *equation*. Indeed, from (2.32) it follows that J_{k+1} should be chosen in such a way that the vector $J_{k+1}s^k$ approximates $\Phi'(x^{k+1})s^k$. At the same time,

$$r^k = \int_0^1 \Phi'(x^k + ts^k)s^k dt,$$

and implicitly assuming that the matrix $\Phi'(x^k + ts^k)$ in the right-hand side of the last equality approximates $\Phi'(x^{k+1})$ (which is automatic provided the sequence $\{x^k\}$ converges), the idea to impose the equality (2.33) comes naturally.

Therefore, having at hand a nonsingular matrix J_k and vectors s^k and r^k , it is suggested to choose a matrix J_{k+1} satisfying the quasi-Newton equation (2.33). However, such a choice would clearly be not unique. Having in mind stability considerations, it is natural to additionally require the matrix change $J_{k+1} - J_k$ to be “minimal” in some sense: from one iteration to another, the variation of J_k should not be too large. Different understandings of “minimal” lead to different specific quasi-Newton methods. For instance, consider the case when the correction $J_{k+1} - J_k$ is minimal in the Frobenius norm. Taking into account that linearity of constraints is a CQ, by applying the Lagrange principle (Theorem 1.11), we immediately obtain the following.

Proposition 2.10. *For any elements $s^k \in \mathbf{R}^n \setminus \{0\}$ and $r^k \in \mathbf{R}^n$, and for any matrix $J_k \in \mathbf{R}^{n \times n}$, the unique (global) solution of the problem*

$$\begin{aligned} & \text{minimize} \quad \|J - J_k\|_F^2 \\ & \text{subject to} \quad Js^k = r^k \end{aligned}$$

is given by

$$J_{k+1} = J_k + \frac{(r^k - J_k s^k)s_k^T}{\|s^k\|^2}. \quad (2.34)$$

Proposition 2.10 motivates *Broyden's method*, which is one of the popular quasi-Newton methods for systems of equations: J_0 is an arbitrary nonsingular matrix (e.g., $J_0 = I$), and for each k , the matrix J_{k+1} is computed according to (2.34).

If $n = 1$, formula (2.34) reduces to the following:

$$J_{k+1} = \frac{\Phi(x^{k+1}) - \Phi(x^k)}{x^{k+1} - x^k} = J_k \left(1 - \frac{\Phi(x^{k+1})}{\Phi(x^k)} \right), \quad (2.35)$$

which corresponds to the classical *secant method*.

For an excellent survey of practical quasi-Newton methods for nonlinear equations, see [191].

We proceed to an a priori analysis for the cases when the perturbation term has certain structure. The sequence $\{x^k\}$ is not regarded as given anymore, and the role of the perturbation terms $\{\omega^k\}$ is now primary with respect to $\{x^k\}$.

In many practical algorithms based on (2.11), ω^k depends linearly on x , which is only natural: it is highly desirable to preserve linearity of the iteration system of the pure Newton method in its modifications (so that it remains relatively easy to solve). Let $\omega^k = \omega^k(x) = \Omega_k(x - x^k)$, $x \in \mathbf{R}^n$, where $\Omega_k \in \mathbf{R}^{n \times n}$ for each k . Thus, we consider now the process with the iteration system of the form

$$\Phi(x^k) + (\Phi'(x^k) + \Omega_k)(x - x^k) = 0. \quad (2.36)$$

Note that quasi-Newton methods formally fit this instance of perturbed Newton method by setting $\Omega_k = J_k - \Phi'(x^k)$ (It should be remarked, however, that in what follows we assume that the sequence $\{\Omega_k\}$ is at least bounded, a property which is not automatic for quasi-Newton methods).

Theorem 2.11. *Under the assumptions of Theorem 2.2, it holds that for any fixed $\theta \in (0, \|(\Phi'(\bar{x}))^{-1}\|^{-1}/2)$ there exists $\delta > 0$ such that for any sequence of matrices $\{\Omega_k\} \subset \mathbf{R}^{n \times n}$ satisfying*

$$\|\Omega_k\| \leq \theta \quad \forall k = 0, 1, \dots, \quad (2.37)$$

any $x^0 \in B(\bar{x}, \delta)$ uniquely defines the iterative sequence $\{x^k\} \subset B(\bar{x}, \delta)$ such that for each $k = 0, 1, \dots$, the point x^{k+1} satisfies the relation (2.11) with $\omega^k = \Omega_k(x^{k+1} - x^k)$; this sequence converges to \bar{x} , and the rate of convergence is (at least) linear. Specifically, there exists $q(\theta) \in (0, 1)$ such that $q(\theta) = O(\theta)$ as $\theta \rightarrow 0$, and either $x^k = \bar{x}$ for all k large enough, or

$$\limsup_{k \rightarrow \infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} \leq q(\theta). \quad (2.38)$$

Moreover, the rate of convergence is superlinear if $\{\Omega_k\} \rightarrow 0$ as $k \rightarrow \infty$. The rate of convergence is quadratic, provided the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} and $\Omega_k = O(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$ as $k \rightarrow \infty$.

Proof. Employing Lemma A.6, by (2.37) and the restriction on θ we obtain that there exists $\delta > 0$ such that for all $x \in B(\bar{x}, \delta)$ and all $k = 0, 1, \dots$, it holds that

$\Phi'(x) + \Omega_k$ is nonsingular,

$$\|(\Phi'(x) + \Omega_k)^{-1}\| \leq \frac{\|(\Phi'(\bar{x}))^{-1}\|}{1 - (\theta + \|\Phi'(x) - \Phi'(\bar{x})\|)\|(\Phi'(\bar{x}))^{-1}\|}.$$

Thus, for any $k = 0, 1, \dots$, if $x^k \in B(\bar{x}, \delta)$, then the equation (2.36) has the unique solution x^{k+1} , and

$$\begin{aligned} \|\omega^k\| &= \|\Omega_k(x^{k+1} - x^k)\| \\ &\leq \|\Omega_k\| \|(\Phi'(x^k) + \Omega_k)^{-1} \Phi(x^k)\| \\ &\leq \|\Omega_k\| \|(\Phi'(x^k) + \Omega_k)^{-1} (\Phi(\bar{x}) + \Phi'(x^k)(x^k - \bar{x}))\| + o(\|x^k - \bar{x}\|) \\ &\leq \|\Omega_k\| \|x^k - \bar{x} - (\Phi'(x^k) + \Omega_k)^{-1} \Omega_k(x^k - \bar{x})\| + o(\|x^k - \bar{x}\|) \\ &\leq \|\Omega_k\| \left(1 + \frac{\theta \|(\Phi'(\bar{x}))^{-1}\|}{1 - \theta \|(\Phi'(\bar{x}))^{-1}\|} \right) \|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|) \\ &\leq \frac{\theta}{1 - \theta \|(\Phi'(\bar{x}))^{-1}\|} \|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|) \end{aligned} \quad (2.39)$$

as $x^k \rightarrow \bar{x}$, where (2.37) was again taken into account. It follows that

$$\|(\Phi'(\bar{x}))^{-1} \omega^k\| \leq \|(\Phi'(\bar{x}))^{-1}\| \|\omega^k\| \leq q(\theta) \|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|),$$

where $q(\theta) = \theta \|(\Phi'(\bar{x}))^{-1}\| / (1 - \theta \|(\Phi'(\bar{x}))^{-1}\|)$. Note that by the restriction on θ , it holds that $q(\theta) < 1$, and for any $\varepsilon \in (0, 1 - q(\theta))$ the inequality

$$\|(\Phi'(\bar{x}))^{-1} \omega^k\| \leq (q(\theta) + \varepsilon) \|x^k - \bar{x}\|$$

is valid provided δ is small enough. By assertion (a) of Proposition 2.6, this implies the inclusion $x^{k+1} \in B(\bar{x}, \delta)$, perhaps for a smaller δ . It follows that any starting point $x^0 \in B(\bar{x}, \delta)$ uniquely defines the iterative sequence $\{x^k\}$ such that for each $k = 0, 1, \dots$, the point x^{k+1} satisfies (2.11), and this sequence is contained in $B(\bar{x}, \delta)$ and converges to \bar{x} . Moreover, by assertion (b) of Proposition 2.6, the rate of convergence is at least linear; more precisely, either $x^k = \bar{x}$ for all k large enough, or

$$\limsup_{k \rightarrow \infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} \leq q(\theta) + \varepsilon.$$

Since ε can be taken arbitrarily small at the price of reducing δ , and since $\{x^k\}$ converges to \bar{x} (and hence, the tail of $\{x^k\}$ is contained in $B(\bar{x}, \delta)$ no matter how small δ is), the latter implies (2.38).

Finally, by the next to last inequality in (2.39), we obtain that if it holds that $\{\Omega_k\} \rightarrow 0$, then (2.20) (and, hence, (2.16)) are valid, and the superlinear convergence rate follows from Proposition 2.4.

Similarly, if $\Omega_k = O(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$ as $k \rightarrow \infty$, then (2.17) holds, and Proposition 2.4 gives the quadratic convergence rate. \square

The simplest case of a linear perturbation term is when Ω_k is just constant, i.e., $\Omega_k = \Omega$ for all $k = 0, 1, \dots$, with some fixed $\Omega \in \mathbf{R}^{n \times n}$. Having in mind faster convergence, it is natural to choose Ω_k in such a way that $\Phi'(x^k) + \Omega_k$ is some approximation of $\Phi'(\bar{x})$. One of the possibilities is $\Omega_k = \Phi'(x^0) - \Phi'(x^k)$ for a given starting point $x^0 \in \mathbf{R}^n$. Assuming that $\Phi'(x^0)$ is nonsingular, this iterative scheme can be written in the form

$$x^{k+1} = x^k - (\Phi'(x^0))^{-1} \Phi(x^k), \quad k = 0, 1, \dots \quad (2.40)$$

The iteration cost of the basic Newton method is thus reduced, since the derivative of Φ is computed only once (at x^0) and all the iteration linear systems have the same matrix $\Phi'(x^0)$, which has to be factorized also only once (if factorization is used). From Theorem 2.11, it readily follows that the scheme (2.40) possesses local convergence to a solution with a nonsingular Jacobian. The rate of convergence is only linear, though the closer x^0 is to \bar{x} the higher is the rate of convergence, becoming superlinear in the limit. In practice, one can use a modification of this scheme, with $\Phi'(x^k)$ being computed not only for $k = 0$ but on some subsequence of iterations (but not on all iterations). Such compromise between the basic Newton method and method (2.40) is intended for reducing the iteration costs of the former while increasing the rate of convergence of the latter.

It is also sometimes useful to take $\Omega_k = \Omega(x^k)$, $k = 0, 1, \dots$, where the mapping $\Omega : \mathbf{R}^n \rightarrow \mathbf{R}^{n \times n}$ is such that $\Omega(x) \rightarrow 0$ as $x \rightarrow \bar{x}$. According to Theorem 2.11, any method of this kind possesses local superlinear convergence to a solution \bar{x} whenever $\Phi'(\bar{x})$ is nonsingular.

A particular construction of $\Omega(\cdot)$ in the case when the explicit expression for $\Phi'(\cdot)$ is available can be based on the following observation: if some terms in the expression for $\Phi'(\cdot)$ are known to vanish at a solution, such terms can be dropped (set to zero) in a Newton-type method from the very beginning.

Consider, for example, an over-determined system

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

where $\Psi : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is twice differentiable near a solution \bar{x} , with its second derivative continuous at \bar{x} , and with m generally bigger than n . This problem can be reduced to the standard form (2.1), with the number of equations equal to the number of the unknowns, by setting $\Phi(x) = (\Psi'(x))^T \Psi(x)$, $x \in \mathbf{R}^n$. Moreover, if \bar{x} satisfies the condition $\ker \Psi'(\bar{x}) = \{0\}$ (sufficient for \bar{x} to be an isolated solution; see Proposition 1.32), then $\Phi'(\bar{x}) = (\Psi'(\bar{x}))^T \Psi'(\bar{x})$ is nonsingular. At points that are not solutions, the derivative of Ψ depends not only on the first but also on the second derivative of Φ :

$$\Phi'(x)\xi = (\Psi'(x))^T \Psi'(x)\xi + (\Psi''(x)[\xi])^T \Psi(x), \quad x, \xi \in \mathbf{R}^n,$$

which makes the use of the basic Newton method even more costly in this setting. Fortunately, the last term in the expression for $\Phi'(\cdot)$ vanishes at a

solution. Dropping this term, we obtain the *Gauss–Newton method*: for a given $x^k \in \mathbf{R}^n$, the next iterate $x^{k+1} \in \mathbf{R}^n$ is computed as a solution of the iteration system

$$(\Psi'(x^k))^T \Psi(x^k) + (\Psi'(x^k))^T \Psi'(x^k)(x - x^k) = 0, \quad (2.41)$$

which corresponds to (2.36) with the linear perturbation term $\Omega_k = \Omega(x^k)$ defined by

$$\Omega(x)\xi = -(\Psi''(x)[\xi])^T \Psi(x), \quad x, \xi \in \mathbf{R}^n.$$

Note that if $n = m$, then this iterative process generates the same iterative sequence as the basic Newton method. Note also that the expression in the left-hand side of (2.41) is precisely the gradient of the quadratic objective function of the following linear least-squares problem:

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|\Psi(x^k) + \Psi'(x^k)(x - x^k)\|^2 \\ & \text{subject to} && x \in \mathbf{R}^n. \end{aligned}$$

The latter can be solved by special algorithms for linear least-squares problems [208, Sect. 10.2], or by conjugate gradient methods [208, Sect. 5.1], without explicitly computing the product $(\Psi'(x^k))^T \Psi'(x^k)$, which could be too expensive.

Local superlinear convergence of the Gauss–Newton method under the assumption $\ker \Psi'(\bar{x}) = \{0\}$ is ensured by Theorem 2.11, according to the discussion above.

Even though keeping the iteration system linear is certainly a reasonable approach, it will be seen below that there exist some practical algorithms (for constrained optimization) fitting the perturbed Newton method framework for which the dependence of the perturbation term on the variables is not necessarily linear. Instead, it satisfies some smoothness-like assumptions, still allowing an a priori analysis via the use of the implicit function theorem. One such example is the linearly constrained augmented Lagrangian method for optimization, discussed in Sect. 4.1.2. This motivates the following results, dealing with nonlinear dependence of perturbations on the problem variables.

Theorem 2.12. *Under the hypotheses of Theorem 2.2, let $\omega : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}^n$ satisfy the following assumptions:*

$$\omega(x, \xi^1) - \omega(x, \xi^2) = o(\|\xi^1 - \xi^2\|) \quad (2.42)$$

as $\xi^1, \xi^2 \in \mathbf{R}^n$ tend to 0, uniformly in $x \in \mathbf{R}^n$ close enough to \bar{x} , and there exists $\theta \in (0, \|(\Phi'(\bar{x}))^{-1}\|^{-1})$ such that the inequality

$$\|\omega(x, 0)\| \leq \theta \|x - \bar{x}\| \quad (2.43)$$

holds for all $x \in \mathbf{R}^n$ close enough to \bar{x} .

Then there exists $\delta > 0$ such that any starting point $x^0 \in \mathbf{R}^n$ close enough to \bar{x} uniquely defines the iterative sequence $\{x^k\} \subset \mathbf{R}^n$ such that x^{k+1} satisfies (2.11) with $\omega^k = \omega(x^k, x^{k+1} - x^k)$ for each $k = 0, 1, \dots$, and $\|x^{k+1} - x^k\| \leq \delta$; this sequence converges to \bar{x} , and the rate of convergence is (at least) linear. Specifically, there exists $q(\theta) \in (0, 1)$ such that (2.38) holds, and $q(\theta) = O(\theta)$ as $\theta \rightarrow 0$.

Moreover, the rate of convergence is superlinear if

$$\omega(x, 0) = o(\|x - \bar{x}\|) \quad (2.44)$$

as $x \rightarrow \bar{x}$. The rate of convergence is quadratic provided the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} and

$$\omega(x, \xi) = O(\|\xi\|^2 + \|x - \bar{x}\|^2) \quad (2.45)$$

as $x \rightarrow \bar{x}$ and $\xi \rightarrow 0$.

Proof. Define the mapping $\Psi : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}^n$,

$$\Psi(x, \xi) = \Phi(x) + \Phi'(x)\xi + \omega(x, \xi).$$

By the assumptions (2.42) and (2.43), the implicit function theorem (Theorem 1.22) is applicable to this mapping at $(x, \xi) = (\bar{x}, 0)$ (here, x is regarded as a parameter). Hence, there exist $\delta > 0$ and $\tilde{\delta} > 0$ such that for each $x \in B(\bar{x}, \tilde{\delta})$ the equation

$$\Psi(x, \xi) = 0$$

has the unique solution $\xi(x) \in B(0, \delta)$, and this solution satisfies the estimate

$$\|\xi(x)\| = O(\|\Psi(x, 0)\|) = O(\|\Phi(x)\|) + O(\|\omega(x, 0)\|) = O(\|x - \bar{x}\|) \quad (2.46)$$

as $x \rightarrow \bar{x}$. Then for any $x^k \in B(\bar{x}, \tilde{\delta})$, the point $x^{k+1} = x^k + \xi(x^k)$ is the only one in $B(x^k, \delta)$ satisfying (2.11) with $\omega^k = \omega(x^k, x^{k+1} - x^k)$. Furthermore,

$$\begin{aligned} \|\omega^k\| &= \|\omega(x^k, \xi(x^k))\| \\ &\leq \|\omega(x^k, \xi(x^k)) - \omega(x^k, 0)\| + \|\omega(x^k, 0)\| \\ &= \|\omega(x^k, 0)\| + o(\|\xi(x^k)\|) \\ &= \|\omega(x^k, 0)\| + o(\|x^k - \bar{x}\|) \\ &\leq \theta\|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|) \end{aligned} \quad (2.47)$$

as $x^k \rightarrow \bar{x}$, where (2.42) and (2.43) were employed again. It follows that

$$\|(\Phi'(\bar{x}))^{-1}\omega^k\| \leq \|(\Phi'(\bar{x}))^{-1}\|\|\omega^k\| \leq q(\theta)\|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|),$$

where $q(\theta) = \theta\|(\Phi'(\bar{x}))^{-1}\|$. Note that by the restriction on θ , it holds that $q(\theta) < 1$.

The rest of the proof almost literally repeats the corresponding part of the proof of Theorem 2.11. In particular, convergence follows from Proposition 2.6.

The superlinear convergence rate under the assumption (2.44) follows by the third equality in (2.47), and by Proposition 2.4. Moreover, assuming (2.45), the estimate (2.47) can be sharpened as follows:

$$\omega^k = \omega(x^k, \xi(x^k)) = O(\|\xi(x^k)\|^2 + \|x^k - \bar{x}\|^2) = O(\|x^k - \bar{x}\|^2)$$

as $x^k \rightarrow \bar{x}$, where the last equality is by (2.46). Proposition 2.4 now gives quadratic convergence rate, provided the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} . \square

Note that the case discussed above when $\omega^k = \omega^k(x) = \Omega(x^k)(x - x^k)$, $k = 0, 1, \dots$, with some mapping $\Omega : \mathbf{R}^n \rightarrow \mathbf{R}^{n \times n}$ such that $\Omega(x) \rightarrow 0$ as $x \rightarrow \bar{x}$ (in particular, the Gauss–Newton method), can be treated both by Theorem 2.11 or 2.12. More interesting examples of the use of Theorem 2.12 will be provided below (see Sects. 4.1, 4.2).

The next result is, in a sense, intermediate between a priori and a posteriori characterizations of perturbed Newton method. We present it here mainly because of the conceptual importance of this kind of analysis for Newton-type methods in the setting of variational problems, where the existence of solutions of subproblems can be guaranteed in general only under rather strong assumptions; see Sect. 3.1. In such cases, it may be useful just to *assume* solvability of subproblems, having in mind that this can be verifiable separately, for more specific algorithms and/or problem classes.

For this analysis, it is natural to replace (2.11) by the generalized equation (GE)

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) + \Omega(x^k, x - x^k) \ni 0, \quad (2.48)$$

with a multifunction Ω from $\mathbf{R}^n \times \mathbf{R}^n$ to the subsets of \mathbf{R}^n .

Theorem 2.13. *Under the assumptions of Theorem 2.2, let Ω be a multifunction from $\mathbf{R}^n \times \mathbf{R}^n$ to the subsets of \mathbf{R}^n , satisfying the following assumptions: for each $x \in \mathbf{R}^n$ close enough to \bar{x} , the GE*

$$\Phi(x) + \Phi'(x)\xi + \Omega(x, \xi) \ni 0 \quad (2.49)$$

has a solution $\xi(x)$ such that $\xi(x) \rightarrow 0$ as $x \rightarrow \bar{x}$, and there exist $\theta_1, \theta_2 \geq 0$ such that $2\theta_1 + \theta_2 \leq \|(\Phi'(\bar{x}))^{-1}\|^{-1}$ and the inequality

$$\|\Phi(x) + \Phi'(x)\xi\| \leq \theta_1 \|\xi\| + \theta_2 \|x - \bar{x}\| \quad (2.50)$$

holds for all $x \in \mathbf{R}^n$ close enough to \bar{x} and all $\xi \in \mathbf{R}^n$ close enough to zero, satisfying (2.49).

Then there exists $\delta > 0$ such that for any starting point $x^0 \in \mathbf{R}^n$ close enough to \bar{x} , there exists a sequence $\{x^k\} \subset \mathbf{R}^n$ such that x^{k+1} is a solution of the GE (2.48) for each $k = 0, 1, \dots$, satisfying

$$\|x^{k+1} - x^k\| \leq \delta; \quad (2.51)$$

any such sequence converges to \bar{x} , and the rate of convergence is (at least) linear. Specifically, there exists $q(\theta) \in (0, 1)$, $\theta = \theta_1 + \theta_2$, such that (2.38) holds, and $q(\theta) = O(\theta)$ as $\theta \rightarrow 0$.

Moreover, the rate of convergence is superlinear if (2.50) can be replaced by the stronger condition

$$\Phi(x) + \Phi'(x)\xi = o(\|\xi\| + \|x - \bar{x}\|) \quad (2.52)$$

as $x \rightarrow \bar{x}$ and $\xi \rightarrow 0$. The rate of convergence is quadratic provided the derivative of Φ is locally Lipschitz-continuous with respect to \bar{x} , and provided (2.50) can be replaced by the even stronger condition

$$\Phi(x) + \Phi'(x)\xi = O(\|\xi\|^2 + \|x - \bar{x}\|^2). \quad (2.53)$$

Proof. Under the assumptions of the theorem, there exist $\delta > 0$ and $\tilde{\delta} > 0$ such that for any $x^k \in B(\bar{x}, \tilde{\delta})$, there exists $x^{k+1} \in B(x^k, \delta)$ (specifically, $x^{k+1} = x^k + \xi(x^k)$) satisfying (2.48). Assuming that δ and $\tilde{\delta}$ are small enough, for any such x^{k+1} , by setting $\omega^k = -\Phi(x^k) - \Phi'(x^k)(x^{k+1} - x^k)$, we obtain that (2.11) holds with $x = x^{k+1}$, and

$$\|\omega^k\| \leq \theta_1 \|x^{k+1} - x^k\| + \theta_2 \|x^k - \bar{x}\|,$$

where (2.50) was employed. It follows that

$$\begin{aligned} \|(\Phi'(\bar{x}))^{-1}\omega^k\| &\leq \|(\Phi'(\bar{x}))^{-1}\| \|\omega^k\| \\ &\leq q_1(\theta_1) \|x^{k+1} - x^k\| + q_2(\theta_2) \|x^k - \bar{x}\|, \end{aligned}$$

where $q_j(\theta_j) = \theta_j \|(\Phi'(\bar{x}))^{-1}\|$, $j = 1, 2$, satisfy $2q_1(\theta_1) + q_2(\theta_2) < 1$.

The rest of the proof again almost literally repeats the corresponding part of the proof of Theorem 2.11. Convergence follows by Proposition 2.6, and (2.38) holds with $q(\theta) = (q_1(\theta_1) + q_2(\theta_2))/(1 - q_1(\theta_1))$. The superlinear/quadratic convergence rate under the corresponding additional assumptions follows by Proposition 2.4. \square

We complete this section with a brief discussion of the case when $\Phi'(\bar{x})$ is not necessarily nonsingular. Such cases will be treated in detail later in this book for (generalized) equations possessing some special (primal-dual) structure, arising from optimization and variational problems. For general equations without any special structure, the behavior of Newton-type methods near solutions with singular Jacobians, as well as various modifications of these methods intended for preserving the efficiency despite singularity, was studied in [151, 152]. Here, we limit the discussion to some comments which may give an initial understanding of the effect of singularity.

Consider the scalar equation

$$x^s = 0,$$

where $s \geq 2$ is an integer parameter. The Newton method iterations for this equation are given by $x^{k+1} = (1 - 1/s)x^k$, and the sequence $\{x^k\}$ converges to the unique solution $\bar{x} = 0$ from any starting point, but the rate of convergence is only linear. This happens because \bar{x} is a singular solution: the derivative at \bar{x} is zero. At the same time, if we modify the Newton method by introducing the stepsize parameter equal to s , the method hits the exact solution in one step, for any starting point x^0 .

More generally, the following fact was established in [242]. Let a function $\Phi : \mathbf{R} \rightarrow \mathbf{R}$ be s times differentiable at $\bar{x} \in \mathbf{R}$, $s \geq 2$, where \bar{x} is a root of multiplicity s of the equation (2.1), i.e.,

$$\Phi(\bar{x}) = \Phi'(\bar{x}) = \dots = \Phi^{(s-1)}(\bar{x}) = 0, \quad \Phi^{(s)}(\bar{x}) \neq 0.$$

Then the Newton method iterates locally converge to \bar{x} at a linear rate, while the method modified by introducing the stepsize parameter equal to s gives the superlinear convergence rate.

2.1.2 Newton Method for Unconstrained Optimization

Consider now the unconstrained optimization problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in \mathbf{R}^n, \end{aligned} \tag{2.54}$$

with a twice differentiable objective function $f : \mathbf{R}^n \rightarrow \mathbf{R}$. Stationary points of this problem are characterized by the equation (2.1) with $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ being the *gradient mapping* of f :

$$\Phi(x) = f'(x).$$

Thus, one strategy to compute stationary points of the optimization problem (2.54) is to apply some Newton-type method to the equation (2.1) with Φ being the gradient of f .

In the case of the basic Newton method for (2.54), given $x^k \in \mathbf{R}^n$, the next iterate x^{k+1} is computed as a solution of the linear system

$$f'(x^k) + f''(x^k)(x - x^k) = 0. \tag{2.55}$$

Assuming that the Hessian $f''(x^k)$ is nonsingular, the Newton method can be written in the form of the explicit iterative scheme

$$x^{k+1} = x^k - (f''(x^k))^{-1} f'(x^k), \quad k = 0, 1, \dots$$

This iteration allows for the following interpretation that puts to the foreground the optimization nature of the original problem. Near the current iterate x^k , the objective function f is naturally approximated by its second-order expansion or, in other words, the original problem (2.54) is approximated by the following subproblem:

$$\begin{aligned} &\text{minimize} && f(x^k) + \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle f''(x^k)(x - x^k), x - x^k \rangle \\ &\text{subject to} && x \in \mathbf{R}^n. \end{aligned} \quad (2.56)$$

Since (2.55) is precisely the equation defining stationary points of (2.56), the basic *Newton method* for unconstrained optimization can be presented as follows.

Algorithm 2.14 Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $f'(x^k) = 0$, stop.
2. Compute $x^{k+1} \in \mathbf{R}^n$ as a stationary point of problem (2.56).
3. Increase k by 1 and go to step 1.

Local convergence result for Newton method for unconstrained optimization follows immediately from Theorem 2.2 on local convergence of Newton method for equations.

Theorem 2.15. *Let a function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its Hessian being continuous at \bar{x} . Let \bar{x} be a stationary point of problem (2.54), and assume that this point satisfies the SOSC*

$$\langle f''(\bar{x})\xi, \xi \rangle > 0 \quad \forall \xi \in \mathbf{R}^n \setminus \{0\} \quad (2.57)$$

(thus, according to Theorem 1.9, \bar{x} is a strict local solution of problem (2.54)).

Then any starting point $x^0 \in \mathbf{R}^n$ close enough to \bar{x} uniquely defines the iterative sequence of Algorithm 2.14; this sequence converges to \bar{x} , and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the Hessian of f is locally Lipschitz-continuous with respect to \bar{x} .

As one specificity of Newton method for optimization, let us mention that a Hessian of a twice differentiable function is a symmetric matrix. Under the assumptions of Theorem 2.15, perhaps the most natural general strategy for solving the iteration system (2.55) appears to be the so-called Cholesky factorization, which provides the LL^T -decomposition of a positive definite symmetric matrix (L is a lower triangular matrix with positive diagonal elements) at a price of $n^3/6$ multiplications and the same amount of additions (see, e.g., [100], [261, Lecture 23], [103, Sect.4.2]). More details on special tools of numerical linear algebra for iteration systems arising in optimization can be found, e.g., in [29, 208].

Note also that the assertion of Theorem 2.15 remains valid if the SOSC (2.57) is replaced by the weaker assumption that $f''(\bar{x})$ is nonsingular. In this respect, Newton method does not distinguish local minimizers from other stationary points of the problem (including the maximizers).

The main advantage of Newton method is its high convergence rate (superlinear, under natural assumptions). However, the basic Newton method has also serious drawbacks, which we discuss next.

First, each step of the Newton method requires computing the Hessian and solving the corresponding linear system, which can be too costly, or simply impossible in some applications. Regarding this issue, we note that perturbed Newton methods for equations discussed in Sect. 2.1.1 can be directly adapted for unconstrained optimization. Indeed, all these methods can be applied to the equation defined by the gradient of f . This may help to reduce the iteration costs significantly. One important example is the class of quasi-Newton methods for unconstrained optimization, discussed in Sect. 2.2.

The second inevitable drawback of pure Newton-type methods is that they possess only local convergence: in all results presented above, a starting point close enough to a solution is required. An iterative sequence of Newton method defined by an inappropriate starting point may not have stationary points of the problem among its accumulations points. In fact, this may happen even in the case of a strongly convex objective function (so that its stationary point is unique, and it is the unique global minimizer).

Example 2.16. Consider the function $f : \mathbf{R} \rightarrow \mathbf{R}$,

$$f(x) = \begin{cases} -\frac{x^4}{4\sigma^3} + \left(1 + \frac{3}{\sigma}\right) \frac{x^2}{2} & \text{if } |x| \leq \sigma, \\ \frac{x^2}{2} + 2|x| - \frac{3\sigma}{4} & \text{if } |x| > \sigma, \end{cases}$$

where $\sigma > 0$ is a parameter. It can be easily checked that for any such σ , the function f is twice continuously differentiable and strongly convex on \mathbf{R} , and problem (2.54) with this objective function has the unique stationary point $\bar{x} = 0$. In particular, $f''(\bar{x}) = 1 + 3/\sigma > 0$, and all the assumptions of Theorem 2.15 are satisfied. Take $x^0 = \sigma$. The corresponding iterative sequence $\{x^k\}$ of Algorithm 2.14 is then given by $x^k = 2(-1)^k$, $k = 1, 2, \dots$, and \bar{x} is not an accumulation point of $\{x^k\}$, no matter how small σ is.

Strategies for globalization of convergence of Newton-type methods for unconstrained optimization is the subject of the rest of this chapter. In particular, linesearch quasi-Newton methods (to be discussed in Sect. 2.2) serve not only for reducing the iteration costs but also for enforcing global convergence of Newton-type methods. (This is the main reason why we present quasi-Newton methods for unconstrained optimization in the context of linesearch methods.)

2.2 Linesearch Methods, Quasi-Newton Methods

In this section, we consider the unconstrained optimization problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in \mathbf{R}^n, \end{aligned} \tag{2.58}$$

with a differentiable objective function $f : \mathbf{R}^n \rightarrow \mathbf{R}$. One of the most natural approaches to solving (2.58) is the following. For the given iterate, compute a descent direction for f at this point, and make a step of some length along this direction so that the value of f is (sufficiently) reduced. Repeat the procedure for the obtained new iterate, etc. We refer to methods of this kind as descent methods. Evidently, efficiency of any such method depends on two choices: that of the descent direction, and that of the stepsize. Perhaps the most practically important example of good choices for both is the class of linesearch quasi-Newton methods.

2.2.1 Descent Methods

We start with a formal definition of descent directions.

Definition 2.17. A vector $p \in \mathbf{R}^n$ is said to be a *descent direction* for the function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ at $x \in \mathbf{R}^n$ if for all $t > 0$ small enough it holds that $f(x + tp) < f(x)$.

The set of all descent directions for f at $x \in \mathbf{R}^n$ is a cone, which will be denoted by $\mathcal{D}_f(x)$. Therefore, $p \in \mathcal{D}_f(x)$ if and only if any sufficiently small displacement of x in the direction p results in a reduction of the function value with respect to $f(x)$. The next statement is elementary.

Lemma 2.18. Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be differentiable at $x \in \mathbf{R}^n$.

Then the following assertions are valid:

- (a) For any $p \in \mathcal{D}_f(x)$ it holds that $\langle f'(x), p \rangle \leq 0$.
- (b) If for $p \in \mathbf{R}^n$ it holds that $\langle f'(x), p \rangle < 0$, then $p \in \mathcal{D}_f(x)$.

The class of *descent methods* is then given by iterative schemes of the form

$$x^{k+1} = x^k + \alpha_k p^k, \quad p^k \in \mathcal{D}_f(x^k), \quad \alpha_k > 0, \quad k = 0, 1, \dots, \tag{2.59}$$

where the *stepsize parameters* $\alpha_k > 0$ are chosen in such a way that, at least,

$$f(x^{k+1}) < f(x^k). \tag{2.60}$$

That is, the sequence $\{f(x^k)\}$ must be monotonically decreasing. (If $\mathcal{D}_f(x^k) = \emptyset$ or if an element of $\mathcal{D}_f(x^k)$ cannot be computed by the prescribed tools,

the process is terminated.) Note that the inclusion $p^k \in \mathcal{D}_f(x^k)$ implies that the inequality (2.60) holds for all $\alpha_k > 0$ small enough. However, (2.60) is obviously not enough to guarantee convergence: the reduction property must be appropriately quantified.

As mentioned above, a specific descent method is characterized by a specific rule for choosing descent directions, and a specific procedure for computing the appropriate values of the stepsize parameter. Procedures for choosing a stepsize are based on exploring the restriction of the objective function f to the ray spanned by p^k , with its origin at x^k . For this reason, such procedures are usually called *linesearch*. It is interesting to point out the following common feature of optimization algorithms: a choice of search directions p^k is typically based on some approximate model of the objective function f (see below), while linesearch procedures are normally performed for f itself.

By Lemma 2.18, if $f'(x^k) \neq 0$, then one can always take the descent direction $p^k = -f'(x^k)$. The corresponding descent methods (sometimes called *steepest descent methods*) are easy to implement, and their convergence and rate of convergence properties can be fully characterized theoretically. However, such methods are completely impractical: this choice of descent directions usually turns out to be extremely inefficient.

Much more practical descent methods are obtained within the following more general framework. Given $x^k \in \mathbf{R}^n$ take $p^k = -Q_k f'(x^k)$, where $Q_k \in \mathbf{R}^{n \times n}$ is a symmetric positive definite matrix. The matrices, of course, must be chosen in some clever way. Good choices of Q_k will be discussed in Sect. 2.2.2. Right now, we note only that by Lemma 2.18, if $f'(x^k) \neq 0$, then $p^k = -Q_k f'(x^k)$ with a positive definite Q_k is clearly a descent direction for f at x^k , since

$$\langle f'(x^k), p^k \rangle = -\langle Q_k f'(x^k), f'(x^k) \rangle < 0. \quad (2.61)$$

The “limiting,” in some sense, choices for Q_k are $Q_k = (f''(x^k))^{-1}$ corresponding to the (expensive) Newton direction (see Sect. 2.1.2) and $Q_k = I$ corresponding to the (cheap) steepest descent direction. We note that the latter can still be useful sometimes, but only as a “last resort,” in those cases when for some reasons more sophisticated options fail.

We next discuss the most important linesearch procedures, assuming that for a given iterate x^k a direction $p^k \in \mathcal{D}_f(x^k)$ is already chosen and fixed. It may seem natural to take the stepsize parameter $\alpha_k > 0$ as a global minimizer of $f(x^k + \alpha p^k)$ over all $\alpha \geq 0$. This *exact linesearch rule* is, formally, ideal: it provides the maximal possible progress in decreasing f along the given direction. If f is a quadratic function with a positive definite Hessian, then such α_k is given by an explicit formula. But beyond the quadratic case, exact linesearch is too expensive and usually impossible anyway. Moreover, even searching for a local minimizer of $f(x^k + \alpha p^k)$ (or, e.g., for the local minimizer closest to zero) is usually not worthwhile—afterall, the eventual

goal is to minimize f on the entire space rather than on the given ray. For this reason, much cheaper inexact linesearch rules are used in practice. These rules ensure *sufficient decrease* of the objective function value, instead of searching for (local or global) minimizers of f along the given descent direction.

Armijo rule. Choose the parameters $C > 0$, $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Set $\alpha = C$.

1. Check the inequality

$$f(x^k + \alpha p^k) \leq f(x^k) + \sigma \alpha \langle f'(x^k), p^k \rangle. \quad (2.62)$$

2. If (2.62) does not hold, replace α by $\theta \alpha$ and go to step 1. Otherwise, set $\alpha_k = \alpha$.

Thus, α_k is the first α of the form $C\theta^j$, $j = 0, 1, \dots$, satisfying (2.62) (the needed value is computed by a backtracking procedure starting with the initial trial value C). The quantity $\alpha \langle f'(x^k), p^k \rangle$ in the right-hand side of (2.62) plays the role of “predicted” (by the linear model of f) reduction of the objective function value for the step of length α in the direction p^k . Therefore, inequality (2.62) means that the actual reduction must be no less than a given fraction (defined by the choice of $\sigma \in (0, 1)$) of the “predicted” reduction. Armijo linesearch is illustrated in Fig. 2.3.

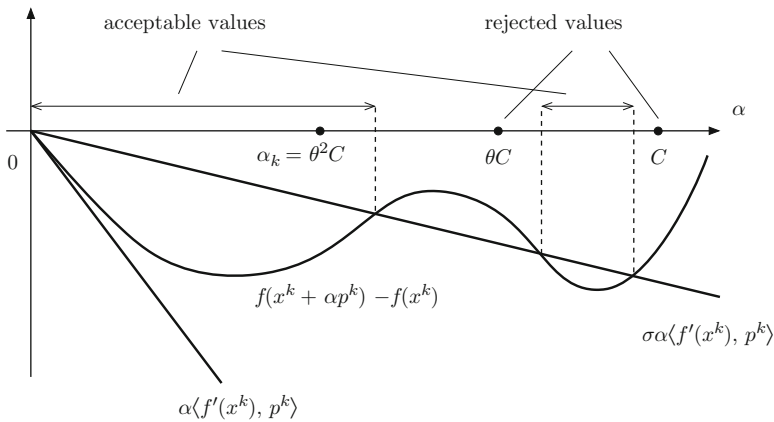


Fig. 2.3 Armijo rule

The next lemma demonstrates that if p^k satisfies the sufficient condition for a descent direction stated in Lemma 2.18, i.e., if

$$\langle f'(x^k), p^k \rangle < 0, \quad (2.63)$$

then the backtracking procedure in the Armijo rule is finite.

Lemma 2.19. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be differentiable at $x^k \in \mathbf{R}^n$.*

Then for any $p^k \in \mathbf{R}^n$ satisfying (2.63), inequality (2.62) holds for all $\alpha > 0$ small enough.

Proof. It holds that

$$\begin{aligned} f(x^k + \alpha p^k) - f(x^k) &= \langle f'(x^k), \alpha p^k \rangle + o(\alpha) \\ &= \sigma \alpha \langle f'(x^k), p^k \rangle + (1 - \sigma) \alpha \langle f'(x^k), p^k \rangle + o(\alpha) \\ &= \sigma \alpha \langle f'(x^k), p^k \rangle + \alpha \left((1 - \sigma) \langle f'(x^k), p^k \rangle + \frac{o(\alpha)}{\alpha} \right) \\ &\leq \sigma \alpha \langle f'(x^k), p^k \rangle, \end{aligned}$$

because $(1 - \sigma) \langle f'(x^k), p^k \rangle + o(\alpha)/\alpha < 0$ for any $\alpha > 0$ small enough. \square

Evidently, if (2.63) holds, then choosing α_k according to the Armijo rule guarantees the descent property (2.60). Moreover, the inequality (2.62) with $\alpha = \alpha_k$ gives a quantitative estimate of by how much $f(x^{k+1})$ is smaller than $f(x^k)$, and this estimate (unlike (2.60)) is sufficient for establishing convergence under natural assumptions. However, convergence proof is significantly simplified when one can show that the backtracking is finite uniformly with respect to k , i.e., when α_k is separated from zero by some threshold independent of k .

Lemma 2.20. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be differentiable on \mathbf{R}^n , and suppose that its gradient is Lipschitz-continuous on \mathbf{R}^n with constant $L > 0$.*

Then for any $x^k \in \mathbf{R}^n$ and $p^k \in \mathbf{R}^n$ satisfying (2.63), the inequality (2.62) holds for all $\alpha \in (0, \bar{\alpha}_k]$, where

$$\bar{\alpha}_k = \frac{2(\sigma - 1) \langle f'(x^k), p^k \rangle}{L \|p^k\|^2} > 0. \quad (2.64)$$

Proof. By Lemma A.11, for all $\alpha > 0$ it holds that

$$f(x^k + \alpha p^k) - f(x^k) - \langle f'(x^k), \alpha p^k \rangle \leq \frac{L}{2} \alpha^2 \|p^k\|^2.$$

Hence, for all $\alpha \in (0, \bar{\alpha}_k]$ we have that

$$\begin{aligned} f(x^k + \alpha p^k) - f(x^k) &\leq \langle f'(x^k), \alpha p^k \rangle + \frac{L}{2} \alpha^2 \|p^k\|^2 \\ &= \alpha \left(\langle f'(x^k), p^k \rangle + \frac{L}{2} \alpha \|p^k\|^2 \right) \\ &\leq \sigma \alpha \langle f'(x^k), p^k \rangle, \end{aligned}$$

where the last inequality follows from (2.64). \square

Lemma 2.21. *Under the assumptions of Lemma 2.20, let $\{Q_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of symmetric matrices satisfying*

$$\langle Q_k \xi, \xi \rangle \geq \gamma \|\xi\|^2 \quad \forall \xi \in \mathbf{R}^n, \quad \|Q_k\| \leq \Gamma \quad \forall k, \quad (2.65)$$

with some $\gamma > 0$ and $\Gamma > 0$.

Then there exists a constant $c > 0$ such that for any point $x^k \in \mathbf{R}^n$ and for $p^k = -Q_k f'(x^k)$, the value α_k obtained by the Armijo rule satisfies

$$\alpha_k \geq c. \quad (2.66)$$

Proof. By (2.65), we obtain that

$$\frac{\langle f'(x^k), p^k \rangle}{\|p^k\|^2} = -\frac{\langle f'(x^k), Q_k f'(x^k) \rangle}{\|Q_k f'(x^k)\|^2} \leq -\frac{\gamma}{\Gamma^2}.$$

Hence, according to (2.64),

$$\bar{\alpha}_k \geq \frac{2(1 - \sigma)\gamma}{L\Gamma^2} > 0.$$

The needed assertion now follows from Lemma 2.20. \square

The Armijo rule is simple, clear, and easy to implement. Convergence results presented below refer to this rule. However, more sophisticated linesearch techniques, with better theoretical and practical properties, are often used in practice.

Goldstein rule consists of choosing the stepsize parameter satisfying the inequalities

$$\sigma_1 \leq \frac{f(x^k + \alpha p^k) - f(x^k)}{\alpha \langle f'(x^k), p^k \rangle} \leq \sigma_2, \quad (2.67)$$

with fixed $0 < \sigma_1 < \sigma_2 < 1$.

The first inequality in (2.67) is just the Armijo inequality (2.62) with $\sigma = \sigma_1$; it guarantees sufficient decrease of the objective function. Recall that according to Lemma 2.19, this inequality holds for all $\alpha > 0$ small enough. By contrast, the second inequality in (2.67) is evidently violated for all $\alpha > 0$ close to zero. The reason for introducing the second inequality is precisely to avoid stepsize parameters that are too small. The idea is to take larger steps, i.e., prevent the method from slowing down. Goldstein linesearch is illustrated in Fig. 2.4.

Wolfe rule is another realization of the same idea, but instead of (2.67) it employs the inequalities

$$f(x^k + \alpha p^k) \leq f(x^k) + \sigma_1 \alpha \langle f'(x^k), p^k \rangle, \quad (2.68)$$

$$\langle f'(x^k + \alpha p^k), p^k \rangle \geq \sigma_2 \langle f'(x^k), p^k \rangle. \quad (2.69)$$

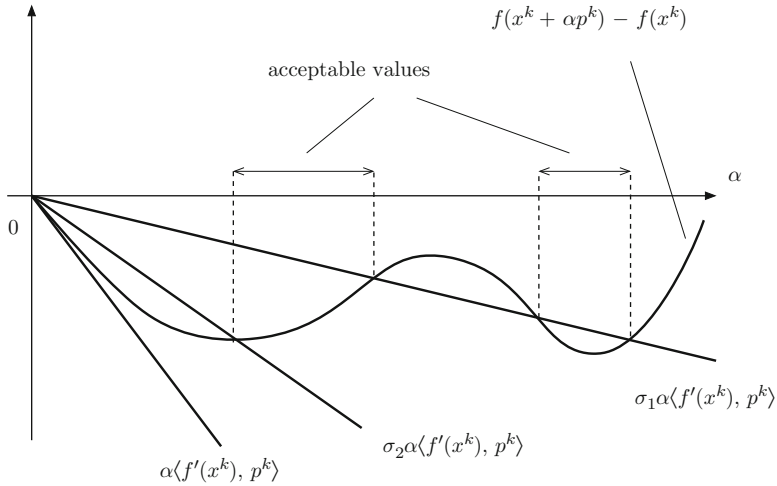


Fig. 2.4 Goldstein rule

Again, (2.68) is the Armijo inequality (2.62) with $\sigma = \sigma_1$. Evidently, analogously to (2.67), the second inequality in (2.69) also does not allow stepsize values that are too small. Note that it involves the gradient of f not only at x^k but also at the trial points $x^k + \alpha p^k$, which entails some additional computational cost. However, when computation of the gradient is not too expensive, the Wolfe rule is often regarded as the most efficient among currently known linesearch options. One important property of this rule is related to quasi-Newton methods; see Sect. 2.2.2. Wolfe linesearch is illustrated in Fig. 2.5.

We next give a simple algorithmic implementation of the Wolfe rule. (The Goldstein rule can be implemented along the same lines.) Let $0 < \sigma_1 < \sigma_2 < 1$ be fixed. Set $c = C = 0$, and choose an initial trial value $\alpha > 0$.

1. Check the inequalities (2.68) and (2.69). If both do hold, go to step 6.
2. If (2.68) does not hold, set $C = \alpha$, and go to step 5.
3. If (2.69) does not hold, set $c = \alpha$.
4. If $C = 0$, choose a new trial value $\alpha > c$ (“extrapolation”), and go to step 1.
5. Choose a new trial value $\alpha \in (c, C)$ (“interpolation”), and go to step 1.
6. Set $\alpha_k = \alpha$.

Violation of (2.68) basically means that the current trial value α is “too large,” while violation of (2.69) means that it is “too small.” The procedure just described works as follows. Extrapolation steps are performed first, until C becomes positive. Once this happened, interpolation steps are performed. In the course of interpolation C may only decrease, remaining positive, while c may only increase, staying smaller than C .

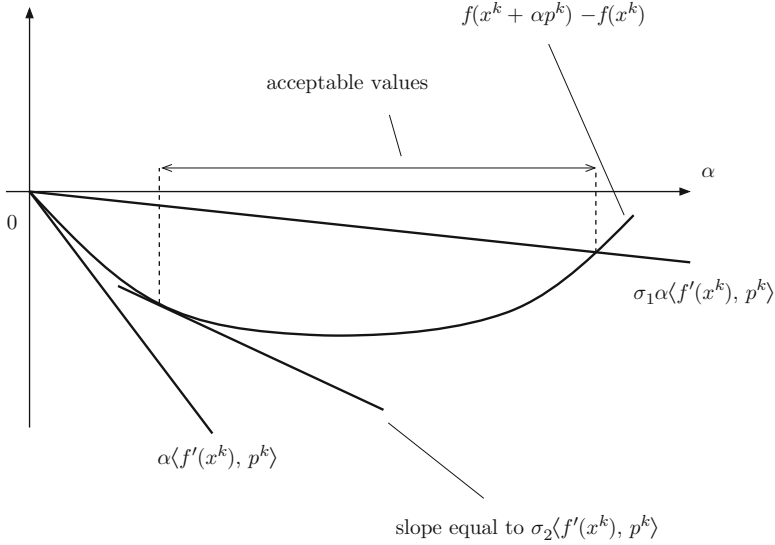


Fig. 2.5 Wolfe rule

Extrapolation and interpolation in the presented procedure can be organized in many ways. For example, one can fix $\theta_1 > 1$, $\theta_2 \in (0, 1)$, and replace α by $\theta_1 \alpha$ in the case of extrapolation, and set $\alpha = (1 - \theta_2)c + \theta_2 C$ in the case of interpolation. More sophisticated options are discussed, e.g., in [29, Chap. 3]. From the theoretical viewpoint, it is important to guarantee the following property: in the case of infinite number of extrapolation steps c must be increasing to infinity, while in the case of infinite number of interpolation steps $(C - c)$ must be tending to zero.

Lemma 2.22. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be continuously differentiable and bounded below on \mathbf{R}^n .*

Then for any $x^k \in \mathbf{R}^n$ and $p^k \in \mathbf{R}^n$ satisfying (2.63), the procedure implementing the Wolfe rule such that $c \rightarrow +\infty$ in the case of infinite number of extrapolation steps and $(C - c) \rightarrow 0$ in the case of infinite number of interpolation steps, is finite.

Proof. Suppose first that there is an infinite number of extrapolation steps. Then the procedure generates an increasing to infinity sequence of values of c , and for each of these values it holds that

$$f(x^k + cp^k) \leq f(x^k) + \sigma_1 c \langle f'(x^k), p^k \rangle. \quad (2.70)$$

But according to inequality (2.63), the latter contradicts the assumption that f is bounded below. Therefore, the number of extrapolation steps is finite.

Suppose now that the number of interpolation steps is infinite. Then the monotone sequences of values of c and of C converge to a common limit $\bar{\alpha}$. The elements of the first sequence satisfy (2.70) and

$$\langle f'(x^k + cp^k), p^k \rangle < \sigma_2 \langle f'(x^k), p^k \rangle, \quad (2.71)$$

while the elements of the second sequence satisfy

$$f(x^k + Cp^k) > f(x^k) + \sigma_1 C \langle f'(x^k), p^k \rangle. \quad (2.72)$$

By passing onto the limit in (2.70) and (2.72), we obtain the equality

$$f(x^k + \bar{\alpha}p^k) = f(x^k) + \sigma_1 \bar{\alpha} \langle f'(x^k), p^k \rangle. \quad (2.73)$$

Taking into account (2.72) and monotone decrease of the values of C , it follows that these values always remain bigger than $\bar{\alpha}$. Employing (2.73), we can rewrite inequality (2.72) in the form

$$\begin{aligned} f(x^k + Cp^k) &> f(x^k) + \sigma_1 \bar{\alpha} \langle f'(x^k), p^k \rangle + (C - \bar{\alpha}) \langle f'(x^k), p^k \rangle \\ &= f(x^k + \bar{\alpha}p^k) + \sigma_1 (C - \bar{\alpha}) \langle f'(x^k), p^k \rangle. \end{aligned}$$

Taking into account the inequality $C - \bar{\alpha} > 0$, the latter implies

$$\frac{f(x^k + Cp^k) - f(x^k + \bar{\alpha}p^k)}{C - \bar{\alpha}} > \sigma_1 \langle f'(x^k), p^k \rangle.$$

Passing onto the limit, and employing the inequalities $\sigma_1 < \sigma_2$ and (2.63), we obtain that

$$\langle f'(x^k + \bar{\alpha}p^k), p^k \rangle \geq \sigma_1 \langle f'(x^k), p^k \rangle > \sigma_2 \langle f'(x^k), p^k \rangle. \quad (2.74)$$

On the other hand, passing onto the limit in (2.71) results in the inequality

$$\langle f'(x^k + \bar{\alpha}p^k), p^k \rangle \leq \sigma_2 \langle f'(x^k), p^k \rangle,$$

which is in a contradiction with (2.74). \square

We conclude this section by mentioning the so-called nonmonotone linesearch methods; see [110]. Allowing an increase of the objective function value on some iterations, these methods tend to produce longer steps. Roughly speaking, the choice of α_k in nonmonotone methods is based on comparison of $f(x^k + \alpha p^k)$ not with $f(x^k)$ but rather with the maximum (or average) value of f along some fixed number of previous iterations. There is computational evidence that such methods may be more efficient in some applications than the usual descent methods.

2.2.2 Quasi-Newton Methods

From now on, we consider descent methods of the specific form

$$x^{k+1} = x^k - \alpha_k Q_k f'(x^k), \quad \alpha_k > 0, \quad k = 0, 1, \dots, \quad (2.75)$$

where for each k , $Q_k \in \mathbf{R}^{n \times n}$ is a symmetric positive definite matrix, and the stepsize parameter α_k is chosen by linesearch.

Algorithm 2.23 Choose the parameters $C > 0$, $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $f'(x^k) = 0$, stop.
2. Choose a symmetric positive definite matrix $Q_k \in \mathbf{R}^{n \times n}$, and compute α_k according to the Armijo rule, employing the direction $p^k = -Q_k f'(x^k)$.
3. Set $x^{k+1} = x^k - \alpha_k Q_k f'(x^k)$.
4. Increase k by 1 and go to step 1.

We first show that the algorithm possesses global convergence (in a certain sense) to stationary points of problem (2.58).

Theorem 2.24. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be differentiable on \mathbf{R}^n , and suppose that its gradient is Lipschitz-continuous on \mathbf{R}^n . Assume further that there exist $\gamma > 0$ and $\Gamma > 0$ such that the matrices Q_k in Algorithm 2.23 satisfy condition (2.65).*

Then for any starting point $x^0 \in \mathbf{R}^n$, Algorithm 2.23 generates an iterative sequence $\{x^k\}$ such that each of its accumulation points is a stationary point of problem (2.58). Moreover, if an accumulation point exists, or if f is bounded below on \mathbf{R}^n , then

$$\{f'(x^k)\} \rightarrow 0. \quad (2.76)$$

Proof. The fact that Algorithm 2.23 is well defined follows from Lemma 2.19. Moreover (under the standing assumption that $f'(x^k) \neq 0 \forall k$), the sequence $\{f(x^k)\}$ is monotonically decreasing.

If the sequence $\{x^k\}$ has an accumulation point $\bar{x} \in \mathbf{R}^n$, then $f(\bar{x})$ is an accumulation point of $\{f(x^k)\}$, by the continuity of f . In this case, monotonicity of $\{f(x^k)\}$ implies that the whole sequence $\{f(x^k)\}$ converges to $f(\bar{x})$. If f is bounded below, then the monotone sequence $\{f(x^k)\}$ is bounded below. In this case, $\{f(x^k)\}$ converges even when $\{x^k\}$ does not have any accumulation points.

Since $p^k = -Q_k f'(x^k)$, by the Armijo rule, taking into account Lemma 2.21 and the first inequality in (2.65), we obtain that for all k it holds that

$$f(x^k) - f(x^{k+1}) \geq \sigma \alpha_k \langle Q_k f'(x^k), f'(x^k) \rangle \geq \sigma c \gamma \|f'(x^k)\|^2, \quad (2.77)$$

where $c > 0$ is the constant in the right-hand side of (2.66). Since the left-hand side in the relation above tends to zero as $k \rightarrow \infty$, we conclude that (2.76) holds. The assertion follows. \square

Somewhat more subtle analysis allows to replace Lipschitz-continuity of the gradient of f on the entire \mathbf{R}^n (which is rather restrictive) by simple continuity. The difficulty here is, of course, that under this weaker assumption one cannot guarantee that the values of the stepsize parameter are bounded away from zero.

Theorem 2.25. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be continuously differentiable on \mathbf{R}^n . Assume further that there exist $\gamma > 0$ and $\Gamma > 0$ such that the matrices Q_k in Algorithm 2.23 satisfy condition (2.65).*

Then for any starting point $x^0 \in \mathbf{R}^n$ Algorithm 2.23 generates an iterative sequence $\{x^k\}$ such that each of its accumulation points is a stationary point of problem (2.58). Moreover, if the sequence $\{x^k\}$ is bounded, then (2.76) holds.

Proof. The fact that Algorithm 2.23 is well defined follows from Lemma 2.19, as before. Suppose that the sequence $\{x^k\}$ has an accumulation point $\bar{x} \in \mathbf{R}^n$, and let a subsequence $\{x^{k_j}\}$ be convergent to \bar{x} as $j \rightarrow \infty$. The case when the corresponding subsequence $\{\alpha_{k_j}\}$ is bounded away from zero is dealt with the same way as in Theorem 2.24 (the only difference is that $\{x^k\}$ in the argument should be replaced by $\{x^{k_j}\}$). Therefore, we consider the case when $\{\alpha_{k_j}\} \rightarrow 0$ as $j \rightarrow \infty$.

In the latter case, for each j large enough, in the process of backtracking when computing α_{k_j} the initial trial value C was reduced at least once, which means that the value $\alpha = \alpha_{k_j}/\theta$ had been tried and found not to satisfy the Armijo inequality (2.62), i.e.,

$$f\left(x^{k_j} - \frac{\alpha_{k_j}}{\theta} Q_{k_j} f'(x^{k_j})\right) > f(x^{k_j}) - \sigma \frac{\alpha_{k_j}}{\theta} \langle Q_{k_j} f'(x^{k_j}), f'(x^{k_j}) \rangle.$$

Denoting $\tilde{\alpha}_{k_j} = \alpha_{k_j} \|Q_{k_j} f'(x^{k_j})\|/\theta$ and $\tilde{p}^{k_j} = -Q_{k_j} f'(x^{k_j})/\|Q_{k_j} f'(x^{k_j})\|$, the last inequality can be written in the form

$$f(x^{k_j} + \tilde{\alpha}_{k_j} \tilde{p}^{k_j}) > f(x^{k_j}) + \sigma \tilde{\alpha}_{k_j} \langle f'(x^{k_j}), \tilde{p}^{k_j} \rangle. \quad (2.78)$$

Recalling the second inequality in (2.65), we conclude that $\{\tilde{\alpha}_{k_j}\} \rightarrow 0$ as $j \rightarrow \infty$. Extracting further subsequences if necessary, we may assume that $\{\tilde{p}^{k_j}\}$ converges to some $\tilde{p} \in \mathbf{R}^n \setminus \{0\}$. With these observations, employing the mean-value theorem (see Theorem A.10, (a)), dividing both sides of (2.78) by $\tilde{\alpha}_{k_j}$ and passing onto the limit as $j \rightarrow \infty$, we obtain the inequality

$$\langle f'(\bar{x}), \tilde{p} \rangle \geq \sigma \langle f'(\bar{x}), \tilde{p} \rangle,$$

which implies that $\langle f'(\bar{x}), \tilde{p} \rangle \geq 0$. Then, by (2.65),

$$\begin{aligned} 0 &\geq -\langle f'(\bar{x}), \tilde{p} \rangle \\ &= \lim_{j \rightarrow \infty} \langle f'(x^{k_j}), -\tilde{p}^{k_j} \rangle \\ &= \lim_{j \rightarrow \infty} \frac{\langle Q_{k_j} f'(x^{k_j}), f'(x^{k_j}) \rangle}{\|Q_{k_j} f'(x^{k_j})\|} \\ &\geq \lim_{j \rightarrow \infty} \frac{\gamma \|f'(x^{k_j})\|^2}{\Gamma \|f'(x^{k_j})\|} \\ &= \frac{\gamma}{\Gamma} \|f'(\bar{x})\|, \end{aligned}$$

which is possible only when $f'(\bar{x}) = 0$.

The last assertion of the theorem can be easily derived from the assertion proven above. \square

We note that for Algorithm 2.23 with the Armijo linesearch rule replaced by Goldstein or Wolfe rules, global convergence statements analogous to Theorem 2.25 can be obtained.

Observe that neither Theorem 2.24 nor Theorem 2.25 claims the existence of accumulation points for iterative sequences of Algorithm 2.23. However, the latter is evidently guaranteed when f is coercive, since any sequence $\{x^k\} \subset \mathbf{R}^n$ generated by any descent method for problem (2.58) is contained in the level set $\{x \in \mathbf{R}^n \mid f(x) \leq f(x^0)\}$.

The results presented above suggest to try to combine, within a single algorithm, the attractive global convergence properties of descent methods with high convergence rate of Newton-type methods. For that purpose, the Newton-type method should be modified by introducing a stepsize parameter α_k computed by an appropriate linesearch rule. If this rule allows for the full Newton-type step near a qualified solution (i.e., the value $\alpha_k = 1$ is accepted for all k large enough), one can expect that high convergence rate of the Newton-type method would be inherited by the globalized algorithm. At the same time, far from solutions, full Newton-type steps can be too long to ensure monotone decrease of the sequence of the objective function values (and, as a consequence, convergence may not be guaranteed). Far from a solution, the step should therefore be shortened when necessary (i.e., $\alpha_k = 1$ should be reduced). The rest of this section is devoted to formal development of this idea.

Generally, Algorithm 2.23 is referred to as a *quasi-Newton method* for problem (2.58) if, assuming convergence of its iterative sequence to a solution \bar{x} , the directions $Q_k f'(x^k)$ approximate Newton directions $(f''(x^k))^{-1} f'(x^k)$ in the sense of the *Dennis–Moré* [57, 58] *condition* (2.80) (or (2.81); cf. (2.32)) stated below. We remark that it is quite natural to discuss quasi-Newton methods for unconstrained optimization in the context of linesearch methods, as it is possible to ensure positive definiteness of Hessian approximations when using some specific quasi-Newton update formulas and the Wolfe rule for computing the stepsize. The resulting algorithms thus fall within the class of descent methods.

The following version of the Dennis–Moré Theorem deals with a linesearch quasi-Newton method, for which the acceptance of full stepsize can be established rather than assumed (see also Theorem 2.29 below).

Theorem 2.26. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its second derivative being continuous at \bar{x} . Let \bar{x} be a stationary point of problem (2.58). Let $\{x^k\}$ be an iterative sequence of Algorithm 2.23, where $C = 1$ and $\sigma \in (0, 1/2)$, and assume that $\{x^k\}$ converges to \bar{x} .*

If the rate of convergence of $\{x^k\}$ is superlinear, then the condition

$$(\alpha_k Q_k - (f''(x^k))^{-1})f'(x^k) = o(\|f'(x^k)\|)$$

holds as $k \rightarrow \infty$.

Conversely, if \bar{x} satisfies the SOSC

$$\langle f''(\bar{x})\xi, \xi \rangle > 0 \quad \forall \xi \in \mathbf{R}^n \setminus \{0\}, \quad (2.79)$$

and the condition

$$(Q_k - (f''(x^k))^{-1})f'(x^k) = o(\|f'(x^k)\|) \quad (2.80)$$

holds as $k \rightarrow \infty$, then $\alpha_k = 1$ for all k large enough, and the rate of convergence of $\{x^k\}$ to \bar{x} is superlinear.

Remark 2.27. It can be easily checked that under the assumptions of Theorem 2.26, condition (2.80) is equivalent to

$$(Q_k^{-1} - f''(x^k))(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|). \quad (2.81)$$

Proof. According to Theorem 2.9 and the fact stated in Remark 2.27, we only need to prove that $\alpha_k = 1$ for all k large enough provided (2.79) and (2.80) hold.

From (2.80) and from the convergence of $\{x^k\}$ to \bar{x} , it evidently follows that

$$Q_k f'(x^k) = O(\|f'(x^k)\|) \quad (2.82)$$

as $k \rightarrow \infty$.

By the mean-value theorem for scalar-valued functions (see Theorem A.10, (a)), for each k there exists $\tilde{t}_k \in (0, 1)$ such that

$$\begin{aligned} f(x^k - Q_k f'(x^k)) &= f(x^k) - \langle f'(x^k), Q_k f'(x^k) \rangle \\ &\quad + \frac{1}{2} \langle f''(\tilde{x}^k) Q_k f'(x^k), Q_k f'(x^k) \rangle, \end{aligned}$$

where $\tilde{x}^k = x^k - \tilde{t}_k Q_k f'(x^k)$. It suffices to show that for all k large enough

$$\langle f'(x^k), Q_k f'(x^k) \rangle - \frac{1}{2} \langle f''(\tilde{x}^k) Q_k f'(x^k), Q_k f'(x^k) \rangle \geq \sigma \langle f'(x^k), Q_k f'(x^k) \rangle,$$

that is,

$$(1 - \sigma) \langle f'(x^k), Q_k f'(x^k) \rangle - \frac{1}{2} \langle f''(\tilde{x}^k) Q_k f'(x^k), Q_k f'(x^k) \rangle \geq 0. \quad (2.83)$$

Note that $\{\tilde{x}^k\} \rightarrow \bar{x}$, because $\{x^k\} \rightarrow \bar{x}$ and $\{Q_k f'(x^k)\} \rightarrow 0$ (the latter relation is an immediate consequence of (2.82) and of $\{x^k\} \rightarrow \bar{x}$). According to (2.80) and (2.82), we then derive that

$$\begin{aligned}\langle f'(x^k), Q_k f'(x^k) \rangle &= \langle f'(x^k), (f''(x^k))^{-1} f'(x^k) \rangle + o(\|f'(x^k)\|^2) \\ &= \langle f'(x^k), (f''(\bar{x}))^{-1} f'(x^k) \rangle + o(\|f'(x^k)\|^2),\end{aligned}$$

and

$$\begin{aligned}\langle f''(\bar{x}^k) Q_k f'(x^k), Q_k f'(x^k) \rangle &= \langle f''(x^k) Q_k f'(x^k), Q_k f'(x^k) \rangle \\ &\quad + o(\|f'(x^k)\|^2) \\ &= \langle f'(x^k), Q_k f'(x^k) \rangle + o(\|f'(x^k)\|^2) \\ &= \langle f'(x^k), (f''(x^k))^{-1} f'(x^k) \rangle + o(\|f'(x^k)\|^2) \\ &= \langle f'(x^k), (f''(\bar{x}))^{-1} f'(x^k) \rangle + o(\|f'(x^k)\|^2),\end{aligned}$$

where the nonsingularity of $f''(\bar{x})$ was taken into account. Hence,

$$\begin{aligned}(1 - \sigma) \langle f'(x^k), Q_k f'(x^k) \rangle &- \frac{1}{2} \langle f''(x^k) Q_k f'(x^k), Q_k f'(x^k) \rangle \\ &= \left(\frac{1}{2} - \sigma \right) \langle (f''(\bar{x}))^{-1} f'(x^k), f'(x^k) \rangle + o(\|f'(x^k)\|^2).\end{aligned}$$

The latter implies that (2.83) holds for all k large enough, as $\sigma \in (0, 1/2)$ and $(f''(\bar{x}))^{-1}$ is positive definite (by positive definiteness of $f''(\bar{x})$). \square

We note that for Algorithm 2.23 with the Armijo linesearch rule replaced by the Goldstein rule (with $0 < \sigma_1 < 1/2 < \sigma_2 < 1$) or the Wolfe rule (with $0 < \sigma_1 < 1/2$, $\sigma_1 < \sigma_2 < 1$) with the initial trial value of the stepsize parameter $\alpha = 1$, results analogous to Theorem 2.26 can be established.

In Theorem 2.26, convergence of the iterates is *assumed*. To obtain a complete result affirming global and locally superlinear convergence, it remains to show that under the assumptions of Theorem 2.24 on global convergence, if the iterates enter a neighborhood of a solution satisfying the SOSC (2.79), then they converge to this solution. Then, if the sequence of matrices $\{Q_k\}$ satisfies the Dennis–Moré condition, Theorem 2.26 guarantees that the rate of convergence is superlinear.

Before stating the needed result, we prove the following local growth property for the norm of the gradient of the objective function, complementing the quadratic growth property in Theorem 1.9.

Lemma 2.28. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be differentiable near $\bar{x} \in \mathbf{R}^n$ and twice differentiable at \bar{x} . Let \bar{x} be a stationary point of problem (1.10) satisfying the SOSC (1.14) or, equivalently, satisfying*

$$f(x) - f(\bar{x}) \geq \rho \|x - \bar{x}\|^2 \quad \forall x \in U \quad (2.84)$$

for some neighborhood U of \bar{x} and some $\rho > 0$.

Then for any $\nu \in (0, 4)$, there exists a neighborhood $V \subset U$ of \bar{x} such that

$$\|f'(x)\|^2 \geq \nu \rho (f(x) - f(\bar{x})) \quad \forall x \in V. \quad (2.85)$$

Proof. Indeed, for $x \in \mathbf{R}^n$ we have that

$$f'(x) = f'(x) - f'(\bar{x}) = f''(\bar{x})(x - \bar{x}) + o(\|x - \bar{x}\|)$$

as $x \rightarrow \bar{x}$, so that

$$\begin{aligned} f(x) - f(\bar{x}) &= \frac{1}{2} \langle f''(\bar{x})(x - \bar{x}), x - \bar{x} \rangle + o(\|x - \bar{x}\|^2) \\ &= \frac{1}{2} \langle f'(x), x - \bar{x} \rangle + o(\|x - \bar{x}\|^2), \end{aligned}$$

i.e.,

$$\langle f'(x), x - \bar{x} \rangle = 2(f(x) - f(\bar{x})) + o(\|x - \bar{x}\|^2).$$

Using (2.84) from Theorem 1.9, we then obtain that for all $x \in U$ close enough to \bar{x} it holds that

$$\begin{aligned} \langle f'(x), x - \bar{x} \rangle - \sqrt{\nu}(f(x) - f(\bar{x})) &= (2 - \sqrt{\nu})(f(x) - f(\bar{x})) + o(\|x - \bar{x}\|^2) \\ &\geq (2 - \sqrt{\nu})\rho\|x - \bar{x}\|^2 + o(\|x - \bar{x}\|^2) \\ &\geq 0, \end{aligned}$$

and therefore,

$$\langle f'(x), x - \bar{x} \rangle \geq \sqrt{\nu}(f(x) - f(\bar{x})). \quad (2.86)$$

Combining the latter inequality again with (2.84), we obtain that

$$\|f'(x)\|\|x - \bar{x}\| \geq \langle f'(x), x - \bar{x} \rangle \geq \sqrt{\nu}(f(x) - f(\bar{x})) \geq \sqrt{\nu}\rho\|x - \bar{x}\|^2,$$

i.e.,

$$\|f'(x)\| \geq \sqrt{\nu}\rho\|x - \bar{x}\|.$$

Hence, using (2.86),

$$\|f'(x)\|^2 \geq \sqrt{\nu}\rho\|x - \bar{x}\|\|f'(x)\| \geq \sqrt{\nu}\rho\langle f'(x), x - \bar{x} \rangle \geq \nu\rho(f(x) - f(\bar{x})),$$

which completes the proof. \square

Theorem 2.29. *Suppose that the assumptions of Theorem 2.24 are satisfied. Assume, in addition, that $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is twice differentiable at $\bar{x} \in \mathbf{R}^n$, which is a stationary point of problem (2.58) satisfying the SOSC (2.79).*

Then if on some iteration k Algorithm 2.23 generates an iterate x^k close enough to \bar{x} , it holds that the whole sequence $\{x^k\}$ converges to \bar{x} , and the rate of convergence is (at least) geometric.

Moreover, if f is twice differentiable in a neighborhood of \bar{x} , with its second derivative being continuous at \bar{x} , and if in Algorithm 2.23 we take $C = 1$, $\sigma \in (0, 1/2)$, and $\{Q_k\}$ satisfying the Dennis–Moré condition (2.80), then the convergence rate is superlinear.

Proof. By Theorem 1.9 and Lemma 2.28, there exists a neighborhood U of \bar{x} such that the growth conditions

$$f(x) - f(\bar{x}) \geq \rho \|x - \bar{x}\|^2 \quad \forall x \in U \quad (2.87)$$

and

$$\|f'(x)\|^2 \geq \nu \rho (f(x) - f(\bar{x})) \quad \forall x \in U \quad (2.88)$$

hold with some $\rho > 0$ and $\nu \in (0, 4)$. Note also that, by (2.87), it holds that

$$\begin{aligned} \|f'(x)\| &= \|f'(x) - f'(\bar{x})\| \leq L \|x - \bar{x}\| \\ &\leq L \sqrt{\frac{f(x) - f(\bar{x})}{\rho}} \quad \forall x \in U, \end{aligned} \quad (2.89)$$

where $L > 0$ is a Lipschitz constant of the gradient of f .

From Lemma 2.21, it follows that (2.77) holds, where $c > 0$ is the constant in the right-hand side of (2.66). Suppose that $x^k \in U$ for some k . Then by (2.77) and (2.88), we have that

$$\begin{aligned} f(x^{k+1}) - f(\bar{x}) &\leq f(x^k) - f(\bar{x}) - \sigma c \gamma \|f'(x^k)\|^2 \\ &\leq (1 - \sigma c \gamma \nu \rho) (f(x^k) - f(\bar{x})) \\ &= q (f(x^k) - f(\bar{x})), \end{aligned} \quad (2.90)$$

where $q = 1 - \sigma c \gamma \nu \rho < 1$.

We next show that if x^k is close enough to \bar{x} , then all the subsequent iterates do not leave the neighborhood U of \bar{x} . Fix $r > 0$ such that $B(\bar{x}, r) \subset U$, and define $\delta > 0$ satisfying

$$\delta + \frac{LC \sqrt{(f(x) - f(\bar{x}))/\rho}}{1 - \sqrt{|q|}} \leq r \quad \forall x \in B(\bar{x}, \delta), \quad (2.91)$$

where $C > 0$ is the first trial stepsize value in the Armijo rule. Note that $\alpha_k \leq C$ and $\delta \leq r$. Let $x^k \in B(\bar{x}, \delta)$. In this case, by (2.89) and (2.91),

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &\leq \|x^k - \bar{x}\| + \|x^{k+1} - x^k\| \\ &\leq \delta + C \|f'(x^k)\| \\ &\leq \delta + LC \sqrt{\frac{f(x^k) - f(\bar{x})}{\rho}} \\ &\leq r, \end{aligned}$$

i.e., $x^{k+1} \in B(\bar{x}, r)$. From this, using also (2.87), it follows that $q \geq 0$ (as otherwise, (2.90) would not hold).

Suppose that $x^j \in B(\bar{x}, r) \quad \forall j = k, \dots, s$, for some integer $s \geq k$. Then, by (2.90), we have that

$$\begin{aligned}
f(x^j) - f(\bar{x}) &\leq q(f(x^{j-1}) - f(\bar{x})) \\
&\vdots \\
&\leq q^{j-k}(f(x^k) - f(\bar{x})) \quad \forall j = k, \dots, s.
\end{aligned}$$

Therefore, using also (2.89), we have that

$$\begin{aligned}
\|x^{j+1} - x^j\| &\leq C\|f'(x^j)\| \\
&\leq LC\sqrt{\frac{f(x^j) - f(\bar{x})}{\rho}} \\
&\leq (\sqrt{q})^{j-k} LC\sqrt{\frac{f(x^k) - f(\bar{x})}{\rho}} \quad \forall j = k, \dots, s.
\end{aligned}$$

From the latter and (2.91), it follows that

$$\begin{aligned}
\|x^{s+1} - \bar{x}\| &\leq \|x^s - \bar{x}\| + \|x^{s+1} - x^s\| \\
&\vdots \\
&\leq \|x^k - \bar{x}\| + \sum_{l=k}^s \|x^{l+1} - x^l\| \\
&\leq \delta + LC\sqrt{\frac{f(x^k) - f(\bar{x})}{\rho}} \sum_{l=k}^s (\sqrt{q})^{l-k} \\
&\leq \delta + \frac{LC\sqrt{(f(x^k) - f(\bar{x}))/\rho}}{1 - \sqrt{q}} \\
&\leq r,
\end{aligned}$$

i.e., $x^{s+1} \in B(\bar{x}, r)$.

We have thus established that $x^j \in U$ for all $j = k, k+1, \dots$. In particular, (2.90) holds for all k (large enough), which shows that $\{f(x^k)\}$ converges to $f(\bar{x})$ at a linear rate. Then (2.87) implies that $\{x^k\}$ converges to \bar{x} geometrically.

Superlinear convergence rate under the Dennis–Moré condition (2.80) now follows from Theorem 2.26. \square

As a direct consequence of Theorem 2.29, we obtain that the usual Newton method with linesearch (for which the Dennis–Moré condition is automatic) is superlinearly convergent whenever its sequence enters a neighborhood of a minimizer satisfying SOSC.

The crucial conclusion from the Dennis–Moré Theorem is that in order to construct a fast optimization method, it is *indispensable* to employ the “second-order information” about the problem either explicitly, or to construct objects describing second-order behavior using first-order information. For the basic Newton method (see Sect. 2.1.2), condition (2.80) is

automatic provided the sequence $\{x^k\}$ converges. However, beyond the case of a strongly convex f , there are no reasons to expect $f''(x^k)$ (and hence, $Q_k = (f''(x^k))^{-1}$) to be positive definite for all k . It will indeed be positive definite for x^k close to a solution satisfying the SOSC (2.79) but, in general, not when x^k is far from such a solution. Moreover, even if $f''(x^k)$ is positive definite for all k but the Hessian of f is singular at some (other) points, there is still no guarantee that the iterative sequence would not get stuck near a point of degeneracy, which by no means has to be a stationary point of f . The question concerning the possibility of such behavior was posed in [89] and answered in the affirmative in [196], where an example of nonconvergence of the basic Newton method with the Wolfe linesearch rule is constructed.

The observations above indicate that the basic Newtonian choice of Q_k may be inadequate from the point of view of global convergence, even when exact Hessians are available. Perhaps even more importantly, it turns out that the needed “second-order information” can be constructed without direct computation of Hessians. The main idea of quasi-Newton methods is to completely avoid computing $f''(x^k)$ and solving the corresponding linear system, and instead to approximate the Newton step itself in the sense of the Dennis–Moré condition (2.80). It is important that this approximation does not subsume that $\|Q_k - (f''(x^k))^{-1}\| \rightarrow 0$ as $k \rightarrow \infty$ and, in fact, this relation indeed does not hold for quasi-Newton methods (in general). The needed approximations must be computed according to some recursive formulas, without using any information about the second derivative of f . Fortunately, such construction can be accomplished, and in many ways.

For each k , define

$$s^k = x^{k+1} - x^k, \quad r^k = f'(x^{k+1}) - f'(x^k). \quad (2.92)$$

Note that these two vectors are already known by the time when Q_{k+1} should be computed, and the goal to achieve (2.81) (which is equivalent to (2.80)) can be modeled as the *quasi-Newton equation*

$$Q_{k+1}r^k = s^k. \quad (2.93)$$

Taking into account that $s^k = -Q_k f'(x^k)$ (by (2.75)), the motivation behind (2.93) is the same as behind the quasi-Newton equation (2.33) for systems of equations; see Sect. 2.1.1.

Therefore, having at hand a symmetric positive definite matrix Q_k and vectors r^k and s^k , it is suggested to choose a symmetric positive definite matrix Q_{k+1} satisfying the quasi-Newton equation (2.93). However, such a choice would be clearly not unique. As in the case of systems of equations, it is natural to additionally require the difference between Q_k and Q_{k+1} to be “minimal” in some sense: from one iteration to another, the variation of Q_k should not be too large. Similarly to quasi-Newton methods for systems

of equations, a natural approach is to define Q_{k+1} as a *symmetric* matrix minimizing some matrix norm of $Q_{k+1} - Q_k$ or $Q_{k+1}^{-1} - Q_k^{-1}$. Different norms lead to different specific quasi-Newton methods.

Historically, the first quasi-Newton method is the *Davidon–Fletcher–Powell* (DFP) *method*, in which Q_0 is an arbitrary symmetric positive definite matrix (e.g., $Q_0 = I$), and for each k

$$Q_{k+1} = Q_k + \frac{s^k (s^k)^T}{\langle r^k, s^k \rangle} - \frac{(Q_k r^k)(Q_k r^k)^T}{\langle Q_k r^k, r^k \rangle}. \quad (2.94)$$

Note that the matrices generated this way remain symmetric and satisfy the quasi-Newton equation (2.93):

$$\begin{aligned} Q_{k+1} r^k &= Q_k r^k + s^k \frac{\langle r^k, s^k \rangle}{\langle r^k, s^k \rangle} - Q_k r^k \frac{\langle Q_k r^k, r^k \rangle}{\langle Q_k r^k, r^k \rangle} \\ &= Q_k r^k + s^k - Q_k r^k = s^k. \end{aligned}$$

Moreover, the corresponding Q_{k+1}^{-1} minimizes the weighted Frobenius norm of the correction $Q_{k+1}^{-1} - Q_k^{-1}$ over all the symmetric matrices $Q_{k+1} \in \mathbf{R}^{n \times n}$ satisfying the quasi-Newton equation (2.93); see, e.g., [208, Sect. 11.1] for details. Furthermore, the correction $Q_{k+1} - Q_k$ is a matrix whose rank cannot be greater than 2 (since $\ker(Q_{k+1} - Q_k)$ contains all vectors orthogonal to both r^k and $Q_k s^k$), so the correction is “small” in this sense as well.

Regarding positive definiteness of Q_{k+1} , this depends not only on the quasi-Newton formula used for computing this matrix but also on the choice of the stepsize parameter in (2.75). Specifically, we have the following.

Proposition 2.30. *Let $Q_k \in \mathbf{R}^{n \times n}$ be a symmetric positive definite matrix, and let $s^k, r^k \in \mathbf{R}^n$.*

Then formula (2.94) is well defined and the matrix Q_{k+1} is positive definite if and only if the following inequality holds:

$$\langle r^k, s^k \rangle > 0. \quad (2.95)$$

Proof. The necessity follows immediately from the quasi-Newton equation (2.93), according to which

$$\langle r^k, s^k \rangle = \langle Q_{k+1} r^k, r^k \rangle.$$

Note also that formula (2.94) is not well defined when $r^k = 0$.

We proceed with sufficiency. From (2.95) it follows that $r^k \neq 0$, and hence, positive definiteness of Q_k implies the inequality $\langle Q_k r^k, r^k \rangle > 0$. Combining the latter with (2.95), we obtain that the matrix Q_{k+1} is well defined.

Furthermore, for an arbitrary $\xi \in \mathbf{R}^n$, by (2.94) we derive

$$\begin{aligned}
\langle Q_{k+1}\xi, \xi \rangle &= \langle Q_k\xi, \xi \rangle + \frac{\langle s^k, \xi \rangle^2}{\langle r^k, s^k \rangle} - \frac{\langle Q_k r^k, \xi \rangle^2}{\langle Q_k r^k, r^k \rangle} \\
&= \frac{\langle s^k, \xi \rangle^2}{\langle r^k, s^k \rangle} + \frac{\|Q_k^{1/2}\xi\|^2 \|Q_k^{1/2}s^k\|^2 - \langle Q_k^{1/2}\xi, Q_k^{1/2}r^k \rangle^2}{\|Q_k^{1/2}r^k\|^2},
\end{aligned}$$

where both terms in the right-hand side are nonnegative, according to (2.95) and the Cauchy–Schwarz inequality. Moreover, the equality $\langle Q_{k+1}\xi, \xi \rangle = 0$ may hold only when both terms above are equal to zero, i.e., when

$$\langle s^k, \xi \rangle = 0, \quad (2.96)$$

and

$$\|Q_k^{1/2}\xi\| \|Q_k^{1/2}r^k\| = |\langle Q_k^{1/2}\xi, Q_k^{1/2}r^k \rangle|.$$

The second equality means that $Q_k^{1/2}\xi = tQ_k^{1/2}r^k$ with some $t \in \mathbf{R}$, and since $Q_k^{1/2}$ is nonsingular, this leads to the equality $\xi = tr^k$. Then, by (2.96), $t\langle r^k, s^k \rangle = \langle \xi, s^k \rangle = 0$, and according to (2.95), the latter is possible only when $t = 0$, i.e., when $\xi = 0$. \square

In particular, the inequality (2.95) is always valid if the stepsize parameter in (2.75) is chosen according to the Wolfe rule, while the Armijo rule and the Goldstein rule do not possess this property. This is one of the reasons why the Wolfe rule is recommended for quasi-Newton methods.

Currently, the *Broyden–Fletcher–Goldfarb–Shanno* (BFGS) method is regarded as the most efficient general purpose quasi-Newton method. For each k , it defines

$$\begin{aligned}
Q_{k+1} &= Q_k + \frac{(r^k - Q_k s^k)(r^k)^T + r^k(r^k - Q_k s^k)^T}{\langle r^k, s^k \rangle} \\
&\quad - \frac{\langle r^k - Q_k s^k, s^k \rangle r^k (r^k)^T}{\langle r^k, s^k \rangle^2}. \quad (2.97)
\end{aligned}$$

It can be immediately verified that (as for the DFP method) the matrices generated according to this formula remain symmetric and satisfy the quasi-Newton equation (2.93), and the rank of corrections $Q_{k+1} - Q_k$ cannot be greater than 2. Moreover, it can be shown that this Q_{k+1} minimizes the weighted Frobenius norm of the correction $Q_{k+1} - Q_k$ over all symmetric matrices $Q_{k+1} \in \mathbf{R}^{n \times n}$ satisfying quasi-Newton equation (2.93); see [208, Sect. 11.1]. For a recent survey of variational origins of the DFP and the BFGS updates, see [111].

Remark 2.31. It can be easily checked that the DFP and BFGS methods can be regarded as “dual” with respect to each other in the following sense. For any symmetric positive definite matrix $Q_k \in \mathbf{R}^{n \times n}$, set $H_k = Q_k^{-1}$. Let Q_{k+1} be generated according to (2.97), let H_{k+1} be generated according to the formula

$$H_{k+1} = H_k + \frac{s^k(s^k)^T}{\langle r^k, s^k \rangle} - \frac{(H_k r^k)(H_k r^k)^T}{\langle H_k r^k, r^k \rangle}$$

(cf. (2.94)), and suppose that the matrix H_{k+1} is nonsingular. Then the matrix Q_{k+1} is also nonsingular, and $H_{k+1} = Q_{k+1}^{-1}$. From this fact it immediately follows that a counterpart of Proposition 2.30 is valid for the BFGS method as well.

It can be shown that for a quadratic function f with a positive definite Hessian, if α_k is chosen according to the exact linesearch rule, then the DFP and BFGS methods find the unique critical point of f (which is the global solution of problem (2.58), by necessity) from any starting point after no more than $k \leq n$ iterations. Moreover, Q_k would coincide with the inverse of the Hessian of f ; see, e.g., [18, 19] for details. Recall that for quadratic functions, the exact linesearch rule reduces to an explicit formula. In the non-quadratic case, convergence and rate of convergence results for the DFP and BFGS methods can be found, e.g., in [29, 89, 208]. This analysis is highly nontrivial and is concerned with overcoming serious technical difficulties. In particular, the condition (2.65) is not automatic for the DFP and BFGS methods, and in order to apply Theorems 2.24 or 2.25 one has to verify (2.65), which normally requires some additional assumptions. Here, we present only some general comments.

Known (full) global convergence results for the DFP and BFGS methods are concerned with the case of convex f . The theory of quasi-Newton methods for nonconvex problems is far from being complete, though rich numerical practice puts in evidence that these methods are highly efficient in the nonconvex case as well (especially the BFGS).

Of course, in the non-quadratic case, one cannot expect finite termination of quasi-Newton methods at a solution. However, the rate of convergence usually remains very high. Proving superlinear convergence of a specific quasi-Newton method reduces to the (usually highly nontrivial) verification of the Dennis–Moré condition (2.80), and application of Theorem 2.26.

Quasi-Newton methods are very popular among the users of optimization methods, because they combine high convergence rate with low computational cost per iteration. It is difficult to overestimate the practical value of these methods. For general principles of constructing and analyzing quasi-Newton methods, see [19, 89, 208].

2.2.3 Other Linesearch Methods

Let us briefly mention some other ideas for developing linesearch Newton-type methods, different from the quasi-Newton class. One possibility is to take Q_k as the inverse matrix of a positive definite modification of the Hessian $f''(x^k)$, when the latter is not (sufficiently) positive definite. Specifically, in

the process of factorization of $f''(x^k)$, one could replace this matrix by a positive definite matrix if needed, say, of the form $f''(x^k) + \nu_k I$, with the regularization parameter $\nu_k > 0$ large enough. More details on methods of this kind can be found in [208, Sect. 3.4]; for recent contributions, including smart rules for controlling ν_k , see, e.g., [262] and references therein. Note also that instead of modifying the Hessian this way every time when it is not (sufficiently) positive definite, one may choose to do so only when the generated direction p^k is not a descent direction or is not a “good enough” descent direction (positive definiteness of Q_k is sufficient but not necessary for $p^k = -Q_k f'(x^k)$ to be a direction of descent). Other possibilities arise from combining the linesearch strategy with perturbed Newton methods considered in Sect. 2.1.1, and in particular, with the truncated Newton method; see [208, Sect. 7.1] and the corresponding discussion for systems of equations below. See also Sects. 5.1, 6.2, where various ideas of linesearch globalization are applied to constrained optimization and more general variational problems.

We complete this section with a brief discussion of how linesearch can be used to globalize Newton-type methods for a system of equations presented in Sect. 2.1.1. We note that the underlying principles for problems which are more general, or different from unconstrained optimization, are often quite similar nevertheless—the task is to construct an appropriate *merit function* that measures the quality of approximation to a solution in some sense, and for which the direction of a given Newton-type method is of descent, so that linesearch can be applied. We shall get back to those issues more systematically in Sect. 5.1.

Consider the equation

$$\Phi(x) = 0, \quad (2.98)$$

where $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is a smooth mapping. In this case, it is natural to measure the quality of approximation by a given point $x \in \mathbf{R}^n$ of a solution of (2.98) by the value of the residual $\|\Phi(x)\|$ or, to preserve smoothness, by the value of $\|\Phi(x)\|^2$. Thus, from the optimization point of view, while solving (2.98) we are trying to minimize the merit function $f : \mathbf{R}^n \rightarrow \mathbf{R}_+$ given by

$$f(x) = \frac{1}{2} \|\Phi(x)\|^2. \quad (2.99)$$

It is therefore natural to globalize Newton method for (2.98) by introducing linesearch for this merit function f . With this choice,

$$f'(x) = (\Phi'(x))^T \Phi(x), \quad x \in \mathbf{R}^n,$$

and if $\bar{x} \in \mathbf{R}^n$ is a critical point of f , and $\Phi'(\bar{x})$ is a nonsingular matrix, then \bar{x} is a solution of the equation (2.98). On the other hand, if $x \in \mathbf{R}^n$ is not a solution of (2.98), and $\Phi'(x)$ is a nonsingular matrix, then the basic Newtonian direction

$$p = -(\Phi'(x))^{-1} \Phi(x)$$

is a descent direction for the function f at x . This follows from Lemma 2.18, because

$$\langle f'(x), p \rangle = -\langle (\Phi'(x))^T \Phi(x), (\Phi'(x))^{-1} \Phi(x) \rangle = -\|\Phi(x)\|^2 < 0.$$

Moreover, it holds that

$$p = -(\Phi'(x))^{-1}((\Phi'(x))^T)^{-1}(\Phi'(x))^T \Phi(x) = -Q(x)f'(x), \quad (2.100)$$

where the matrix

$$Q(x) = ((\Phi'(x))^T \Phi'(x))^{-1} \quad (2.101)$$

is symmetric and positive definite. Therefore, the corresponding linesearch method can be written in the form (2.75).

Furthermore, the perturbed Newtonian direction has the form

$$p = -(\Phi'(x))^{-1}(\Phi(x) + \omega), \quad (2.102)$$

with some $\omega \in \mathbf{R}^n$ (see (2.11)). Then

$$\begin{aligned} \langle f'(x), p \rangle &= -\langle (\Phi'(x))^T \Phi(x), (\Phi'(x))^{-1}(\Phi(x) + \omega) \rangle \\ &\leq -\|\Phi(x)\|^2 + \|\Phi(x)\|\|\omega\| \\ &\leq -\|\Phi(x)\|(\|\Phi(x)\| - \|\omega\|), \end{aligned}$$

which is negative provided $\|\omega\| < \|\Phi(x)\|$. Hence, according to Lemma 2.18, in this case p is still a descent direction for f at x . In particular, this will always be the case for a truncated Newton method, characterized by (2.26), if the forcing sequence satisfies $\{\theta_k\} \subset [0, \theta]$ with some $\theta \in (0, 1)$. More precisely, if $\|\omega\| \leq \theta\|\Phi(x)\|$, then

$$\langle f'(x), p \rangle \leq -(1 - \theta)\|\Phi(x)\|^2 < 0.$$

As another example of a perturbed Newtonian direction, consider

$$p = -(\Phi'(x) + \Omega)^{-1} \Phi(x), \quad (2.103)$$

with some $\Omega \in \mathbf{R}^{n \times n}$; see (2.36). Such p satisfies (2.102) with

$$\begin{aligned} \omega &= -\Phi(x) + \Phi'(x)(\Phi'(x) + \Omega)^{-1} \Phi(x) \\ &= (\Phi'(x)(\Phi'(x) + \Omega)^{-1} - I) \Phi(x) \\ &= -\Omega(\Phi'(x) + \Omega)^{-1} \Phi(x). \end{aligned}$$

According to the discussion above, it then follows that p is a descent direction for f at x provided $\|\Omega\| \|(\Phi'(x) + \Omega)^{-1}\| < 1$. Employing Lemma A.6, one can easily check that the latter condition is satisfied if $\|\Omega\| < \|(\Phi'(x))^{-1}\|^{-1}/2$.

Note, however, that perturbations resulting from quasi-Newton approximations J of $\Phi'(x)$ may not fit this framework: since such J need not be close to $\Phi'(x)$, there is no reason to expect the corresponding direction

$$p = -J^{-1}\Phi'(x)$$

to be a descent direction for f at \bar{x} . This is actually a serious limitation of quasi-Newton methods in the context of nonlinear equations. Observe that this limitation does not exist in the context of optimization, where the objective function serves as a natural merit function.

Note further that writing the Newtonian direction in the form (2.100), (2.101), the associated method can be developed for $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^m$ with m possibly bigger than n . Assuming that $\ker \Phi'(x) = \{0\}$ holds for a given $x \in \mathbf{R}^n$ (which generalizes the assumption that $\Phi'(x)$ is nonsingular to the case of $m \geq n$), this direction is well defined. It is, in fact, the direction produced by the step of the Gauss–Newton method (see (2.41)). According to the discussion above, since $Q(x)$ is positive definite, the resulting linesearch method fits the scheme (2.75).

Thus, one can develop descent methods for systems of equations combining the Newton-type methods from Sect. 2.1.1 with linesearch procedures for the function f defined in (2.99). It should be mentioned, however, that such methods can get stuck near points of singularity of the Jacobian of Φ which need not be solutions of (2.98), in general. In fact, they may not even be critical points of the function f . The possibility of such an undesirable scenario is demonstrated in [35]. In particular, even if the iterative sequence $\{x^k\}$ does not hit any point x where $\ker \Phi'(x) \neq \{0\}$, this sequence can accumulate around such points. Consequently, there is no guarantee that the second inequality in (2.65) will be satisfied with any fixed $\Gamma > 0$ by $Q_k = Q(x^k)$ defined according to (2.101). Therefore, there is no guarantee of global convergence under any reasonable assumptions, unless it is assumed that $\ker \Phi'(x) \neq \{0\} \forall x \in \mathbf{R}^n$, which is generally too much to expect.

It should be emphasized, however, that appropriate perturbations of the Newtonian directions may not only not harm the descent properties, but in fact can significantly improve them, as well as the overall global behavior of the corresponding linesearch methods. If $n = m$, one can use perturbed directions of the form (2.103), where Ω is selected in such a way that it moves $\Phi'(x) + \Omega$ away from singularity. However, suggesting some general practical rules for defining such Ω could be rather difficult.

The situation changes if we consider the Gauss–Newton direction, which can be naturally perturbed as follows:

$$p = -((\Phi'(x))^T \Phi'(x) + \nu I)^{-1} (\Phi'(x))^T \Phi(x), \quad (2.104)$$

where $\nu \geq 0$ is the regularization parameter. This direction can be regarded as a blend of the pure Gauss–Newton direction, corresponding to $\nu = 0$, and the (anti)gradient direction for the function f defined in (2.99), to which p turns asymptotically as ν tends to $+\infty$.

Algorithms employing directions of the form (2.104) are known under the common name of *Levenberg–Marquardt methods*. Note that these directions can be computed by solving the unconstrained optimization problem

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|\Phi'(x)p + \Phi(x)\|^2 + \frac{\nu}{2} \|p\|^2 \\ & \text{subject to} && p \in \mathbf{R}^n, \end{aligned}$$

whose quadratic objective function is strongly convex provided $\nu > 0$. Needless to say, a smart control of the parameter ν is essential for the efficiency of this approach. From the global convergence viewpoint, it is desirable to select ν so that the eigenvalues of $(\Phi'(x))^T \Phi'(x) + \nu I$ be separated away from zero by some constant not depending on x . On the other hand, in order to eventually achieve high convergence rate, it is essential that ν must tend to zero as x approaches a solution (so that the method eventually fits the local perturbed Newton framework of Sect. 2.1). An important practical rule for choosing ν arises in the context of trust-region methods; see Sect. 2.3. Observe also that in (2.104), the use of quasi-Newton approximations instead of the exact $\Phi'(x)$ can be considered.

2.3 Trust-Region Methods

Apart from linesearch, another natural strategy to globalize a local method is the following. As discussed above, an iteration of a Newton-type method consists of minimizing a quadratic approximation of the objective function, where the approximation is computed at the current iterate x^k . It is intuitively clear that this approximation can only be “trusted” locally, in some neighborhood of x^k . It is therefore reasonable to minimize the approximation in question in the neighborhood where it is trusted, and not on the whole space. Those considerations give rise to the so-called *trust-region methods*.

Consider again the unconstrained problem

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && x \in \mathbf{R}^n, \end{aligned} \tag{2.105}$$

where $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is differentiable (twice differentiable in the case of trust-region Newton methods). Let $x^k \in \mathbf{R}^n$ be the current iterate. As discussed earlier, subproblem of the Newton method for problem (2.105) is given by

$$\begin{aligned} & \text{minimize} && \psi_k(x) \\ & \text{subject to} && x \in \mathbf{R}^n, \end{aligned} \tag{2.106}$$

where

$$\psi_k(x) = f(x^k) + \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle H_k(x - x^k), x - x^k \rangle, \tag{2.107}$$

with $H_k = f''(x^k)$ for the basic choice. Having in mind possible approximations of the Hessian, we shall also consider the case when H_k is a generic matrix, not necessarily $f''(x^k)$. An iteration of the trust-region version of the method consists in solving (possibly approximately) the subproblem

$$\begin{aligned} & \text{minimize} && \psi_k(x) \\ & \text{subject to} && \|x - x^k\| \leq \delta_k, \end{aligned} \quad (2.108)$$

where $\delta_k > 0$ is the parameter defining a region (in this case, a ball) where the model ψ_k is trusted to be a good approximation of f . Intelligent control of this parameter is clearly the principal issue in this approach.

In what follows, we shall consider only quadratic models ψ_k . Much more general choices are possible, see [45]. In fact, the linesearch approach can also be viewed from the trust-region perspective, making the “extreme” choice of the simple one-dimensional model given by restricting the objective function to the direction of linesearch.

Trust-region methods operate as follows. If a (possibly approximate) solution of the subproblem (2.108) provides a sufficient decrease of the objective function (with respect to what is predicted by the model, see (2.109)), it is accepted as the next iterate. Otherwise, the trust-region parameter is reduced, essentially for similar reasons as reducing the stepsize in linesearch procedures. A very basic version of a trust-region method is the following.

Algorithm 2.32 Choose the parameters $C > 0$ and $\sigma, \theta \in (0, 1)$, and a starting point $x^0 \in \mathbf{R}^n$. Set $k = 0$.

1. Choose $\delta \geq C$ and a symmetric matrix $H_k \in \mathbf{R}^{n \times n}$.
2. For $\delta_k = \delta$, compute $\tilde{x}^k \in \mathbf{R}^n$, a (possibly approximate) solution of (2.108).
3. If $\tilde{x}^k = x^k$, stop. Otherwise, check the inequality

$$f(\tilde{x}^k) - f(x^k) \leq \sigma(\psi_k(\tilde{x}^k) - f(x^k)). \quad (2.109)$$

- If (2.109) is satisfied, go to step 4; otherwise, set $\delta = \theta\delta_k$ and go to step 2.
4. Set $x^{k+1} = \tilde{x}^k$, increase k by 1 and go to step 1.

We note that it is very important in practice to allow the trust-region radius to increase after successful iterations. While step 1 in Algorithm 2.32 allows this, we shall not go into details of how exactly this should be done. Similarly, when the trust-region is reduced, more sophisticated rules than the simple one in step 3 are often used. We refer the readers to [45] for all the rich details.

Any comparison of efficiency of linesearch and trust-region methods is a complicated issue. In principle, both strategies have some advantages and disadvantages. One could state that trust-region methods became ever more popular over the years. What is clear is that trust-region methods are more robust. Indeed, it is enough to think of the situation when $f'(x^k) = 0$ but there exists a “second-order” descent direction for f from the point x^k , called

direction of negative curvature, i.e., some $\xi \in \mathbf{R}^n$ such that $\langle f''(x^k)\xi, \xi \rangle < 0$. In such a case, linesearch in the direction $p^k = -Q_k f'(x^k) = 0$ would not make any progress, no matter which matrix Q_k to choose. On the other hand, it is clear that solving the trust-region subproblem (2.108) with $H_k = f''(x^k)$, a point different from x^k would be obtained, and it can be seen that the value of f at this point is smaller than $f(x^k)$; see Proposition 2.37 below. This difference is reflected even in the theoretical properties of the methods: while linesearch descent methods converge to first-order stationary points (see Theorems 2.24, 2.25), the trust-region Newton method converges to second-order stationary points (see Theorem 2.38).

Of course, a trust-region iteration is generally more costly than simple linesearch, especially considering that more than one subproblem of the form (2.108) may need to be solved per iteration (when the trust-region parameter δ_k has to be reduced). However, very efficient special methods have been developed for this purpose [45, 208]. Also, good control rules for δ_k usually avoid the need of solving many subproblems per iteration.

We shall not go into much detail concerning specific methods for solving the trust-region subproblems. But some comments are in order. The following characterization of global solutions of (2.108), due to [202], is the key to the construction.

Proposition 2.33. *For any $g \in \mathbf{R}^n$, any symmetric matrix $H \in \mathbf{R}^{n \times n}$, and any $\delta > 0$, a point $\tilde{\xi}$ is a global solution of the problem*

$$\begin{aligned} & \text{minimize} && \langle g, \xi \rangle + \frac{1}{2} \langle H\xi, \xi \rangle \\ & \text{subject to} && \|\xi\| \leq \delta \end{aligned} \quad (2.110)$$

if and only if there exists $\nu \in \mathbf{R}$ such that

$$g + (H + \nu I)\tilde{\xi} = 0, \quad \nu \geq 0, \quad \|\tilde{\xi}\| \leq \delta, \quad \nu(\|\tilde{\xi}\| - \delta) = 0, \quad (2.111)$$

and the matrix $H + \nu I$ is positive semidefinite.

Proof. Denote the objective function of problem (2.110) by

$$q(\xi) = \langle g, \xi \rangle + \frac{1}{2} \langle H\xi, \xi \rangle, \quad \xi \in \mathbf{R}^n.$$

Suppose first that $\tilde{\xi}$ is a global solution of problem (2.110). If we rewrite the constraint in (2.110) in the equivalent smooth form $\|\xi\|^2/2 \leq \delta^2/2$, we observe that it satisfies the LICQ at every feasible point, and the existence of ν satisfying (2.111) follows from the KKT optimality conditions of Theorem 1.14. Moreover, the same holds even for a local solution $\tilde{\xi}$ of problem (2.110), with ν uniquely defined by (2.111).

It remains to show that $H + \nu I$ is positive semidefinite. If $\|\tilde{\xi}\| < \delta$, this follows from the SONC given by Theorem 1.19. (Note, in the passing, that in

this case the last condition in (2.111) implies that $\nu = 0$. Hence, H is positive semidefinite. Thus, in this case, (2.110) is a convex problem. In particular, any local solution of this problem is actually global.)

We proceed with the case when $\|\tilde{\xi}\| = \delta$. Since $\tilde{\xi}$ is a global solution of problem (2.110), for any $\xi \in \mathbf{R}^n$ satisfying $\|\xi\| = \delta$ we obtain that

$$\begin{aligned} \langle g, \tilde{\xi} \rangle + \frac{1}{2} \langle (H + \nu I) \tilde{\xi}, \tilde{\xi} \rangle &= q(\tilde{\xi}) + \frac{\nu}{2} \|\tilde{\xi}\|^2 \\ &= q(\tilde{\xi}) + \frac{\nu}{2} \delta^2 \\ &\leq q(\xi) + \frac{\nu}{2} \delta^2 \\ &= q(\xi) + \frac{\nu}{2} \|\xi\|^2 \\ &= \langle g, \xi \rangle + \frac{1}{2} \langle (H + \nu I) \xi, \xi \rangle. \end{aligned}$$

Employing the first relation in (2.111), we then derive that

$$-\frac{1}{2} \langle (H + \nu I) \tilde{\xi}, \tilde{\xi} \rangle \leq -\langle (H + \nu I) \tilde{\xi}, \xi \rangle + \frac{1}{2} \langle (H + \nu I) \xi, \xi \rangle,$$

which can be further transformed to

$$\langle (H + \nu I)(\xi - \tilde{\xi}), \xi - \tilde{\xi} \rangle \geq 0.$$

Evidently, the closure of the cone spanned by all $\xi - \tilde{\xi}$ with $\|\xi\| = \delta$ is a half-space defined by a hyperplane containing 0 and orthogonal to $\tilde{\xi}$. Thus, the quadratic form defined by $H + \nu I$ is nonnegative on this half-space, which implies that it is nonnegative on the entire space. Thus, $H + \nu I$ is positive semidefinite.

Suppose now that (2.111) holds with some ν , and the matrix $H + \nu I$ is positive semidefinite. Then $\tilde{\xi}$ is an unconstrained global minimizer of the convex quadratic function $\tilde{q} : \mathbf{R}^n \rightarrow \mathbf{R}$ defined by

$$\tilde{q}(\xi) = \langle g, \xi \rangle + \frac{1}{2} \langle (H + \nu I) \xi, \xi \rangle = q(\xi) + \frac{\nu}{2} \|\xi\|^2,$$

because $\tilde{q}'(\tilde{\xi}) = g + (H + \nu I)\tilde{\xi} = 0$. Therefore, for any $\xi \in \mathbf{R}^n$ satisfying $\|\xi\| \leq \delta$ it holds that

$$q(\tilde{\xi}) \leq q(\xi) + \frac{\nu}{2} (\|\xi\|^2 - \|\tilde{\xi}\|^2) = q(\xi) + \frac{\nu}{2} (\|\xi\|^2 - \delta^2) \leq q(\xi),$$

where the equality is by the last relation in (2.111). Thus, $\tilde{\xi}$ is a global solution of problem (2.110). \square

Let H_k be positive definite, and define

$$p(\nu) = -(H_k + \nu I)^{-1} f'(x^k), \quad \nu \geq 0.$$

According to Proposition 2.33, solution of the subproblem (2.108) is given by $\tilde{x}^k = x^k + p(\nu_k)$ for a certain $\nu_k \geq 0$. If $\|H_k^{-1} f'(x^k)\| \leq \delta_k$, then $\nu_k = 0$; otherwise, $\nu_k \geq 0$ is uniquely defined by the equation

$$\|p(\nu)\| = \delta_k.$$

In particular, computing \tilde{x}^k reduces to the search for the appropriate $\nu_k \geq 0$ (for reasons of conditioning, usually the equivalent equation $1/\|p(\nu)\| = 1/\delta_k$ is solved).

Instead of solving the trust-region subproblems for different values of δ_k , one can try to approximate the solutions curve $p(\cdot)$ directly. Note that changes in δ_k are inverse to the changes in ν , and varying the latter $p(\nu)$ moves in a continuous manner from the Newtonian step $p(0) = -H_k^{-1} f'(x^k)$ to the step in the almost steepest descent direction: $p(\nu) \approx -f'(x^k)/\nu$ for large values $\nu > 0$. The so-called *dog-leg strategy* consists in minimizing the model ψ_k along the piecewise linear trajectory connecting the current iterate x^k , the minimizer x_U^k of ψ_k in the direction of the steepest descent, and the unconstrained minimizer x_N^k of ψ_k (the Newton point), subject to the trust-region constraint. This is illustrated in Fig. 2.6, where the point computed by the dog-leg strategy is denoted by \tilde{x}^k and the exact solution of the trust-region subproblem (2.108) for the given $\delta_k > 0$ is \bar{x}^k .

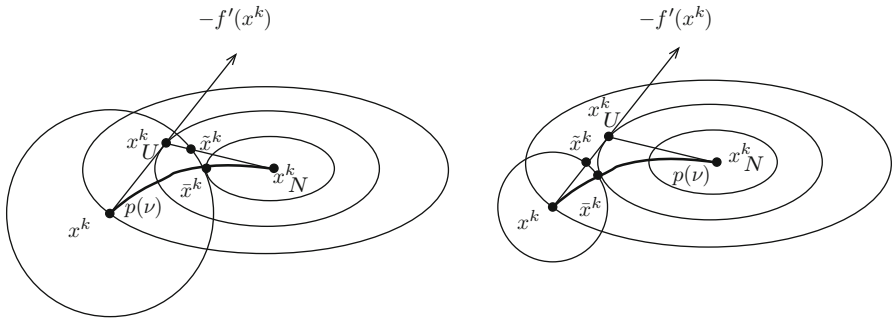


Fig. 2.6 The point computed by the dog-leg strategy

The theory concerning inexact solution of trust-region subproblems is based on the so-called *Cauchy point*, which is the minimizer of the model ψ_k in the direction of steepest descent within the trust-region. Specifically, let $x_C^k = x^k - \alpha_k f'(x^k)$, where $\alpha_k \in \mathbf{R}_+$ is the solution of the problem

$$\begin{aligned} & \text{minimize} && q_k(\alpha) \\ & \text{subject to} && \alpha \|f'(x^k)\| \leq \delta_k, \alpha \geq 0, \end{aligned} \quad (2.112)$$

with the objective function $q_k : \mathbf{R} \rightarrow \mathbf{R}$ defined by

$$q_k(\alpha) = \psi_k(x^k - \alpha f'(x^k)).$$

It turns out that any point \tilde{x}^k that decreases the model ψ_k at least as much as the Cauchy point x_C^k can be employed as an approximate solution of the trust-region subproblem (2.108) in Algorithm 2.32. If in the given implementation x_C^k is not computed, the direct verification of the property $\psi_k(\tilde{x}^k) \leq \psi_k(x_C^k)$ is not possible. The following result gives a bound of the progress obtained by the Cauchy point; this bound is always readily available and can serve as a benchmark for judging whether a candidate for approximate solution of the trust-region subproblem (2.108) is acceptable.

Proposition 2.34. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be differentiable at $x^k \in \mathbf{R}^n$. Then the Cauchy point x_C^k satisfies*

$$f(x^k) - \psi_k(x_C^k) \geq \frac{1}{2} \|f'(x^k)\| \min \left\{ \delta_k, \frac{\|f'(x^k)\|}{1 + \|H_k\|} \right\}. \quad (2.113)$$

Proof. If $f'(x^k) = 0$, the statement holds trivially. Let $f'(x^k) \neq 0$. Note that

$$q_k(\alpha) = f(x^k) - \alpha \|f'(x^k)\|^2 + \frac{\alpha^2}{2} \langle H_k f'(x^k), f'(x^k) \rangle.$$

If $\langle H_k f'(x^k), f'(x^k) \rangle \leq 0$, then the function q_k is strictly decreasing on \mathbf{R}_+ . It follows that the solution α_k of (2.112) is on the boundary of the trust-region, i.e., $\alpha_k = \delta_k / \|f'(x^k)\|$. Hence, according to (2.107),

$$\begin{aligned} f(x^k) - \psi_k(x_C^k) &= \alpha_k \|f'(x^k)\|^2 - \frac{\alpha_k^2}{2} \langle H_k f'(x^k), f'(x^k) \rangle \\ &\geq \alpha_k \|f'(x^k)\|^2 \\ &= \delta_k \|f'(x^k)\|, \end{aligned}$$

which verifies that (2.113) holds in this case.

Suppose that $\langle H_k f'(x^k), f'(x^k) \rangle > 0$. In this case the function q_k is strongly convex and has the unique unconstrained global minimizer

$$\bar{\alpha}_k = \frac{\|f'(x^k)\|^2}{\langle H_k f'(x^k), f'(x^k) \rangle}.$$

If $\bar{\alpha}_k \leq \delta_k / \|f'(x^k)\|$, then $\alpha_k = \bar{\alpha}_k$. In that case,

$$\begin{aligned}
f(x^k) - \psi_k(x_C^k) &= \bar{\alpha}_k \|f'(x^k)\|^2 - \frac{\bar{\alpha}_k^2}{2} \langle H_k f'(x^k), f'(x^k) \rangle \\
&= \frac{\|f'(x^k)\|^4}{2 \langle H_k f'(x^k), f'(x^k) \rangle} \\
&\geq \frac{\|f'(x^k)\|^2}{2 \|H_k\|},
\end{aligned}$$

which again verifies (2.113). Finally, if $\bar{\alpha}_k > \delta_k / \|f'(x^k)\|$, then

$$\alpha_k = \frac{\delta_k}{\|f'(x^k)\|} < \frac{\|f'(x^k)\|^2}{\langle H_k f'(x^k), f'(x^k) \rangle},$$

so that

$$\alpha_k^2 \langle H_k f'(x^k), f'(x^k) \rangle < \alpha_k \|f'(x^k)\|^2 = \delta_k \|f'(x^k)\|.$$

Hence,

$$\begin{aligned}
f(x^k) - \psi_k(x_C^k) &= \alpha_k \|f'(x^k)\|^2 - \frac{\alpha_k^2}{2} \langle H_k f'(x^k), f'(x^k) \rangle \\
&\geq \delta_k \|f'(x^k)\| - \frac{1}{2} \delta_k \|f'(x^k)\| \\
&= \frac{1}{2} \delta_k \|f'(x^k)\|,
\end{aligned}$$

which verifies (2.113) in the last case and completes the proof. \square

The next statement shows that if each trust-region subproblem (2.108) is solved to the precision enough to achieve the (computable) bound of reduction guaranteed by the Cauchy point, then the method is well defined. Specifically, a new iterate with a lower objective function value is obtained after reducing the trust-region radius at most a finite number of times.

Let $\gamma \in (0, 1/2]$. For fixed $x^k \in \mathbf{R}^n$ and $H_k \in \mathbf{R}^{n \times n}$, and for each $\delta > 0$, let $\tilde{x}(\delta)$ be some (any) approximate solution of the subproblem (2.108) for $\delta_k = \delta$, satisfying $\|\tilde{x}(\delta) - x^k\| \leq \delta$ and

$$f(x^k) - \psi_k(\tilde{x}(\delta)) \geq \gamma \|f'(x^k)\| \min \left\{ \delta, \frac{\|f'(x^k)\|}{1 + \|H_k\|} \right\}. \quad (2.114)$$

Points satisfying this property certainly exist, by Proposition 2.34. Moreover, they exist on the piecewise linear dog-leg trajectory, because the Cauchy point itself belongs to this trajectory (see Fig. 2.6, where in the first case x_C^k is the point x_U^k and in the second case it is the dog-leg point \tilde{x}^k). Note also that if $\tilde{x}(\delta) = x^k$ so that Algorithm 2.32 stops, by (2.114) we have that $f'(x^k) = 0$.

Proposition 2.35. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be differentiable at $x^k \in \mathbf{R}^n$. Suppose that $f'(x^k) \neq 0$.*

Then if in Algorithm 2.32 approximate solutions $\tilde{x}^k = \tilde{x}(\delta)$ of trust-region subproblems (2.108) satisfy $\|\tilde{x}(\delta) - x^k\| \leq \delta$ and (2.114) with some $\gamma > 0$ and with $\delta = \delta_k$, an iterate x^{k+1} is generated such that $f(x^{k+1}) < f(x^k)$.

Proof. Since the right-hand side of (2.114) is positive, the assertion would be valid if we show that the acceptance criterion (2.109) is guaranteed to be satisfied after a finite number of reductions of the trust-region radius δ . Because in (2.109) we have $\sigma \in (0, 1)$, it is evidently enough to prove that

$$\lim_{\delta \rightarrow 0+} \rho(\delta) = 1, \quad (2.115)$$

where the function $\rho : \mathbf{R}_+ \rightarrow \mathbf{R}$ is given by

$$\rho(\delta) = \frac{f(\tilde{x}(\delta)) - f(x^k)}{\psi_k(\tilde{x}(\delta)) - f(x^k)}. \quad (2.116)$$

Note that

$$|\rho(\delta) - 1| = \left| \frac{f(\tilde{x}(\delta)) - \psi_k(\tilde{x}(\delta))}{f(x^k) - \psi_k(\tilde{x}(\delta))} \right|.$$

We have that

$$\begin{aligned} f(\tilde{x}(\delta)) - \psi_k(\tilde{x}(\delta)) &= f(\tilde{x}(\delta)) - f(x^k) - \langle f'(x^k), \tilde{x}(\delta) - x^k \rangle \\ &\quad - \frac{1}{2} \langle H_k(\tilde{x}(\delta) - x^k), \tilde{x}(\delta) - x^k \rangle \\ &= o(\delta) \end{aligned}$$

as $\delta \rightarrow 0$. Combining the latter relation with (2.114), we obtain that

$$|\rho(\delta) - 1| = \frac{o(\delta)}{\delta},$$

which concludes the proof. \square

We are now in position to show convergence of the algorithm to stationary points of the problem.

Theorem 2.36. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be differentiable on \mathbf{R}^n and let its gradient be Lipschitz-continuous on \mathbf{R}^n . Suppose that there exist $\Gamma > 0$ and $\gamma > 0$ such that for all k in Algorithm 2.32 we have that $\|H_k\| \leq \Gamma$, and approximate solutions $\tilde{x}^k = \tilde{x}(\delta_k)$ of the trust-region subproblems satisfy $\|\tilde{x}^k - x^k\| \leq \delta_k$ and (2.114) with $\delta = \delta_k$.*

Then each accumulation point of any sequence $\{x^k\}$ generated by Algorithm 2.32 is a stationary point of problem (2.105).

Proof. By (2.109), we obtain that

$$\begin{aligned}
f(x^k) - f(x^{k+1}) &\geq \sigma(f(x^k) - \psi_k(x^{k+1})) \\
&\geq \gamma\sigma\|f'(x^k)\| \min\left\{\delta_k, \frac{\|f'(x^k)\|}{1+\Gamma}\right\}, \quad (2.117)
\end{aligned}$$

where the second inequality is by (2.114).

Let \bar{x} be any accumulation point of the sequence $\{x^k\}$. Then $\{f(x^k)\}$ also has an accumulation point, by the continuity of f . Since by (2.117) the sequence $\{f(x^k)\}$ is nonincreasing, it follows that it converges. Hence, $\{f(x^k) - f(x^{k+1})\} \rightarrow 0$ as $k \rightarrow \infty$.

Suppose that there exists some $\varepsilon > 0$ such that

$$\|f'(x^k)\| \geq \varepsilon \quad \forall k. \quad (2.118)$$

Then (2.117) implies that

$$\delta_k \rightarrow 0 \text{ as } k \rightarrow \infty. \quad (2.119)$$

For any k and any approximate solution \tilde{x}^k of the trust-region subproblem (2.108), let

$$\rho_k = \frac{f(\tilde{x}^k) - f(x^k)}{\psi_k(\tilde{x}^k) - f(x^k)} \quad (2.120)$$

and let $L > 0$ be the Lipschitz constant of the gradient of f . We have that

$$\begin{aligned}
|f(\tilde{x}^k) - \psi_k(\tilde{x}^k)| &\leq |f(\tilde{x}^k) - f(x^k) - \langle f'(x^k), \tilde{x}^k - x^k \rangle| \\
&\quad + \frac{1}{2} |\langle H_k(\tilde{x}^k - x^k), \tilde{x}^k - x^k \rangle| \\
&\leq \frac{(L + \Gamma)\delta_k^2}{2},
\end{aligned}$$

where Lemma A.11 was used. Employing the latter relation and (2.114), we obtain that

$$\begin{aligned}
|\rho_k - 1| &= \left| \frac{f(\tilde{x}^k) - \psi_k(\tilde{x}^k)}{f(x^k) - \psi_k(\tilde{x}^k)} \right| \\
&\leq \frac{(L + \Gamma)\delta_k^2}{2\gamma\|f'(x^k)\| \min\{\delta_k, \|f'(x^k)\|/(1 + \|H_k\|)\}} \\
&\leq \frac{(L + \Gamma)\delta_k^2}{2\gamma\varepsilon \min\{\delta_k, \varepsilon/(1 + \Gamma)\}} \\
&= \frac{(L + \Gamma)\delta_k}{2\gamma\varepsilon} \\
&\leq 1 - \sigma,
\end{aligned}$$

where the second inequality is by (2.118), and the last two relations are valid for all k large enough, once

$$\delta_k \leq \bar{\delta} = \min \left\{ \frac{\varepsilon}{1 + \Gamma}, \frac{2\gamma\varepsilon(1 - \sigma)}{L + \Gamma} \right\},$$

which holds by (2.119) for all k large enough. It follows that the acceptance criterion (2.109), i.e., $\rho_k \geq \sigma$, must hold for the first trust-region radius trial value that is less than $\bar{\delta} > 0$. This contradicts (2.119). We conclude that the hypothesis (2.118) is not valid. Hence,

$$\liminf_{k \rightarrow \infty} \|f'(x^k)\| = 0. \quad (2.121)$$

Suppose that

$$\limsup_{k \rightarrow \infty} \|f'(x^k)\| > 0,$$

i.e., there exist $\varepsilon > 0$ and a subsequence $\{x^{k_j}\}$ such that

$$\|f'(x^{k_j})\| \geq \varepsilon \quad \forall j.$$

For each j , let $k(j)$ be the first index $k > k_j$ such that $\|f'(x^k)\| \leq \varepsilon/2$ (this index exists due to (2.121)). Then it holds that

$$\frac{\varepsilon}{2} \leq \|f'(x^{k(j)}) - f'(x^{k_j})\| \leq L\|x^{k(j)} - x^{k_j}\| \quad \forall j.$$

Hence,

$$\frac{\varepsilon}{2L} \leq \|x^{k(j)} - x^{k_j}\| \leq \sum_{i=k_j}^{k(j)-1} \|x^{i+1} - x^i\| \leq \sum_{i=k_j}^{k(j)-1} \delta_i. \quad (2.122)$$

Using the same reasoning as in (2.117), and the fact that $\|f'(x^i)\| \geq \varepsilon/2$ for $i = k_j, \dots, k(j) - 1$ (by the definition of $k(j)$), we then obtain that

$$\begin{aligned} f(x^{k_j}) - \lim_{k \rightarrow \infty} f(x^k) &\geq f(x^{k_j}) - f(x^{k(j)}) \\ &= \sum_{i=k_j}^{k(j)-1} (f(x^i) - f(x^{i+1})) \\ &\geq \gamma\sigma \sum_{i=k_j}^{k(j)-1} \|f'(x^i)\| \min \left\{ \delta_i, \frac{\|f'(x^i)\|}{1 + \Gamma} \right\} \\ &\geq \frac{1}{2}\gamma\sigma\varepsilon \sum_{i=k_j}^{k(j)-1} \min \left\{ \delta_i, \frac{\varepsilon}{2(1 + \Gamma)} \right\} \\ &\geq \frac{1}{4}\gamma\sigma\varepsilon^2 \min \left\{ \frac{1}{L}, \frac{1}{1 + \Gamma} \right\}, \end{aligned}$$

where the fourth inequality follows from (2.122). Since the left-hand side in the relation above tends to zero as $i \rightarrow \infty$, we obtain a contradiction. This completes the proof. \square

We next consider the trust-region Newton method, i.e., Algorithm 2.32 with $H_k = f''(x^k)$ and exact solution of subproblems. First note that if for some iteration index k it holds in step 3 that $\tilde{x}^k = x^k$, then this point is stationary for problem (2.105); moreover, it satisfies the second-order necessary condition for optimality. Indeed, since in this case \tilde{x}^k lies in the interior of the feasible set of subproblem (2.108), it is an unconstrained local minimizer of the objective function of this subproblem, and we have that

$$f'(x^k) = f'(x^k) + f''(x^k)(\tilde{x}^k - x^k) = \psi'_k(\tilde{x}^k) = 0,$$

and the matrix

$$f''(x^k) = \psi''_k(\tilde{x}^k)$$

is positive semidefinite (see Theorems 1.7 and 1.8).

Proposition 2.37. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be twice differentiable at $x^k \in \mathbf{R}^n$. Suppose that either $f'(x^k) \neq 0$ or the matrix $f''(x^k)$ is not positive semidefinite.*

Then Algorithm 2.32 with $H_k = f''(x^k)$ and exact solution of subproblems generates an iterate x^{k+1} such that $f(x^{k+1}) < f(x^k)$.

Proof. For each $\delta > 0$, let $\tilde{x}(\delta)$ be some (any) global solution of the subproblem (2.108) with $\delta_k = \delta$, and let $\rho(\delta)$ be given by (2.116). As already seen above, $f'(x^k) \neq 0$ or $f''(x^k)$ not being positive semidefinite imply that $\tilde{x}(\delta) \neq x^k$. Therefore, in particular, it holds that

$$\psi_k(\tilde{x}(\delta)) < \psi_k(x^k) = f(x^k).$$

As in Proposition 2.35, it is sufficient to prove (2.115), as it implies that (2.109) holds for all δ small enough (in which case, an acceptable δ will be computed after a finite number of modifications in step 3 of Algorithm 2.32).

Suppose first that $f'(x^k) \neq 0$. Take any $\xi \in \mathbf{R}^n$ such that $\langle f'(x^k), \xi \rangle < 0$, $\|\xi\| = 1$. Since $x^k + \delta\xi$ is a feasible point of subproblem (2.108), we have that

$$\begin{aligned} \psi_k(\tilde{x}(\delta)) &\leq \psi_k(x^k + \delta\xi) \\ &= f(x^k) + \delta \langle f'(x^k), \xi \rangle + \frac{\delta^2}{2} \langle f''(x^k) \xi, \xi \rangle. \end{aligned}$$

Hence,

$$\begin{aligned} \psi_k(\tilde{x}(\delta)) - f(x^k) &\leq \delta \left(\langle f'(x^k), \xi \rangle + \frac{\delta}{2} \|f''(x^k)\| \right) \\ &\leq \frac{\delta}{2} \langle f'(x^k), \xi \rangle, \end{aligned} \tag{2.123}$$

where the last inequality holds for all $\delta > 0$ small enough.

On the other hand,

$$\begin{aligned} f(\tilde{x}(\delta)) - \psi_k(\tilde{x}(\delta)) &= f(\tilde{x}(\delta)) - f(x^k) - \langle f'(x^k), \tilde{x}(\delta) - x^k \rangle \\ &\quad - \frac{1}{2} \langle f''(x^k)(\tilde{x}(\delta) - x^k), \tilde{x}(\delta) - x^k \rangle \\ &= o(\delta^2). \end{aligned} \quad (2.124)$$

Combining (2.124) and (2.123), we obtain that

$$|\rho(\delta) - 1| = \left| \frac{f(\tilde{x}(\delta)) - \psi_k(\tilde{x}(\delta))}{\psi_k(\tilde{x}(\delta)) - f(x^k)} \right| = o(\delta)$$

as $\delta \rightarrow 0$, which implies (2.115).

Assume now that $f'(x^k) = 0$, but there exists $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$, such that $\langle f''(x^k)\xi, \xi \rangle < 0$. We have that

$$\psi_k(\tilde{x}(\delta)) \leq \psi_k(x^k + \delta\xi) = f(x^k) + \frac{\delta^2}{2} \langle f''(x^k)\xi, \xi \rangle.$$

Therefore,

$$\psi_k(\tilde{x}(\delta)) - f(x^k) \leq \frac{\delta^2}{2} \langle f''(x^k)\xi, \xi \rangle.$$

Combining this relation with (2.124), we obtain

$$|\rho(\delta) - 1| = \frac{o(\delta^2)}{\delta^2}$$

as $\delta \rightarrow 0$, implying (2.115). \square

As already mentioned, the trust-region Newton method has the property of convergence to second-order stationary points; this is the result that we prove next.

Theorem 2.38. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be twice continuously differentiable on \mathbf{R}^n .*

Then each accumulation point of any sequence $\{x^k\}$ generated by Algorithm 2.32, with $H_k = f''(x^k)$ and exact solution of subproblems, is a stationary point of problem (2.105). Moreover, it satisfies the SONC

$$\langle f''(\bar{x})\xi, \xi \rangle \geq 0 \quad \forall \xi \in \mathbf{R}^n, \quad (2.125)$$

stated in Theorem 1.8.

Proof. By Proposition 2.37, the sequence $\{f(x^k)\}$ is nondecreasing. Suppose that the sequence $\{x^k\}$ has an accumulation point $\bar{x} \in \mathbf{R}^n$. Then the monotone sequence $\{f(x^k)\}$ has an accumulation point $f(\bar{x})$, which means that it converges (to $f(\bar{x})$). In particular, it holds that $f(x^{k+1}) - f(x^k) \rightarrow 0$ as $k \rightarrow \infty$. From (2.109) we have that

$$0 \leq \sigma(f(x^k) - \psi_k(x^{k+1})) \leq f(x^k) - f(x^{k+1}),$$

which implies

$$\psi_k(x^{k+1}) - f(x^k) \rightarrow 0 \text{ as } k \rightarrow \infty. \quad (2.126)$$

Let $\{x^{k_j}\}$ be any subsequence of $\{x^k\}$ that converges to \bar{x} . There are two possibilities:

$$\liminf_{j \rightarrow \infty} \delta_{k_j} > 0, \quad (2.127)$$

or

$$\liminf_{j \rightarrow \infty} \delta_{k_j} = 0. \quad (2.128)$$

Consider the first case. Define

$$\bar{\delta} = \liminf_{j \rightarrow \infty} \delta_{k_j} > 0.$$

Let $\bar{y} \in \mathbf{R}^n$ be a global solution of the problem

$$\begin{aligned} & \text{minimize} \quad \langle f'(\bar{x}), x - \bar{x} \rangle + \frac{1}{2} \langle f''(\bar{x})(x - \bar{x}), x - \bar{x} \rangle \\ & \text{subject to} \quad \|x - \bar{x}\| \leq \frac{\bar{\delta}}{4}. \end{aligned} \quad (2.129)$$

For all indices j sufficiently large,

$$\|\bar{y} - x^{k_j}\| \leq \|\bar{y} - \bar{x}\| + \|x^{k_j} - \bar{x}\| \leq \frac{\bar{\delta}}{2} \leq \delta_{k_j}.$$

In particular, the point \bar{y} is feasible in the subproblem (2.108) for $k = k_j$. Therefore,

$$\begin{aligned} \psi_{k_j}(x^{k_j+1}) & \leq \psi_{k_j}(\bar{y}) \\ & = f(x^{k_j}) + \langle f'(x^{k_j}), \bar{y} - x^{k_j} \rangle + \frac{1}{2} \langle f''(x^{k_j})(\bar{y} - x^{k_j}), \bar{y} - x^{k_j} \rangle. \end{aligned}$$

Passing onto the limit when $j \rightarrow \infty$ and taking into account (2.126), we obtain that

$$0 \leq \langle f'(\bar{x}), \bar{y} - \bar{x} \rangle + \frac{1}{2} \langle f''(\bar{x})(\bar{y} - \bar{x}), \bar{y} - \bar{x} \rangle.$$

Since \bar{x} is obviously feasible in (2.129) and yields zero objective function value, the latter relation implies that the optimal value of (2.129) is exactly zero and \bar{x} is also its global solution. Since \bar{x} lies in the interior of the feasible set of problem (2.129), it is an unconstrained local minimizer of the objective function of this subproblem. Theorems 1.7 and 1.8 imply the assertions of the theorem (in the case of (2.127)).

Suppose now that (2.128) holds. Without loss of generality, we can assume that the whole sequence $\{\delta_{k_j}\}$ converges to zero. Then for all indices j large enough, on iterations indexed by k_j the trust-region parameter had been

reduced at least once in step 3 of the algorithm. Hence, for each such j , the value $\tilde{\delta}_{k_j} = \delta_{k_j}/\theta > \delta_{k_j} > 0$ did not satisfy (2.109), i.e., there exists a global solution $\tilde{x}(\tilde{\delta}_{k_j})$ of subproblem (2.108) with $k = k_j$, such that

$$f(\tilde{x}(\tilde{\delta}_{k_j})) > f(x^{k_j}) + \sigma(\psi_{k_j}(\tilde{x}(\tilde{\delta}_{k_j})) - f(x^{k_j})). \quad (2.130)$$

Note that since $\delta_{k_j} \rightarrow 0$, we have that $\tilde{\delta}_{k_j} \rightarrow 0$ and $\{\tilde{x}(\tilde{\delta}_{k_j}) - x^{k_j}\} \rightarrow 0$ as $j \rightarrow \infty$.

The rest of the proof follows the lines of that for Proposition 2.37. Suppose the SONC (2.125) does not hold, i.e., there exists $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$, such that

$$\langle f'(\bar{x}), \xi \rangle < 0, \quad (2.131)$$

or such that

$$f'(\bar{x}) = 0, \quad \langle f''(\bar{x})\xi, \xi \rangle < 0. \quad (2.132)$$

We then obtain that

$$\begin{aligned} \psi_{k_j}(\tilde{x}(\tilde{\delta}_{k_j})) &\leq \psi_{k_j}(x^{k_j} + \tilde{\delta}_{k_j}\xi) \\ &= f(x^{k_j}) + \tilde{\delta}_{k_j} \langle f'(x^{k_j}), \xi \rangle + \frac{\tilde{\delta}_{k_j}^2}{2} \langle f''(x^{k_j})\xi, \xi \rangle. \end{aligned}$$

In the case of (2.131), from the latter relation we conclude, for all j large enough, that

$$\begin{aligned} \psi_{k_j}(\tilde{x}(\tilde{\delta}_{k_j})) - f(x^{k_j}) &\leq \tilde{\delta}_{k_j} \left(\langle f'(x^{k_j}), \xi \rangle + \frac{\tilde{\delta}_{k_j}}{2} \|f''(x^{k_j})\| \right) \\ &\leq \frac{\tilde{\delta}_{k_j}}{2} \langle f'(\bar{x}), \xi \rangle. \end{aligned}$$

Defining

$$\rho_{k_j} = \frac{f(\tilde{x}(\tilde{\delta}_{k_j})) - f(x^{k_j})}{\psi_{k_j}(\tilde{x}(\tilde{\delta}_{k_j})) - f(x^{k_j})},$$

we have that

$$|\rho_{k_j} - 1| = \frac{o(\tilde{\delta}_{k_j}^2)}{\tilde{\delta}_{k_j}}$$

as $j \rightarrow \infty$. But then

$$\rho_{k_j} \rightarrow 1 \text{ as } j \rightarrow \infty, \quad (2.133)$$

in contradiction with (2.130).

In the case of (2.132), the relation (2.133) is obtained analogously (see the proof of Proposition 2.37). This again gives a contradiction with (2.130). \square

It remains to verify whether, under natural assumptions, the trust-region Algorithm 2.32 locally reduces to the pure Newton method, thus inheriting its fast convergence rate.

Theorem 2.39. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its second derivative being continuous at \bar{x} . Let \bar{x} be a stationary point of problem (2.105) that satisfies the SOSC*

$$\langle f''(\bar{x})\xi, \xi \rangle > 0 \quad \forall \xi \in \mathbf{R}^n \setminus \{0\}.$$

In Algorithm 2.32, let $H_k = f''(x^k)$ and the subproblems be solved exactly.

Then there exists a neighborhood U of \bar{x} such that if $x^k \in U$ for some k , then the next iterate x^{k+1} generated by Algorithm 2.32 coincides with the point $x^k - (f''(x^k))^{-1}f'(x^k)$, the sequence $\{x^k\}$ generated by Algorithm 2.32 converges to \bar{x} and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the Hessian of f is locally Lipschitz-continuous with respect to \bar{x} .

Proof. Under the stated assumptions, there exist a neighborhood U of \bar{x} , $\mu > 0$ and $M > 0$ such that if $x^k \in U$, then $f''(x^k)$ is positive definite,

$$\langle (f''(x^k))^{-1}\xi, \xi \rangle \geq \mu \|\xi\|^2 \quad \forall \xi \in \mathbf{R}^n, \quad \|(f''(x^k))^{-1}\| \leq M, \quad (2.134)$$

and in particular, the point $\tilde{x}^k = x^k - (f''(x^k))^{-1}f'(x^k)$ is well defined. Moreover, since this point satisfies $\psi'_k(\tilde{x}^k) = 0$, it is the unique unconstrained global minimizer of the strongly convex quadratic function ψ_k defined in (2.107). Furthermore, since $f'(\bar{x}) = 0$, the neighborhood U can be taken small enough, so that

$$\|\tilde{x}^k - x^k\| = \|(f''(x^k))^{-1}f'(x^k)\| \leq M\|f'(x^k)\| < C,$$

where the second relation in (2.134) was taken into account. In such a case, the global minimizer \tilde{x}^k of ψ_k is feasible in the trust-region subproblem (2.108) for any $\delta_k \geq C$. Recall that at the beginning of each iteration (in step 1 of Algorithm 2.32), we select $\delta \geq C$. Hence, \tilde{x}^k is the unique global solution of (2.108) for the initial choice of trust-region parameter $\delta_k = \delta$, and it remains to show that this point satisfies the sufficient descent condition (2.109), so that the initial choice of δ_k is never reduced.

To this end, observe that by (2.107) and by the definition of \tilde{x}^k , the following holds for x^k close enough to \bar{x} :

$$\begin{aligned}
& f(\tilde{x}^k) - f(x^k) - \sigma(\psi_k(\tilde{x}^k) - f(x^k)) \\
&= f(\tilde{x}^k) - f(x^k) - \sigma \left(\langle f'(x^k), \tilde{x}^k - x^k \rangle + \frac{1}{2} \langle f''(x^k)(\tilde{x}^k - x^k), \tilde{x}^k - x^k \rangle \right) \\
&= (1 - \sigma) \left(-\langle f'(x^k), (f''(x^k))^{-1} f'(x^k) \rangle \right. \\
&\quad \left. + \frac{1}{2} \langle f''(x^k)(f''(x^k))^{-1} f'(x^k), (f''(x^k))^{-1} f'(x^k) \rangle \right) + o(\|f'(x^k)\|^2) \\
&= -\frac{1 - \sigma}{2} \langle (f''(x^k))^{-1} f'(x^k), f'(x^k) \rangle + o(\|f'(x^k)\|^2) \\
&\leq -\frac{1 - \sigma}{2} \mu \|f'(x^k)\|^2 + o(\|f'(x^k)\|^2) \\
&\leq 0,
\end{aligned}$$

where the first inequality is by the first relation in (2.134).

We have thus shown that if a neighborhood U of \bar{x} is small enough and $x^k \in U$, then the point $x^k - (f''(x^k))^{-1} f'(x^k)$ is well defined and is accepted by Algorithm 2.32 as x^{k+1} . The assertions follow now from Theorem 2.15. \square

As already commented, the basic Newton model ψ_k defined in (2.107) can be replaced in trust-region subproblem (2.108) by other models, and in particular, by those originating from perturbed/modified Newtonian methods. As an example, let us recall the Gauss–Newton method for the equation

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

where $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is a smooth mapping, $n \leq m$. According to (2.41), for a given iterate $x^k \in \mathbf{R}^n$, the full Gauss–Newton step is defined by the linear system

$$(\Phi'(x^k))^T \Phi(x^k) + (\Phi'(x^k))^T \Phi'(x^k)(x - x^k) = 0.$$

Note that this system characterizes the stationary points of the linear least-squares problem:

$$\begin{aligned}
& \text{minimize} && \frac{1}{2} \|\Phi(x^k) + \Phi'(x^k)(x - x^k)\|^2 \\
& \text{subject to} && x \in \mathbf{R}^n.
\end{aligned} \tag{2.135}$$

Supplying problem (2.135) with the trust-region constraint, we obtain the subproblem of the Levenberg–Marquardt method:

$$\begin{aligned}
& \text{minimize} && \frac{1}{2} \|\Phi(x^k) + \Phi'(x^k)(x - x^k)\|^2 \\
& \text{subject to} && \|x - x^k\| \leq \delta_k.
\end{aligned} \tag{2.136}$$

The name fully agrees with terminology introduced in Sect. 2.2.3. Indeed, subproblem (2.136) is convex, and according to Proposition 2.33 (applied

with $g = (\Phi'(x^k))^T \Phi(x^k)$ and positive semidefinite $H = (\Phi'(x^k))^T \Phi'(x^k)$, its solution \tilde{x}^{k+1} is characterized by the feasibility condition $\|\tilde{x}^{k+1} - x^k\| \leq \delta_k$ and by the existence of $\nu_k \geq 0$ such that $\nu_k = 0$ provided $\|\tilde{x}^{k+1} - x^k\| < \delta_k$, and

$$(\Phi'(x^k))^T \Phi(x^k) + ((\Phi'(x^k))^T \Phi'(x^k) + \nu_k I)(\tilde{x}^{k+1} - x^k) = 0.$$

The step defined this way corresponds precisely to the direction defined according to (2.104). Observe that if the point \tilde{x}^{k+1} obtained by the full Gauss–Newton step satisfies $\|\tilde{x}^{k+1} - x^k\| < \delta_k$, then this \tilde{x}^{k+1} also solves the trust-region subproblem (2.136).

Convergence analysis of trust-region methods with inexact Newtonian models can be found in [208, Chap. 4].

To conclude, we note that instead of the trust-region subproblem (2.108), where a quadratic function is minimized in a ball, one could consider the subproblem

$$\begin{aligned} & \text{minimize} && \psi_k(x) \\ & \text{subject to} && \|x - x^k\|_\infty \leq \delta_k, \end{aligned}$$

where the minimization is carried out in a box. In other words, instead of the Euclidean norm $\|\cdot\| = \|\cdot\|_2$, the trust-region is defined using the norm $\|\cdot\|_\infty$ (from computational point of view, other norms can hardly be useful). With this change, essentially the same results can be obtained. As a practical matter, minimization of quadratic functions subject to bound constraints also can be performed efficiently, using specialized software.

Both linesearch and trust-region techniques play important roles for globalizing Newton-type methods in the more general settings of constrained optimization and variational problems, introducing merit functions to measure progress of a given algorithm.

2.4 Semismooth Newton Method

In this section we extend Newton-type methods to nonsmooth equations and to unconstrained optimization problems with nonsmooth derivatives. Applications of the resulting methods to variational problems will be the subject of Sect. 3.2.

2.4.1 Semismooth Newton Method for Equations

Consider the equation

$$\Phi(x) = 0, \tag{2.137}$$

where $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is no longer assumed to be smooth. Various examples of nonsmooth equations arising from variational and optimization problems are provided in Sect. 3.2 below; see [213] for other examples.

One seemingly appealing idea in this setting is to apply quasi-Newton methods, since the derivatives of Ψ are not involved in their iterations; see Sect. 2.1.1. However, it is known that quasi-Newton methods with standard updates are guaranteed to preserve superlinear convergence only when the equation mapping is actually differentiable at the solution (see [125, 195]), which is an assumption too restrictive.

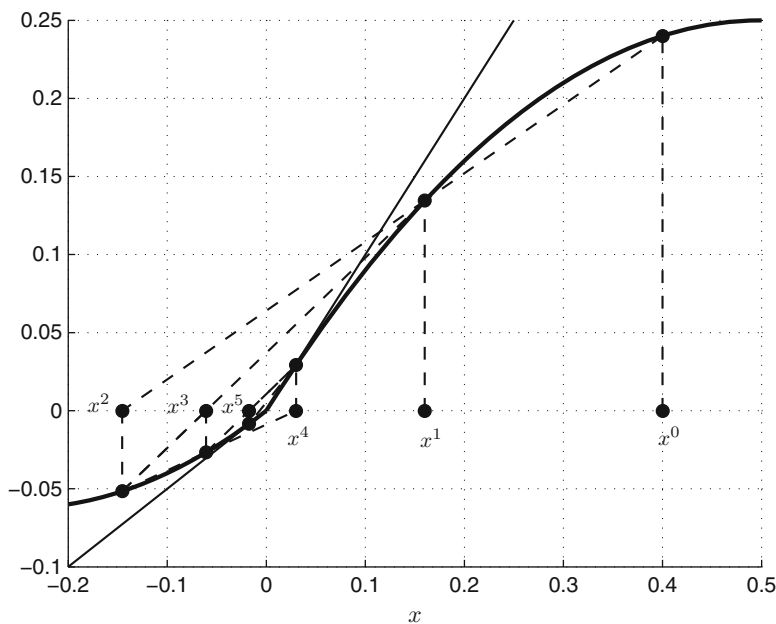


Fig. 2.7 Iterates in Example 2.40

Example 2.40. Let $n = 1$,

$$\Phi(x) = \begin{cases} x/2 + x^2 & \text{if } x \leq 0, \\ x - x^2 & \text{if } x > 0. \end{cases}$$

The solution of interest is $\bar{x} = 0$, and Φ is locally Lipschitz-continuous at 0 and differentiable everywhere except for 0. Evidently, $\partial_B \Phi(0) = \{1/2, 1\}$, $\partial \Phi(0) = [1/2, 1]$, and in particular, Φ is CD -regular at 0.

For the equation (2.137) with this Φ , the secant method (which is the one-dimensional instance of Broyden's method; see Sect. 2.1.1) works as follows: being started from $x^0 > 0$ small enough, it generates $x^1 > 0$, then $x^2 < 0$ and $x^3 < 0$, then $x^4 > 0$, then $x^5 < 0$ and $x^6 < 0$, etc. (see Fig. 2.7). After

computing two subsequent iterates corresponding to the same smooth piece, namely, $x^{k-2} < 0$ and $x^{k-1} < 0$ for some $k \geq 4$, the method generates good secant approximation of the function, and the next iterate $x^k > 0$ is “superlinearly closer” to the solution than x^{k-1} : $x^k = o(|x^{k-1}|)$. Then, according to (2.35), we derive the estimate

$$\begin{aligned} x^{k+1} &= x^k - \frac{\Phi(x^k)(x^k - x^{k-1})}{\Phi(x^k) - \Phi(x^{k-1})} \\ &= x^k - \frac{(x^k - (x^k)^2)(x^k - x^{k-1})}{x^k - (x^k)^2 - x^{k-1}/2 - (x^{k-1})^2} \\ &= x^k - \frac{-x^k x^{k-1} + o(x^k x^{k-1})}{-x^{k-1}/2 + o(x^{k-1})} \\ &= x^k - 2x^k + o(x^k) \\ &= -x^k + o(x^k), \end{aligned}$$

and superlinear decrease of the distance to the solution is lost.

The idea of the *semismooth Newton method*, originating from [175, 176] and independently developed in [222, 225], is to replace the Jacobian in the Newtonian iteration system (2.2) by an element of Clarke’s generalized Jacobian, which may be nonempty even if the true Jacobian does not exist (see Sect. 1.4). Specifically, for the current iterate $x^k \in \mathbf{R}^n$, the next iterate is a solution of the linear equation

$$\Phi(x^k) + J_k(x - x^k) = 0 \quad (2.138)$$

with some $J_k \in \partial\Phi(x^k)$. This is a direct generalization of the Newton method for the differentiable case.

Algorithm 2.41 Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Phi(x^k) = 0$, stop.
2. Compute some $J_k \in \partial\Phi(x^k)$. Compute $x^{k+1} \in \mathbf{R}^n$ as a solution of (2.138).
3. Increase k by 1 and go to step 1.

Assuming that J_k is nonsingular, the semismooth Newton method can be written in the form of the explicit iterative scheme

$$x^{k+1} = x^k - J_k^{-1}\Phi(x^k), \quad J_k \in \partial\Phi(x^k), \quad k = 0, 1, \dots \quad (2.139)$$

Recall that the nondegeneracy assumption needed for local superlinear convergence of the Newton method to the solution $\bar{x} \in \mathbf{R}^n$ the equation (2.137) in the smooth case, consists of nonsingularity of the Jacobian $\Phi'(\bar{x})$ (see Theorem 2.2). In the context of semismooth Newton methods, this condition is replaced by *CD*-regularity of Φ at \bar{x} , as defined in Sect. 1.4.1.

Theorem 2.42. *Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be locally Lipschitz-continuous at $\bar{x} \in \mathbf{R}^n$, and assume that the estimate*

$$\sup_{J \in \partial\Phi(x+\xi)} \|\Phi(x+\xi) - \Phi(x) - J\xi\| = o(\|\xi\|) \quad (2.140)$$

holds as $\xi \in \mathbf{R}^n$ tends to 0. Let \bar{x} be a solution of the equation (2.137), and assume that Φ is CD-regular at \bar{x} .

Then the following assertions are valid:

- (a) *There exists a neighborhood U of \bar{x} and a function $q(\cdot) : U \rightarrow \mathbf{R}$ such that every $J \in \partial\Phi(x)$ is nonsingular for all $x \in U$,*

$$\|x - J^{-1}\Phi(x) - \bar{x}\| \leq q(x)\|x - \bar{x}\| \quad \forall x \in U, \quad (2.141)$$

and

$$q(x) \rightarrow 0 \text{ as } x \rightarrow \bar{x}. \quad (2.142)$$

- (b) *Any starting point $x^0 \in \mathbf{R}^n$ close enough to \bar{x} defines a particular iterative sequence of Algorithm 2.41, any such sequence converges to \bar{x} , and the rate of convergence is superlinear.*
- (c) *If the estimate*

$$\sup_{J \in \partial\Phi(x+\xi)} \|\Phi(x+\xi) - \Phi(x) - J\xi\| = O(\|\xi\|^2) \quad (2.143)$$

holds as $\xi \in \mathbf{R}^n$ tends to 0, then

$$q(x) = O(\|x - \bar{x}\|) \quad (2.144)$$

as $x \rightarrow \bar{x}$, and the rate of convergence is quadratic.

Proof. From Proposition 1.51 and Lemma A.6, it easily follows that there exist a neighborhood U of \bar{x} and $M > 0$ such that

$$J \text{ is nonsingular, } \|J^{-1}\| \leq M \quad \forall J \in \partial\Phi(x), \forall x \in U. \quad (2.145)$$

Combining this with (2.140), we conclude that the inclusions $x \in U$ and $J \in \partial\Phi(x)$ imply

$$\begin{aligned} \|x - J^{-1}\Phi(x) - \bar{x}\| &\leq \|J^{-1}\| \|\Phi(x) - \Phi(\bar{x}) - J(x - \bar{x})\| \\ &\leq q(x)\|x - \bar{x}\| \end{aligned} \quad (2.146)$$

with some $q(x)$ satisfying (2.142), and (2.146) gives (2.141). This completes the proof of assertion (a).

Assertion (b) follows from (a) by the same argument as in the proof of Theorem 2.2.

Finally, if (2.143) holds, then from (2.145) and the first inequality in (2.146) it follows that the inclusions $x \in U$ and $J \in \partial\Phi(x)$ imply the second inequality

with some $q(x)$ satisfying (2.144). The quadratic convergence rate now follows from (2.144) and (2.146). This proves (c). \square

Following the lines of previous development of Newton-type methods in Sect. 2.1.1, some useful variations of the semismooth Newton method can be modeled via its perturbed versions. For a given $x^k \in \mathbf{R}^n$, the next iterate $x^{k+1} \in \mathbf{R}^n$ of a perturbed semismooth Newton method must satisfy the system

$$\Phi(x^k) + J_k(x - x^k) + \omega^k = 0 \quad (2.147)$$

with some $J_k \in \partial\Phi(x^k)$ and some perturbation term $\omega^k \in \mathbf{R}^n$.

The next technical lemma follows from Theorem 2.42 the same way as Lemma 2.3 follows from Theorem 2.2.

Lemma 2.43. *Under the assumptions of Theorem 2.42, there exist a neighborhood U of \bar{x} and $M > 0$ such that for any $x^k \in U$, any $J_k \in \partial\Phi(x^k)$, any $\omega^k \in \mathbf{R}^n$, and any $x^{k+1} \in \mathbf{R}^n$ satisfying the equation (2.147), it holds that*

$$\|x^{k+1} - \bar{x}\| \leq M\omega^k + o(\|x^k - \bar{x}\|) \quad (2.148)$$

as $x^k \rightarrow \bar{x}$. Moreover, if the estimate (2.143) holds, then the estimate (2.148) can be sharpened as follows:

$$\|x^{k+1} - \bar{x}\| \leq M\omega^k + O(\|x^k - \bar{x}\|^2) \quad (2.149)$$

as $x^k \rightarrow \bar{x}$.

The proof of the next a posteriori result mimics the proof of Proposition 2.4, but with (2.140) and (2.143) used instead of smoothness assumptions, and with Lemma 2.43 used instead of Lemma 2.3.

Proposition 2.44. *Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be locally Lipschitz-continuous at a point $\bar{x} \in \mathbf{R}^n$, and assume that the estimate (2.140) holds for $\xi \in \mathbf{R}^n$. Let \bar{x} be a solution of the equation (2.137). Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to \bar{x} and such that x^{k+1} satisfies (2.147) with some $J_k \in \partial\Phi(x^k)$ and $\omega^k \in \mathbf{R}^n$ for all k large enough.*

If the rate of convergence of $\{x^k\}$ is superlinear, then

$$\omega^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \quad (2.150)$$

as $k \rightarrow \infty$.

Conversely, if Φ is CD-regular at \bar{x} , and (2.150) holds, then the rate of convergence of $\{x^k\}$ is superlinear. Moreover, the rate of convergence is quadratic, provided (2.143) holds and

$$\omega^k = O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2)$$

as $k \rightarrow \infty$.

Remark 2.45. By Remark 2.5, we note that in condition (2.150) the right-hand side can be replaced by either $o(\|x^{k+1} - x^k\|)$ or $o(\|x^k - \bar{x}\|)$. The condition modified this way is generally stronger than (2.150). However, if $\{x^k\}$ is assumed to be superlinearly convergent to \bar{x} , these conditions become equivalent.

Remark 2.46. The assumptions of the sufficiency part of Proposition 2.44 can be formally relaxed as follows: instead of assuming that (2.150) holds with $\omega^k = -\Phi(x^k) - J_k(x^{k+1} - x^k)$ for the given $J_k \in \partial\Phi(x^k)$, one can assume that (2.150) holds for some $J_k \in \partial\Phi(x^k)$, which means that

$$\min_{J \in \partial\Phi(x^k)} \|\Phi(x^k) + J(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

as $k \rightarrow \infty$. However, due to the necessity part of Proposition 2.44, taking into account Proposition 1.51, this would actually imply that

$$\max_{J \in \partial\Phi(x^k)} \|\Phi(x^k) + J(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

as $k \rightarrow \infty$. Therefore, this modification does not make Proposition 2.44 any sharper.

Remark 2.47. Local Lipschitz-continuity of Φ at \bar{x} and the estimate (2.140) are the ingredients of semismoothness of Φ at \bar{x} , while estimate (2.143) is the ingredient of strong semismoothness (see Sect. 1.4.2). Therefore, by Theorem 2.42, if Φ is CD -regular at a solution \bar{x} of the equation (2.137), Algorithm 2.41 possesses local superlinear convergence to this solution provided Φ is semismooth at \bar{x} . (That is why we use the name “semismooth Newton method.”) Moreover, the rate of convergence is quadratic provided Φ is strongly semismooth at \bar{x} . The development in [222, 225] relies on semismoothness assumption, while [175, 176] directly employ (2.140), which gives a sharper result. Note also that the condition (2.140) in Theorem 2.42 cannot be dropped; see [68, Example 7.4.1].

Conditions (2.140) and (2.143) can be replaced by semismoothness and strong semismoothness, respectively, in Lemma 2.43 and Proposition 2.44 as well. In the rest of this section, we employ the assumption of semismoothness: this makes the exposition somewhat simpler, even though not all the ingredients of semismoothness are always needed.

As an immediate application of Proposition 2.44, consider the following *semismooth quasi-Newton method*. Let $\{J_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of matrices. For the current iterate $x^k \in \mathbf{R}^n$, let the next iterate x^{k+1} be computed as a solution of the equation (2.138), and assume that $\{J_k\}$ satisfies the Dennis–Moré-type condition (cf. (2.32) and (2.80), (2.81)):

$$\min_{J \in \partial\Phi(x^k)} \|(J_k - J)(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|) \quad (2.151)$$

as $k \rightarrow \infty$. The latter implies the existence of $\tilde{J}_k \in \partial\Phi(x^k)$ such that

$$(J_k - \tilde{J}_k)(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|). \quad (2.152)$$

Therefore, the equation (2.138) can be interpreted as (2.147) with J_k replaced by \tilde{J}_k , and with

$$\omega^k = (J_k - \tilde{J}_k)(x^{k+1} - x^k).$$

By (2.152),

$$\omega^k = o(\|x^{k+1} - x^k\|)$$

as $k \rightarrow \infty$, and hence, Proposition 2.44 implies the following a posteriori result.

Theorem 2.48. *Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be semismooth at a point $\bar{x} \in \mathbf{R}^n$. Let \bar{x} be a solution of the equation (2.137), and let Φ be CD-regular at \bar{x} . Let $\{J_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of matrices, and let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to \bar{x} and such that x^{k+1} satisfies (2.138) for all k large enough. Assume, finally, that condition (2.151) holds.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Remark 2.49. Evidently, condition (2.151) in Theorem 2.48 can be replaced by the following formally stronger condition:

$$\max_{J \in \partial\Phi(x^k)} \|(J_k - J)(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|)$$

as $k \rightarrow \infty$. Moreover, according to Remark 2.46, in the context of Theorem 2.48 this condition is actually not stronger than (2.151).

It would seem natural to state next a counterpart of Proposition 2.6 for the perturbed semismooth Newton method. However, such a statement would be of less interest in this context, because the special selection of the norm $\|\cdot\|_*$ in such a counterpart of Proposition 2.6 would have now to take care of *all* $J \in \partial\Phi(\bar{x})$ simultaneously. Because of that, the consequences would not be as nice as in the smooth case. In particular, in the counterpart of Theorem 2.8 (see Theorem 2.51 below), it is not enough to assume that forcing parameters are bounded above by some (arbitrary fixed) constant $\theta \in (0, 1)$: we can only establish the *existence* of the needed θ .

Therefore, we proceed directly to *truncated semismooth Newton methods*. These are practically important instances of perturbed semismooth Newton methods, with the perturbation terms in the iteration system (2.147) satisfying the condition

$$\|\omega^k\| \leq \theta_k \|\Phi(x^k)\|, \quad k = 0, 1, \dots \quad (2.153)$$

for some forcing sequence $\{\theta_k\}$. Note that condition (2.153) is exactly the same as truncation condition (2.26) employed above for smooth equations.

Theorem 2.50. *Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be semismooth at $\bar{x} \in \mathbf{R}^n$ and let \bar{x} be a solution of the equation (2.137). Let $\{x^k\} \subset \mathbf{R}^n$ be a sequence convergent to \bar{x} , and for each $k = 0, 1, \dots$, let*

$$\omega^k = -\Phi(x^k) - J_k(x^{k+1} - x^k) \quad (2.154)$$

with some $J_k \in \partial\Phi(x^k)$.

If the rate of convergence of $\{x^k\}$ is superlinear, then there exists a sequence $\{\theta_k\} \subset \mathbf{R}$ satisfying condition (2.153), and such that $\theta_k \rightarrow 0$.

Conversely, if Φ is CD-regular at \bar{x} , and there exists a sequence $\{\theta_k\} \subset \mathbf{R}$ satisfying condition (2.153), and such that $\theta_k \rightarrow 0$, then the rate of convergence of $\{x^k\}$ is superlinear. The rate of convergence is quadratic, provided Φ is strongly semismooth at \bar{x} and

$$\theta_k = O(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \quad (2.155)$$

as $k \rightarrow \infty$.

Proof. To prove the first assertion, observe that by Proposition 1.64 and Remark 1.65, for all k large enough it holds that

$$\|x^k - \bar{x}\| = O(\|\Phi(x^k)\|) \quad (2.156)$$

as $k \rightarrow \infty$ (recall that semismoothness of Φ at \bar{x} implies directional differentiability of Φ at \bar{x} in every direction, so that Proposition 1.64 is applicable). By Proposition 2.44 and Remark 2.45, the superlinear convergence rate of $\{x^k\}$ implies the estimate

$$\omega^k = o(\|x^k - \bar{x}\|) \quad (2.157)$$

as $k \rightarrow \infty$. Combining the latter with (2.156), we have that

$$\omega^k = o(\|\Phi(x^k)\|)$$

as $k \rightarrow \infty$, which means precisely the existence of a sequence $\{\theta_k\}$ with the claimed properties.

The second assertion follows from Proposition 2.44, since the relation (2.153) with $\theta_k \rightarrow 0$ evidently implies (2.157) (and, hence, (2.150)):

$$\omega^k = o(\|\Phi(x^k) - \Phi(\bar{x})\|) = o(\|x^k - \bar{x}\|)$$

as $k \rightarrow \infty$, where local Lipschitz-continuity of Φ at \bar{x} was employed.

Finally, if Φ is strongly semismooth at \bar{x} , and (2.155) holds, quadratic convergence follows by the last assertion of Proposition 2.44. \square

Theorem 2.51. *Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be semismooth at $\bar{x} \in \mathbf{R}^n$, let \bar{x} be a solution of the equation (2.137), and assume that Φ is CD-regular at \bar{x} .*

Then there exists $\theta > 0$ such that for any starting point $x^0 \in \mathbf{R}^n$ close enough to \bar{x} and any sequences $\{x^k\} \subset \mathbf{R}^n$, $\{\omega^k\} \subset \mathbf{R}^n$ and $\{\theta_k\} \subset [0, \theta]$ satisfying (2.153) and (2.154) with some $J_k \in \partial\Phi(x^k)$ for all $k = 0, 1, \dots$, it holds that $\{x^k\}$ converges to \bar{x} and the rate of convergence is (at least) linear. Moreover, the rate of convergence is superlinear if $\theta_k \rightarrow 0$. The rate of convergence is quadratic, provided Φ is strongly semismooth at \bar{x} and provided $\theta_k = O(\|\Phi(x^k)\|)$ as $k \rightarrow \infty$.

Proof. Let $L > 0$ stand for the Lipschitz constant of Φ with respect to \bar{x} . Similarly to the proof of Theorem 2.42, we can choose a neighborhood U of \bar{x} and $M > 0$ such that (2.145) holds. Take any $\theta < 1/(LM)$, and any $q \in (\theta LM, 1)$. By (2.153), (2.154) and the inclusion $\theta_k \in [0, \theta]$, employing assertion (a) of Theorem 2.42, we obtain that there exists $\delta > 0$ such that $B(\bar{x}, \delta) \subset U$, and for any $x^k \in B(\bar{x}, \delta)$ it holds that

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &= \|x^k - J_k^{-1}(\Phi(x^k) + \omega^k) - \bar{x}\| \\ &\leq \|J_k^{-1}\omega^k\| + \|x^k - J_k^{-1}\Phi(x^k) - \bar{x}\| \\ &\leq \theta_k \|J_k^{-1}\| \|\Phi(x^k)\| + o(\|x^k - \bar{x}\|) \\ &\leq \theta M \|\Phi(x^k) - \Phi(\bar{x})\| + o(\|x^k - \bar{x}\|) \\ &\leq \theta LM \|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|) \\ &\leq q \|x^k - \bar{x}\|. \end{aligned} \tag{2.158}$$

Since $q \in (0, 1)$, this estimate implies the inclusion $x^{k+1} \in B(\bar{x}, \delta)$. Therefore, the inclusion $x^0 \in B(\bar{x}, \delta)$ implies that the entire sequence $\{x^k\}$ is contained in $B(\bar{x}, \delta)$, and (2.158) shows convergence of this sequence to \bar{x} at a linear rate.

Superlinear rate of convergence when $\theta_k \rightarrow 0$, and quadratic rate when Φ is strongly semismooth at \bar{x} and $\theta_k = O(\|\Phi(x^k)\|)$ as $k \rightarrow \infty$, follow from Theorem 2.50. \square

The fact that in Theorem 2.51 one cannot simply take an arbitrary $\theta \in (0, 1)$ is confirmed by the following counterexample, borrowed from [68, Example 7.5.6].

Example 2.52. Let $n = 2$,

$$\Phi_1(x) = \begin{cases} 2x_1 & \text{if } x_1 \geq 0, \\ x_1 & \text{if } x_1 < 0, \end{cases} \quad \Phi_2(x) = \begin{cases} x_2 & \text{if } x_2 \geq 0, \\ 2x_2 & \text{if } x_2 < 0. \end{cases}$$

The unique solution of $\Phi(x) = 0$ is $\bar{x} = 0$, and Φ is locally Lipschitz-continuous at 0 and differentiable everywhere except for those points satisfying $x_1 = 0$ or $x_2 = 0$. Evidently, $\partial_B \Phi(0)$ consists of the four matrices

$$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

In particular, Φ is CD -regular at 0.

Fix any $\delta > 0$ and define the sequence $\{x^k\}$ as follows:

$$x^k = \begin{cases} (\delta, \delta/2) & \text{if } k \text{ is even,} \\ (-\delta/2, -\delta) & \text{if } k \text{ is odd.} \end{cases}$$

Then Φ is differentiable at each x^k , and it can be directly checked that for all k

$$\|\Phi(x^k) + \Phi'(x^k)(x^{k+1} - x^k)\| = \sqrt{2}\delta,$$

while at the same time

$$\|\Phi(x^k)\| = \frac{\sqrt{17}\delta}{2}.$$

Therefore, if we take any $\theta \geq 2\sqrt{2}/\sqrt{17}$, then

$$\|\Phi(x^k) + \Phi'(x^k)(x^{k+1} - x^k)\| \leq \theta \|\Phi(x^k)\|.$$

Hence, the sequence $\{x^k\}$ can be generated by the truncated semismooth Newton method with $\theta_k = \theta$ for all k . However, $\{x^k\}$ oscillates between the two points distinct from 0, no matter how small δ is (and hence, no matter how close x^0 is to \bar{x}).

Various important modifications of semismooth Newton methods arise from the natural intention to develop globally convergent schemes. Far from a CD -regular solution, the generalized Jacobian may contain singular matrices. As a remedy, consider the following *semismooth Levenberg–Marquardt method*: for a given $x^k \in \mathbf{R}^n$, the next iterate $x^{k+1} \in \mathbf{R}^n$ is defined by the equation

$$J_k^T \Phi(x^k) + (J_k^T J_k + \nu_k I)(x - x^k) = 0 \quad (2.159)$$

with some $J_k \in \partial\Phi(x^k)$ and some $\nu_k \geq 0$.

For $\nu_k = 0$, the semismooth Levenberg–Marquardt method reduces to the *semismooth Gauss–Newton method* with the iteration system

$$J_k^T \Phi(x^k) + J_k^T J_k(x - x^k) = 0. \quad (2.160)$$

One potential advantage of the latter compared to the basic iteration system (2.138) is the symmetry of the iteration matrix in (2.160), the property that can be efficiently employed in the methods used to solve this system. In particular, solving (2.160) is equivalent to solving the linear least-squares problem

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|\Phi(x^k) + J_k(x - x^k)\|^2 \\ & \text{subject to} && x \in \mathbf{R}^n. \end{aligned}$$

See the related discussion for the Gauss–Newton method for smooth equations in Sect. 2.1.1.

Similarly, solving (2.159) is equivalent to solving the unconstrained optimization problem

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|J_k(x - x^k) + \Phi(x^k)\|^2 + \frac{\nu_k}{2} \|x - x^k\|^2 \\ & \text{subject to} && x \in \mathbf{R}^n, \end{aligned}$$

whose quadratic objective function is strongly convex when $\nu_k > 0$. As in the smooth case, smart selection of ν_k is essential for efficiency of the semismooth Levenberg–Marquardt method. One possibility arises in the context of trust-region methods; see Sect. 2.3. Speaking about local convergence theory, the following result can be derived by combining considerations similar to those in the proof of Theorem 2.11, with the proof of Theorem 2.51. We omit the details.

Theorem 2.53. *Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be semismooth at $\bar{x} \in \mathbf{R}^n$, let \bar{x} be a solution of the equation (2.137), and assume that Φ is CD -regular at \bar{x} .*

Then there exists a constant $\nu > 0$ with the property that for any sequence $\{\nu_k\} \subset [0, \nu]$, any starting point $x^0 \in \mathbf{R}^n$ close enough to \bar{x} defines an iterative sequence $\{x^k\} \subset \mathbf{R}^n$ such that for each $k = 0, 1, \dots$, the point x^{k+1} satisfies (2.159) with some $J_k \in \partial\Phi(x^k)$; any such sequence converges to \bar{x} , and the rate of convergence is (at least) linear. Moreover, the rate of convergence is superlinear if $\nu_k \rightarrow 0$. The rate of convergence is quadratic, provided Φ is strongly semismooth at \bar{x} and $\nu_k = O(\|\Phi(x^k)\|)$ as $k \rightarrow \infty$.

The very natural next step (which we again only outline without going into details) is the development of a truncated semismooth Levenberg–Marquardt method, where the iteration system (2.159) is replaced by

$$J_k^T \Phi(x^k) + (J_k^T J_k + \nu_k I)(x - x^k) + \omega^k = 0$$

with the perturbation terms $\omega^k \in \mathbf{R}^n$ satisfying the truncation condition (2.153) for some forcing sequence $\{\theta_k\}$. Combining the ideas contained in Theorems 2.51 and 2.53, local superlinear convergence to a solution satisfying CD -regularity can be established if both ν_k and θ_k are kept small enough.

We note that there are various other perturbed/modified versions of semismooth Newton methods in the literature; one interesting example is given in [138].

Remark 2.54. The CD -regularity assumption in Theorem 2.42 can be somewhat weakened if we assume that J_k is chosen in step 2 of Algorithm 2.41 not from $\partial\Phi(x^k)$ but from the generally smaller set $\partial_B\Phi(x^k)$. Specifically, CD -regularity can be replaced by BD -regularity of Φ at \bar{x} . The proof of the corresponding counterpart of Theorem 2.42 is exactly the same, but with the reference to the part of Proposition 1.51 concerning B -differentials.

Moreover, sometimes even more special choices of generalized Jacobians can be useful, e.g., employing the specific structure of Φ . For a generic scheme of this kind, suppose that $J_k \in \Delta(x^k)$ for each k , where Δ is a given multifunction from \mathbf{R}^n to the subsets of $\mathbf{R}^{n \times n}$ such that $\Delta(x) \subset \partial\Phi(x)$ for all $x \in \mathbf{R}^n$. Then CD -regularity can be replaced by the assumption of nonsingularity of any matrix in the set

$$\bar{\Delta}(\bar{x}) = \left\{ J \in \mathbf{R}^{n \times n} \mid \begin{array}{l} \exists \{x^k\} \subset \mathbf{R}^n, \{J_k\} \subset \mathbf{R}^{n \times n} : J_k \in \Delta(x^k) \forall k, \\ \{x^k\} \rightarrow \bar{x}, \{J_k\} \rightarrow J \end{array} \right\}.$$

By Proposition 1.51, it holds that $\bar{\Delta}(\bar{x}) \subset \partial\Phi(\bar{x})$. Hence, this regularity assumption is implied by CD -regularity, and it can be strictly weaker than CD -regularity, depending on the choice of $\Delta(\cdot)$.

Similar modifications apply not only to Theorem 2.42, but to all results of this section.

Remark 2.55. An approach originally suggested in [172] is closely related to the semismooth Newton method, but its domain of application is much narrower. Let $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be *piecewise smooth* on an open set $O \subset \mathbf{R}^n$, which means that Φ is continuous and there exists a finite collection of continuously differentiable mappings $\Phi_i : O \rightarrow \mathbf{R}^n$, $i = 1, \dots, m$, called smooth pieces of Φ , such that for any $x \in O$ the set

$$I(x) = \{i = 1, \dots, m \mid \Phi(x) = \Phi_i(x)\}$$

of indices of smooth pieces active at x is nonempty. Then the set

$$\hat{\partial}\Phi(x) = \{\Phi'_i(x) \mid i \in I(x)\}$$

is nonempty as well, and one can define the *piecewise Newton method* for solving the equation (2.137) by the iterative system (2.138) with $J_k \in \hat{\partial}\Phi(x^k)$ (provided $x^k \in O$).

Assuming now that $\bar{x} \in \mathbf{R}^n$ is a solution of the equation (2.137), that Φ is piecewise smooth on some neighborhood O of \bar{x} , and that all the matrices in $\hat{\partial}\Phi(\bar{x})$ are nonsingular, the local superlinear convergence of this method to \bar{x} follows easily from the fact that $I(x) \subset I(\bar{x})$ for all $x \in O$ close enough to \bar{x} . Indeed, this fact implies that, locally, each iteration of the piecewise Newton method can be interpreted as the usual Newtonian iteration for the smooth equation

$$\Phi_i(x) = 0$$

for some $i \in I(\bar{x})$, and \bar{x} is a solution of this equation as well. The convergence result can now be derived from Theorem 2.2.

It is evident that

$$\partial_B\Phi(x) \subset \hat{\partial}\Phi(x) \quad \forall x \in O.$$

Employing this relation, it can be easily seen that piecewise smoothness of Φ on O implies semismoothness of Φ at any point of O , and for equations

with piecewise smooth mappings, the semismooth Newton method employing B -differentials (see Remark 2.54) can be regarded as a special realization of the piecewise smooth Newton method. Computing an element of $\hat{\partial}\Phi(x^k)$ can be much simpler than computing an element of $\partial_B\Phi(x^k)$. The price paid for this is the necessity to assume nonsingularity of all the matrices in $\hat{\partial}\Phi(\bar{x})$, which can be a strictly stronger assumption than BD -regularity of Φ at \bar{x} . This issue will be illustrated later on by Example 3.15 in the context of complementarity problems.

An important particular instance of a piecewise smooth mapping is a *piecewise linear* (*piecewise affine*) mapping, which has linear (affine) pieces Φ_i , $i = 1, \dots, m$.

2.4.2 Semismooth Newton Method for Unconstrained Optimization

One possible application of semismooth Newton methods is an unconstrained optimization problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in \mathbf{R}^n, \end{aligned} \tag{2.161}$$

with a differentiable objective function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ whose gradient is Lipschitz-continuous, but the second derivatives of f may not exist. Similarly to Sect. 2.1.2, stationary points of this problem are characterized by the equation (2.137) with $\Phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ being the gradient mapping of f :

$$\Phi(x) = f'(x).$$

This Φ is Lipschitz-continuous. Thus, one can try to search for stationary points of the optimization problem (2.161) by applying the semismooth Newton-type method to the equation (2.137).

Given the current iterate $x^k \in \mathbf{R}^n$, the next iterate x^{k+1} of the basic semismooth Newton method for (2.161) is computed as a solution of the linear system

$$f'(x^k) + H_k(x - x^k) = 0 \tag{2.162}$$

with H_k being an element of the generalized Hessian of f at x^k , that is, $H_k \in \partial f'(x^k)$. Assuming that H_k is nonsingular, the semismooth Newton method can be written in the form of the explicit iterative scheme

$$x^{k+1} = x^k - H_k^{-1}f'(x^k), \quad H_k \in \partial f'(x^k), \quad k = 0, 1, \dots$$

As usual, the optimization nature of the original problem can be put to the foreground by observing that solutions of (2.162) coincide with stationary points of the subproblem

$$\begin{aligned} & \text{minimize} && f(x^k) + \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle H_k(x - x^k), x - x^k \rangle \\ & \text{subject to} && x \in \mathbf{R}^n. \end{aligned} \quad (2.163)$$

Therefore, the basic *semismooth Newton method* for unconstrained optimization is stated as follows.

Algorithm 2.56 Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $f'(x^k) = 0$, stop.
2. Compute some $H_k \in \partial f'(x^k)$. Compute $x^{k+1} \in \mathbf{R}^n$ as a stationary point of problem (2.163).
3. Increase k by 1 and go to step 1.

Local convergence properties of the semismooth Newton method for unconstrained optimization follow directly from Theorem 2.42.

Theorem 2.57. *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with its gradient being locally Lipschitz-continuous at \bar{x} , and assume that the estimate*

$$\sup_{H \in \partial f'(x+\xi)} \|f'(x+\xi) - f'(x) - H\xi\| = o(\|\xi\|)$$

holds as $\xi \in \mathbf{R}^n$ tends to 0. Let \bar{x} be a stationary point of problem (2.161), and assume that this point satisfies the SOSC

$$\forall H \in \partial f'(\bar{x}) \quad \langle H\xi, \xi \rangle > 0 \quad \forall \xi \in \mathbf{R}^n \setminus \{0\} \quad (2.164)$$

(thus, according to Theorem 1.81, \bar{x} is a strict local solution of problem (2.161)).

Then for any starting point $x^0 \in \mathbf{R}^n$ close enough to \bar{x} , any iterative sequence generated by Algorithm 2.56 converges to \bar{x} , and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the estimate

$$\sup_{J \in \partial \Phi(x+\xi)} \|\Phi(x+\xi) - \Phi(x) - J\xi\| = O(\|\xi\|^2)$$

holds as $\xi \in \mathbf{R}^n$ tends to 0.

From the symmetry of Hessians of a twice differentiable function, and from the definition of the generalized Hessian, it follows that the latter consists of symmetric matrices as well. This specificity can be used when solving the linear iteration system (2.55); see the related comments in Sect. 2.1.2.

Also, similarly to the smooth case, the assertion of Theorem 2.57 remains valid if the SOSC (2.164) is replaced by the weaker assumption of *CD*-regularity of the gradient mapping f' at \bar{x} . As in the smooth case, in this respect the semismooth Newton method does not distinguish local minimizers from other stationary points of the problem.

Needless to say, the perturbed versions of the semismooth Newton method discussed in Sect. 2.4.1 can be directly adapted for unconstrained optimization, with the same motivation as in Sect. 2.4.1.

Semismooth Newton methods for unconstrained optimization can be globalized along the lines of Sects. 2.2 and 2.3, employing f as the natural merit function. In particular, the global convergence analysis of linesearch methods presented in Sects. 2.2.1 and 2.2.2 is fully applicable in this context. An interesting related observation is stated in [67]. Let f be differentiable in a neighborhood of a stationary point \bar{x} of problem (2.58), with its gradient being semismooth at \bar{x} , and assume that the SOSC (2.164) holds. Let $\{x^k\}$ be an iterative sequence of the semismooth Newton method equipped with the Armijo or Wolfe linesearch rule with the initial trial stepsize value $\alpha = 1$ (plus some additional requirements regarding the parameters involved in those linesearch rules). Then if $\{x^k\}$ converges to \bar{x} , the method eventually accepts the unit stepsize. Hence, the convergence rate is superlinear.

Newton-Type Methods for Optimization and Variational Problems

Izmailov, A.F.; Solodov, M.V.

2014, XIX, 573 p. 30 illus., 1 illus. in color., Hardcover

ISBN: 978-3-319-04246-6