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.

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3_2,

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

$$\Phi(x) = 0, \tag{2.1}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ 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:

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) = 0. \tag{2.2}$$

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.

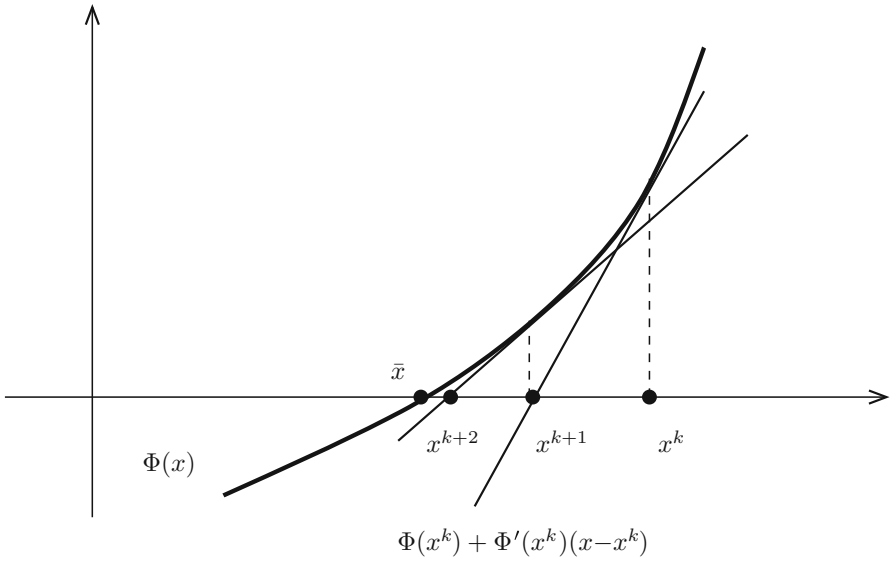

Fig. 2.1 Iterations of the Newton method

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

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

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.

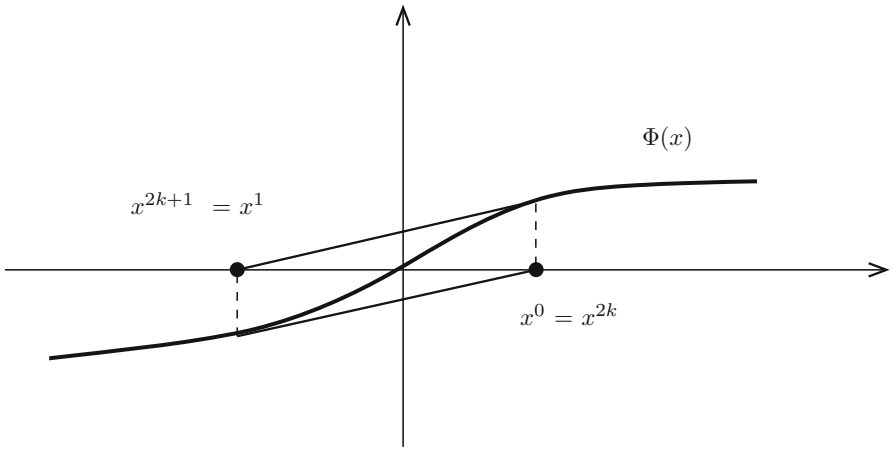

Fig. 2.2 Non-convergence of the Newton method from points far from a solution

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

Theorem 2.2. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ 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 $q(\cdot) : U \to \mathbf{R}$ such that $\Phi'(x)$ is nonsingular for all $x \in U$,

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

and

$$q(x) \to 0 \text{ as } x \to \bar{x}. \tag{2.5}$$

(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

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

as $x \to \bar{x}$, 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

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

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

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

where

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

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

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

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

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

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). □

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):

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

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*

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

it holds that

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

as $x^k \to \bar{x}$. 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:

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

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

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

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. □

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 $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ 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 $k = 0, 1, \ldots$.*

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

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

as $k \to \infty$.

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

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

as $k \to \infty$.

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

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

as $k \to \infty$. 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 $t_k \to 0$ and

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

for all k large enough. This implies that

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

Hence, for all k large enough

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

i.e.,

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

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

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

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

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

and hence, for all k large enough

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

which gives the quadratic convergence rate. □

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:

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

or

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

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 $t_k \to 0$, $\tau_k \to 0$, and

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

$$\|x^{k+1} - \bar{x}\| \le \tau_k \|x^k - \bar{x}\| \tag{2.22}$$

for all k. Then

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

implying that

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

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

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

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

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

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 $q_1, q_2 \ge 0$ 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 $k = 0, 1, \ldots$, it holds that $x^k \in B(\bar{x}, \delta)$, and ω^k defined according to (2.12) satisfies the condition

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

then

$$\|x^{k+1} - \bar{x}\|_* \le \frac{q_1 + q_2 + \varepsilon}{1 - q_1}\|x^k - \bar{x}\|_* \tag{2.24}$$

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* $k = 0, 1, \ldots$, *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*

$$\limsup_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|_*}{\|x^k - \bar{x}\|_*} \le \frac{q_1 + q_2}{1 - q_1}. \tag{2.25}$$

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

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

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). □

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

$$\|\omega^k\| \le \theta_k\|\Phi(x^k)\|, \quad k = 0, 1, \ldots. \tag{2.26}$$

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 $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ 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 $k = 0, 1, \ldots$.*

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 $\theta_k \to 0$.

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 $\theta_k \to 0$, 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

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

as $k \to \infty$.

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

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

as $k \to \infty$. 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,

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

as $k \to \infty$, and therefore, (2.26) with $\theta_k \to 0$ 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 $k \to \infty$. □

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 $k = 0, 1, \ldots$, 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 $\theta_k \to 0$. The rate of convergence is quadratic, provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$ and

$$\theta_k = O(\|\Phi(x^k)\|) \tag{2.30}$$

as $k \to \infty$.

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

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

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 $k = 0, 1, \ldots$. It remains to apply assertion (b) of Proposition 2.6. Superlinear rate of convergence when $\theta_k \to 0$, 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). □

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

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

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]):

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

as $k \to \infty$. (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 $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ 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 $\|J_k - \Phi'(x^k)\| \to 0$ as $k \to \infty$ 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

$$r^k = J_{k+1}s^k, \tag{2.33}$$

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*

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

is given by

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

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:

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

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

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

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*

$$\|\Omega_k\| \le \theta \quad \forall\, k = 0, 1, \ldots, \tag{2.37}$$

any $x^0 \in B(\bar{x}, \delta)$ uniquely defines the iterative sequence $\{x^k\} \subset B(\bar{x}, \delta)$ such that for each $k = 0, 1, \ldots$, 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 $\theta \to 0$, and either $x^k = \bar{x}$ for all k large enough, or

$$\limsup_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} \le q(\theta). \tag{2.38}$$

Moreover, the rate of convergence is superlinear if $\{\Omega_k\} \to 0$ as $k \to \infty$. 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 $k \to \infty$.

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 $k = 0, 1, \ldots$, it holds that

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

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

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

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

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

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

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

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

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 $k = 0, 1, \ldots$, 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

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

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 $\{\Omega_k\} \to 0$, 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 $k \to \infty$, then (2.17) holds, and Proposition 2.4 gives the quadratic convergence rate. □

The simplest case of a linear perturbation term is when Ω_k is just constant, i.e., $\Omega_k = \Omega$ for all $k = 0, 1, \ldots$, 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

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

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 $\Omega_k = \Omega(x^k)$, $k = 0, 1, \ldots$, where the mapping $\Omega : \mathbf{R}^n \to \mathbf{R}^{n\times n}$ is such that $\Omega(x) \to 0$ as $x \to \bar{x}$. 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 $\Psi : \mathbf{R}^n \to \mathbf{R}^m$ 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 $\Phi(x) = (\Psi'(x))^{\mathrm{T}}\Psi(x)$, $x \in \mathbf{R}^n$. 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 $\Phi'(\bar{x}) = (\Psi'(\bar{x}))^{\mathrm{T}}\Psi'(\bar{x})$ 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:

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

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

$$(\Psi'(x^k))^{\mathrm{T}}\Psi(x^k) + (\Psi'(x^k))^{\mathrm{T}}\Psi'(x^k)(x - x^k) = 0, \tag{2.41}$$

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

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

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:

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

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 $(\Psi'(x^k))^{\mathrm{T}}\Psi'(x^k)$, 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* $\omega : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ *satisfy the following assumptions:*

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

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*

$$\|\omega(x, 0)\| \le \theta\|x - \bar{x}\| \tag{2.43}$$

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 $k = 0, 1, \ldots$, 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 $\theta \to 0$.

Moreover, the rate of convergence is superlinear if

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

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

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

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

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

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

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

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

as $x \to \bar{x}$. 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,

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

as $x^k \to \bar{x}$, 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 $x^k \to \bar{x}$, 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 $\bar{x}$. □

Note that the case discussed above when $\omega^k = \omega^k(x) = \Omega(x^k)(x - x^k)$, $k = 0, 1, \ldots$, with some mapping $\Omega : \mathbf{R}^n \to \mathbf{R}^{n \times n}$ such that $\Omega(x) \to 0$ as $x \to \bar{x}$ (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)

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

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*

$$\Phi(x) + \Phi'(x)\xi + \Omega(x, \xi) \ni 0 \tag{2.49}$$

has a solution $\xi(x)$ such that $\xi(x) \to 0$ as $x \to \bar{x}$, 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

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

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 $k = 0, 1, \ldots$, satisfying

$$\|x^{k+1} - x^k\| \le \delta; \tag{2.51}$$

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 $\theta \to 0$.

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

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

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

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

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

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

where (2.50) was employed. It follows that

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

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. □

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 $\Phi : \mathbf{R} \to \mathbf{R}$ 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.,

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

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

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

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

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

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:

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

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 $f : \mathbf{R}^n \to \mathbf{R}$ 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*

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

(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 $f : \mathbf{R} \to \mathbf{R}$,

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

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 $x^k = 2(-1)^k$, $k = 1, 2, \ldots$, 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

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

with a differentiable objective function $f : \mathbf{R}^n \to \mathbf{R}$. 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 $f : \mathbf{R}^n \to \mathbf{R}$ 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 $f : \mathbf{R}^n \to \mathbf{R}$ 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

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

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

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

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

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

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 $C\theta^j$, $j = 0, 1, \ldots$, 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.

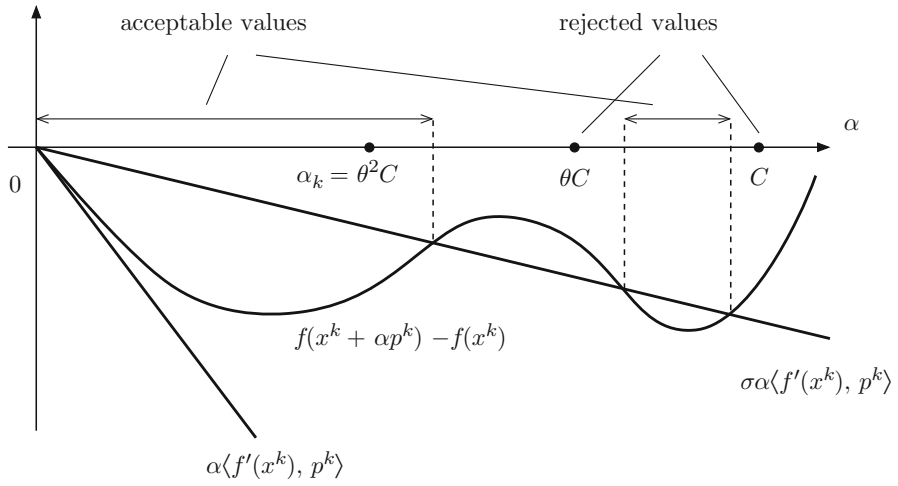

Fig. 2.3 Armijo rule

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

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

then the backtracking procedure in the Armijo rule is finite.

Lemma 2.19. *Let $f : \mathbf{R}^n \to \mathbf{R}$ 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. □

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 $f : \mathbf{R}^n \to \mathbf{R}$ 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

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

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). □

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*

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

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*

$$\alpha_k \geq c. \tag{2.66}$$

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. □

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

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

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

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

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

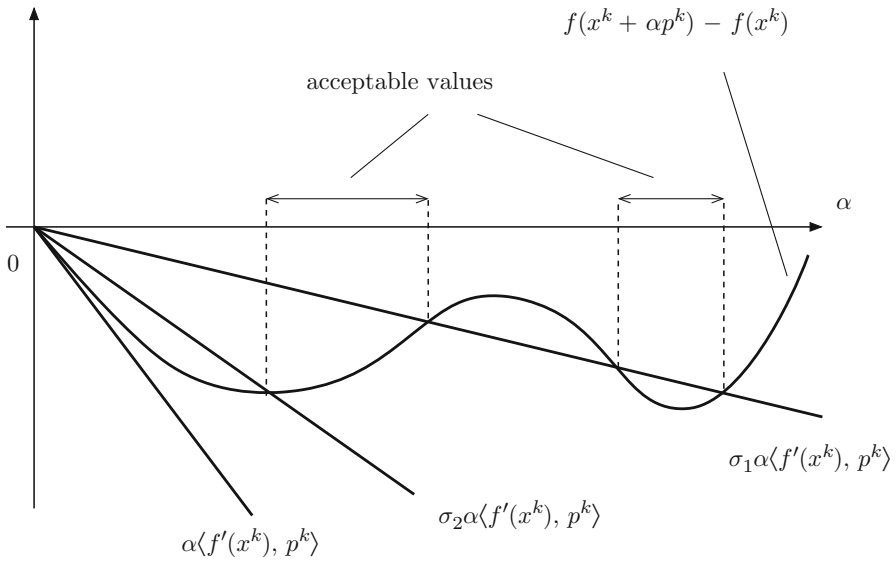

Fig. 2.4 Goldstein rule

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.

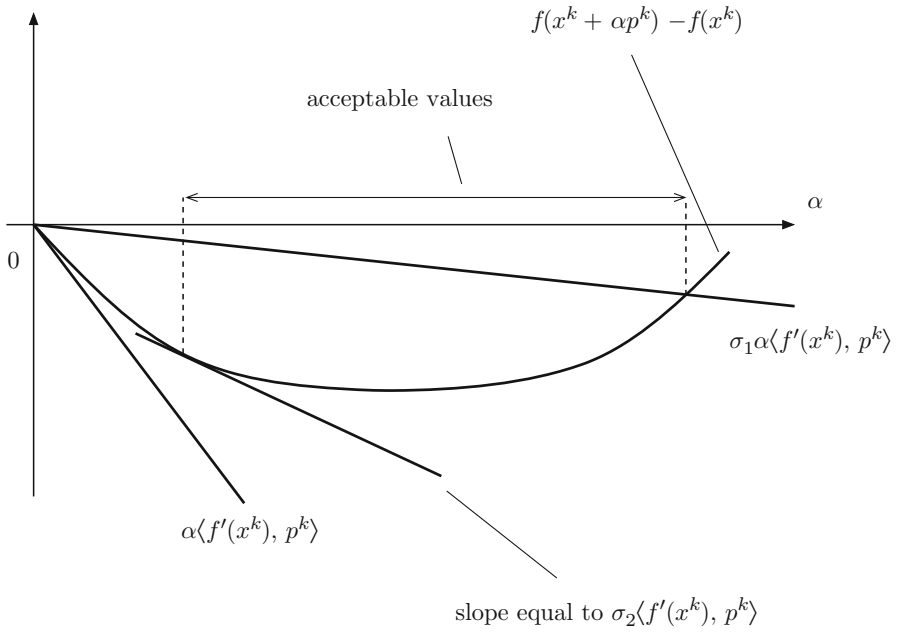

Fig. 2.5 Wolfe rule

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 $f : \mathbf{R}^n \to \mathbf{R}$ 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 $c \to +\infty$ in the case of infinite number of extrapolation steps and $(C - c) \to 0$ 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

$$f(x^k + cp^k) \le f(x^k) + \sigma_1 c \langle f'(x^k), p^k \rangle. \tag{2.70}$$

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

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

while the elements of the second sequence satisfy

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

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

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

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

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

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). □

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

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

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 $f : \mathbf{R}^n \to \mathbf{R}$ 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

$$\{f'(x^k)\} \to 0. \tag{2.76}$$

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

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

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 $k \to \infty$, we conclude that (2.76) holds. The assertion follows. □

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 $f : \mathbf{R}^n \to \mathbf{R}$ 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 $j \to \infty$. 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 $\{\alpha_{k_j}\} \to 0$ as $j \to \infty$.

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

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

Recalling the second inequality in (2.65), we conclude that $\{\tilde{\alpha}_{k_j}\} \to 0$ as $j \to \infty$. 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 $j \to \infty$, 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),

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

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

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

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 $f : \mathbf{R}^n \to \mathbf{R}$ 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 $k \to \infty$.

Conversely, if $\bar{x}$ *satisfies the SOSC*

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

and the condition

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

holds as $k \to \infty$, *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

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

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

$$Q_k f'(x^k) = O(\|f'(x^k)\|) \tag{2.82}$$

as $k \to \infty$.

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

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

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,

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

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

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

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 $f''(\bar{x})$). □

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 $f : \mathbf{R}^n \to \mathbf{R}$ 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*

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

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

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

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 $x \to \bar{x}$, 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,

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

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. □

Theorem 2.29. *Suppose that the assumptions of Theorem 2.24 are satisfied. Assume, in addition, that $f : \mathbf{R}^n \to \mathbf{R}$ 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

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

and

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

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

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

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

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

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

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

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

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

Therefore, using also (2.89), we have that

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

From the latter and (2.91), it follows that

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

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

We have thus established that $x^j \in U$ for all $j = k, k+1, \ldots$. 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. □

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 $\|Q_k - (f''(x^k))^{-1}\| \to 0$ as $k \to \infty$ 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

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

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*

$$Q_{k+1} r^k = s^k. \tag{2.93}$$

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

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

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:

$$\langle r^k,\, s^k \rangle > 0. \tag{2.95}$$

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

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

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

$$\langle s^k,\,\xi\rangle = 0, \tag{2.96}$$

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 $\xi = 0$. □

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

$$Q_{k+1} = Q_k + \frac{(r^k - Q_k s^k)(r^k)^{\mathrm{T}} + r^k(r^k - Q_k s^k)^{\mathrm{T}}}{\langle r^k,\,s^k\rangle}$$
$$- \frac{\langle r^k - Q_k s^k,\,s^k\rangle r^k(r^k)^{\mathrm{T}}}{\langle r^k,\,s^k\rangle^2}. \tag{2.97}$$

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

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

(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

Appendix A
Miscellaneous Material

A.1 Linear Algebra and Linear Inequalities

The following statement can be regarded as a variant of the celebrated Farkas Lemma (e.g., [27, Theorem 2.201]).

Lemma A.1. *For any $A \in \mathbf{R}^{l\times n}$ and $B \in \mathbf{R}^{m\times n}$, for the cone*

$$C = \{x \in \mathbf{R}^n \mid Ax = 0,\ Bx \leq 0\}$$

it holds that

$$C^\circ = \{x \in \mathbf{R}^n \mid x = A^{\mathrm{T}}y + B^{\mathrm{T}}z,\ y \in \mathbf{R}^l,\ z \in \mathbf{R}^m_+\}.$$

Lemma A.1 can be derived as a corollary of the Motzkin Theorem of the Alternatives [186, p. 28], stated next.

Lemma A.2. *For any $A \in \mathbf{R}^{l\times n}$, $B \in \mathbf{R}^{m\times n}$, $B_0 \in \mathbf{R}^{m_0\times n}$, one and only one of the following statements holds: either there exists $x \in \mathbf{R}^n$ such that*

$$Ax = 0, \quad Bx \leq 0, \quad B_0x < 0,$$

or there exists $(y,\ z,\ z^0) \in \mathbf{R}^l \times \mathbf{R}^m \times \mathbf{R}^{m_0}$ such that

$$A^{\mathrm{T}}y + B^{\mathrm{T}}z + B_0^{\mathrm{T}}z^0 = 0, \quad z \geq 0, \quad z^0 \geq 0, \quad z^0 \neq 0.$$

The following simplified version of Lemma A.2, convenient in some applications, is known as the Gordan Theorem of the Alternatives.

Lemma A.3. *For any $B \in \mathbf{R}^{m_0\times n}$, one and only one of the following two alternatives is valid: either there exists $x \in \mathbf{R}^n$ such that*

$$B_0x < 0,$$

or there exists $z^0 \in \mathbf{R}^m$ such that

$$B_0^{\mathrm{T}}z^0 = 0, \quad z^0 \geq 0, \quad z^0 \neq 0.$$

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3,

The following is Hoffman's lemma giving a (global) error bound for linear systems (e.g., [27, Theorem 2.200]).

Lemma A.4. *For any $A \in \mathbf{R}^{l\times n}$, $a \in \mathbf{R}^l$, and $B \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, assume that the set*

$$S = \{x \in \mathbf{R}^n \mid Ax = a,\ Bx \le b\}$$

is nonempty.

Then there exists $c > 0$ such that

$$\operatorname{dist}(x,\ S) \le c(\|Ax - a\| + \|\max\{0,\ Bx - b\}\|) \quad \forall\, x \in \mathbf{R}^n.$$

The following is the finite-dimensional version of the classical Banach Open Mapping Theorem.

Lemma A.5. *For any $A \in \mathbf{R}^{l\times n}$, if $\operatorname{rank} A = l$, then there exists $c > 0$ such that for any $B \in \mathbf{R}^{l\times n}$ close enough to A, and any $y \in \mathbf{R}^l$, the equation*

$$Bx = y$$

has a solution $x(y)$ such that

$$\|x(y)\| \le c\|y\|.$$

This result is complemented by the more exact characterization of invertibility of small perturbations of a nonsingular matrix; see, e.g., [103, Theorem 2.3.4].

Lemma A.6. *Let $A \in \mathbf{R}^{n\times n}$ be a nonsingular matrix.*

Then any matrix $B \in \mathbf{R}^{n\times n}$ satisfying the inequality $\|B - A\| < 1/\|A^{-1}\|$ is nonsingular, and

$$\|B^{-1} - A^{-1}\| \le \frac{\|A^{-1}\|^2\|B - A\|}{1 - \|A^{-1}\|\|B - A\|}.$$

Lemma A.7 below is well known; it is sometimes called the Finsler Lemma [81], or the Debreu Lemma [54]. The similar in spirit Lemma A.8, on the other hand, is not standard, so we have to give its proof (from [150]). As the proofs of Lemmas A.7 and A.8 are somehow related, it makes sense to provide both.

Lemma A.7. *Let $H \in \mathbf{R}^{n\times n}$ be any symmetric matrix and $A \in \mathbf{R}^{l\times n}$ any matrix such that*

$$\langle H\xi,\ \xi\rangle > 0 \quad \forall\, \xi \in \ker A \setminus \{0\}. \tag{A.1}$$

Then the matrix $H + cA^{\mathrm{T}}A$ is positive definite for all $c \ge 0$ large enough.

Proof. We argue by contradiction. Suppose that there exist $\{c_k\} \subset \mathbf{R}$ and $\{\xi^k\} \subset \mathbf{R}^n$ such that $c_k \to +\infty$ as $k \to \infty$, and for all k it holds that $\|\xi^k\| = 1$ and

$$\langle (H + c_k A^{\mathrm{T}} A)\xi^k, \xi^k \rangle \leq 0. \tag{A.2}$$

Without loss of generality, we may assume that $\{\xi^k\} \to \xi$, with some $\xi \in \mathbf{R}^n \setminus \{0\}$. Dividing (A.2) by c_k and passing onto the limit as $k \to \infty$, we obtain that

$$0 \geq \langle A^{\mathrm{T}} A\xi, \xi \rangle = \|A\xi\|^2,$$

i.e., $\xi \in \ker A$.

On the other hand, since for each k it holds that

$$\langle A^{\mathrm{T}} A\xi^k, \xi^k \rangle = \|A\xi^k\|^2 \geq 0,$$

the inequality (A.2) implies that $\langle H\xi^k, \xi^k \rangle \leq 0$. Passing onto the limit as $k \to \infty$, we obtain that $\langle H\xi, \xi \rangle \leq 0$, in contradiction with (A.1). □

Another result, as already commented somewhat similar in nature to the Debreu–Finsler Lemma, is stated next.

Lemma A.8. *Let $H \in \mathbf{R}^{n\times n}$ and $A \in \mathbf{R}^{l\times n}$ be such that*

$$H\xi \notin \operatorname{im} A^{\mathrm{T}} \quad \forall\, \xi \in \ker A \setminus \{0\}. \tag{A.3}$$

Then for any $C > 0$, any $\tilde{H} \in \mathbf{R}^{n\times n}$ close enough to H, and any $\tilde{A} \in \mathbf{R}^{l\times n}$ close enough to A, the matrix $\tilde{H} + c(A + \Omega)^{\mathrm{T}} \tilde{A}$ is nonsingular for all $c \in \mathbf{R}$ such that $|c|$ is large enough, and for all $\Omega \in \mathbf{R}^{l\times n}$ satisfying $\|\Omega\| \leq C/|c|$.

Proof. Suppose the contrary, i.e., that there exist sequences $\{H_k\} \subset \mathbf{R}^{n\times n}$, $\{A_k\} \subset \mathbf{R}^{l\times n}$, $\{\Omega_k\} \subset \mathbf{R}^{l\times n}$, $\{c_k\} \subset \mathbf{R}$ and $\{\xi^k\} \subset \mathbf{R}^n \setminus \{0\}$, such that $\{H_k\} \to H$, $\{A_k\} \to A$, $|c_k| \to \infty$ as $k \to \infty$, and for all k it holds that $\|\Omega_k\| \leq C/|c_k|$ and

$$H_k\xi^k + c_k(A + \Omega_k)^{\mathrm{T}} A_k\xi^k = 0. \tag{A.4}$$

We can assume, without loss of generality, that $\|\xi^k\| = 1$ for all k, and $\{\xi^k\} \to \xi$, with some $\xi \in \mathbf{R}^n \setminus \{0\}$. Then since the right-hand side in

$$A^{\mathrm{T}} A_k\xi^k = -\frac{1}{c_k} H_k\xi^k - \Omega_k^{\mathrm{T}} A_k\xi^k,$$

tends to zero as $k \to \infty$, it must hold that $A^{\mathrm{T}} A\xi = 0$.

It is thus established that $A\xi \in \ker A^{\mathrm{T}}$, and since $A\xi \in \operatorname{im} A = (\ker A^{\mathrm{T}})^{\perp}$, this shows that $A\xi = 0$. Thus, $\xi \in \ker A \setminus \{0\}$.

On the other hand, (A.4) implies that the inclusion

$$H_k\xi^k + c_k\Omega_k^{\mathrm{T}} A_k\xi^k = -c_k A^{\mathrm{T}} A_k\xi^k \in \operatorname{im} A^{\mathrm{T}}$$

holds for all k, where the second term in the left-hand side tends to zero as $k \to \infty$ because $\{c_k \Omega_k\}$ is bounded and $\{A_k \xi^k\} \to A\xi = 0$. Hence, $H\xi \in \operatorname{im} A^{\mathrm{T}}$ by the closedness of $\operatorname{im} A^{\mathrm{T}}$. This gives a contradiction with (A.3). □

We complete this section by the following fact concerned with the existence of the inverse of a block matrix; see [243, Proposition 3.9].

Lemma A.9. *If $A \in \mathbf{R}^{n\times n}$ is a nonsingular matrix, $B \in \mathbf{R}^{n\times m}$, $C \in \mathbf{R}^{m\times n}$, $D \in \mathbf{R}^{m\times m}$, then for the matrix*

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

it holds that

$$\det M = \det A \det(D - CA^{-1}B).$$

Under the assumptions of Lemma A.9, the matrix $D - CA^{-1}B$ is referred to as the *Schur complement* of A in M.

A.2 Analysis

Our use of the big-O and little-o notation employs the following conventions. For a mapping $F : \mathbf{R}^n \to \mathbf{R}^m$ and a function $\varphi : \mathbf{R}^n \to \mathbf{R}_+$, and for a given $\bar{x} \in \mathbf{R}^n$, we write $F(x) = O(\varphi(x))$ as $x \to \bar{x}$ if there exists $c > 0$ such that $\|F(x)\| \le c\varphi(x)$ for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$. We write $F(x) = o(\varphi(x))$ as $x \to \bar{x}$ if for every $\varepsilon > 0$ no matter how small, it holds that $\|F(x)\| \le \varepsilon\varphi(x)$ for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$. For sequences $\{x^k\} \subset \mathbf{R}^n$ and $\{t_k\} \subset \mathbf{R}_+$, by $x^k = O(t_k)$ as $k \to \infty$ we mean that there exists $c > 0$ such that $\|x^k\| \le ct_k$ for all k large enough. Accordingly, $x^k = o(t_k)$ as $k \to \infty$ if for every $\varepsilon > 0$ no matter how small, it holds that $\|x^k\| \le \varepsilon t_k$ for all k large enough. For a sequence $\{\tau_k\} \subset \mathbf{R}$, we write $\tau^k \le o(t_k)$ as $k \to \infty$ if for any $\varepsilon > 0$ no matter how small it holds that $\tau_k \le \varepsilon t_k$ for all k large enough.

Concerning convergence rate estimates, the terminology is as follows. Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to some $\bar{x} \in \mathbf{R}^n$. If there exist $q \in (0, 1)$ and $c > 0$ such that

$$\|x^k - \bar{x}\| \le cq^k$$

for all k large enough (or, in other words, $\|x^k - \bar{x}\| = O(q^k)$ as $k \to \infty$), then we say that $\{x^k\}$ has *geometric convergence rate*. If there exists $q \in (0, 1)$ such that

$$\|x^{k+1} - \bar{x}\| \le q\|x^k - \bar{x}\| \tag{A.5}$$

for all k large enough, then we say that $\{x^k\}$ has *linear convergence rate*. Linear rate implies geometric rate, but the converse is not true. If for every $q \in (0, 1)$, no matter how small, the inequality (A.5) holds for all k large

enough, then we say that $\{x^k\}$ has *superlinear convergence rate*. To put it in other words, superlinear convergence means that

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

as $k \to \infty$. A particular case of the superlinear rate is *quadratic convergence rate*, meaning that there exists $c > 0$ such that

$$\|x^{k+1} - \bar{x}\| \leq c\|x^k - \bar{x}\|^2$$

for all k large enough or, in other words,

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

as $k \to \infty$. Unlike superlinear or quadratic rate, linear convergence rate depends on the norm: linear convergence rate in some norm in $\mathbf{R}^n$ does not necessarily imply linear convergence rate in a different norm.

We next state some facts and notions of differential calculus for mappings (generally vector-valued and with vector variable). It is assumed that the reader is familiar with differential calculus for scalar-valued functions in a scalar variable.

The mapping $F : \mathbf{R}^n \to \mathbf{R}^m$ is said to be *differentiable at* $x \in \mathbf{R}^n$ if there exists a matrix $J \in \mathbf{R}^{m \times n}$ such that for $\xi \in \mathbf{R}^n$ is holds that

$$F(x + \xi) = F(x) + J\xi + o(\|\xi\|)$$

as $\xi \to 0$. The matrix J with this property is necessarily unique; it coincides with the *Jacobian* $F'(x)$ (the matrix of first partial derivatives of the components of F at x with respect to all the variables), and it is also called the *first derivative of* F *at* x. The rows of the Jacobian are the *gradients* $F_1'(x), \ldots, F_m'(x)$ (vectors of first partial derivatives with respect to all variables) of the components of F at x.

The mapping $F : \mathbf{R}^n \to \mathbf{R}^m$ is (*continuously*) *differentiable on a set* $S \subset \mathbf{R}^n$ if it is differentiable at every point of some open set $O \subset \mathbf{R}^n$ such that $S \subset O$ (and the mapping $F'(\cdot)$ defined on O is continuous at every point of S).

The mapping $F : \mathbf{R}^n \to \mathbf{R}^m$ is said to be *twice differentiable at* $x \in \mathbf{R}^n$ if it is differentiable in a neighborhood of x, and the mapping $F'(\cdot)$ defined on this neighborhood is differentiable at x. The derivative $(F')'(x)$ of $F'(\cdot)$ at x can be regarded as a linear operator from $\mathbf{R}^n$ to $\mathbf{R}^{m \times n}$, or alternatively, as a bilinear mapping $F''(x) : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^m$ defined by

$$F''(x)[\xi^1, \xi^2] = ((F')'(x)\xi^1)\xi^2, \quad \xi^1, \xi^2 \in \mathbf{R}^n.$$

This bilinear mapping is necessarily symmetric, that is,

$$F''(x)[\xi^1, \xi^2] = F''(x)[\xi^2, \xi^1] \quad \forall\, \xi^1, \xi^2 \in \mathbf{R}^n.$$

The mapping in question is called the *second derivative of F at x*, and it is comprised by the *Hessians* $F_1''(x), \ldots, F_m''(x)$ (the matrices of second partial derivatives) of the components of F at x:

$$F''(x)[\xi^1, \xi^2] = (\langle F_1''(x)\xi^1, \xi^2\rangle, \ldots, \langle F_m''(x)\xi^1, \xi^2\rangle) \quad \forall\, \xi^1, \xi^2 \in \mathbf{R}^n.$$

Note that the symmetry of the bilinear mapping $F''(x)$ is equivalent to the symmetry of the Hessians of the components of F.

If F is twice differentiable at $\bar{x}$, then for $\xi \in \mathbf{R}^n$ it holds that

$$F(x+\xi) = F(x) + F'(x)\xi + \frac{1}{2}F'(x)[\xi, \xi] + o(\|\xi\|^2)$$

as $\xi \to 0$. This fact can be regarded as a particular case of the Taylor formula.

The mapping F is *twice (continuously) differentiable on a set* $S \subset \mathbf{R}^n$ if it is twice differentiable at every point of some open set $O \subset \mathbf{R}^n$ such that $S \subset O$ (and the mapping $F''(\cdot)$ defined on the set O is continuous at every point of S).

Furthermore, the mapping $F : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^m$ is said to be *differentiable at $(x, y) \in \mathbf{R}^n \times \mathbf{R}^l$ with respect to x* if the mapping $F(\cdot, y)$ is differentiable at x. The derivative of the latter mapping at x is called the *partial derivative of F with respect to x at (x, y)*, and it is denoted by $\frac{\partial F}{\partial x}(x, y)$.

Similarly, the mapping $F : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^m$ is said to be *twice differentiable at $(x, y) \in \mathbf{R}^n \times \mathbf{R}^l$ with respect to x* if the mapping $F(\cdot, y)$ is twice differentiable at x. The second derivative of the latter mapping at x is called the *second partial derivative of F with respect to x at (x, y)*, and it is denoted by $\frac{\partial^2 F}{\partial x^2}(x, y)$.

In this book, any of the itemized assertions in the next statement is referred to as a mean-value theorem. The first part of item (a) is a rather subtle and not widely known result; it was established in [198]. The other statements are fairly standard.

Theorem A.10. *For any $F : \mathbf{R}^n \to \mathbf{R}^m$ and any $x^1, x^2 \in \mathbf{R}^n$, the following assertions are valid:*

(a) If F is continuous on $[x^1, x^2] = \{tx^1 + (1-t)x^2 \mid t \in [0, 1]\}$ and differentiable on $(x^1, x^2) = \{tx^1 + (1-t)x^2 \mid t \in (0, 1)\}$, then there exist $t_i \in (0, 1)$ and $\theta_i \geq 0$, $i = 1, \ldots, m$, such that $\sum_{i=1}^m \theta_i = 1$ and

$$F(x^1) - F(x^2) = \sum_{i=1}^m \theta_i F'(t_i x^1 + (1-t_i)x^2)(x^1 - x^2),$$

and in particular,

$$\|F(x^1) - F(x^2)\| \leq \sup_{t \in (0,1)} \|F'(tx^1 + (1-t)x^2)\| \|x^1 - x^2\|.$$

(b) If F is continuously differentiable on the line segment $[x^1, x^2]$, then

$$F(x^1) - F(x^2) = \int_0^1 F'(tx^1 + (1-t)x^2)(x^1 - x^2)\, dt.$$

The next fact is an immediate corollary of assertion (b) of Theorem A.10.

Lemma A.11. *For any $F : \mathbf{R}^n \to \mathbf{R}^m$ and any $x^1, x^2 \in \mathbf{R}^n$, if F is differentiable on the line segment $[x^1, x^2]$, with its derivative being Lipschitz-continuous on this segment with a constant $L > 0$, then*

$$\|F(x^1) - F(x^2) - F'(x^2)(x^1 - x^2)\| \le \frac{L}{2}\|x^1 - x^2\|^2.$$

A.3 Convexity and Monotonicity

For a detailed exposition of finite-dimensional convex analysis, we refer to [235]. In this section we only recall some basic definitions and facts used in this book. For details on (maximal) monotone mappings and related issues, we refer to [17, 30] and [239, Chap. 12].

For a finite number of points $x^1, \ldots, x^m \in \mathbf{R}^n$, their *convex combinations* are points of the form $\sum_{i=1}^m t_i x^i$ with some $t_i \ge 0$, $i = 1, \ldots, m$, such that $\sum_{i=1}^m t_i = 1$. In particular, convex combinations of two points $x^1, x^2 \in \mathbf{R}^n$ are points of the form $tx^1 + (1-t)x^2$, $t \in [0, 1]$, and they form a line segment connecting x^1 and x^2.

A set $S \subset \mathbf{R}^n$ is said to be *convex* if for each pair of points $x^1, x^2 \in S$ all convex combinations of these points belong to S (equivalently, for any points $x^1, \ldots, x^m \in S$, where $m \ge 2$, all convex combinations of these points belong to S).

The *convex hull* of a set $S \subset \mathbf{R}^n$, denoted by $\operatorname{conv} S$, is the smallest convex set in $\mathbf{R}^n$ that contains S (equivalently, the set of all convex combinations of points in S).

By a (Euclidean) *projection* of a point $x \in \mathbf{R}^n$ onto a given set $S \subset \mathbf{R}^n$ we mean a point closest to x among all the points in S, i.e., any global solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & \|y - x\| \\ \text{subject to} & y \in S. \end{array} \tag{A.6}$$

As the objective function in (A.6) is coercive, projection of any point onto any nonempty closed set in $\mathbf{R}^n$ exists. If, in addition, the set is convex, then the following holds.

Lemma A.12. *Let $S \subset \mathbf{R}^n$ be any nonempty closed convex set.*

Then the projection operator onto S, $\pi_S : \mathbf{R}^n \to S$, is well defined and single valued: for any point $x \in \mathbf{R}^n$ its projection $\pi_S(x)$ onto S exists and is unique. Moreover, $\bar{x} = \pi_S(x)$ if and only if

$$\bar{x} \in S, \quad \langle x - \bar{x}, y - \bar{x} \rangle \le 0 \quad \forall\, y \in S.$$

In addition, the projection operator is nonexpansive:

$$\|\pi_S(x^1) - \pi_S(x^2)\| \le \|x^1 - x^2\| \quad \forall\, x^1, x^2 \in \mathbf{R}^n.$$

A set $C \subset \mathbf{R}^n$ is called a *cone* if for each $x \in C$ it contains all points of the form tx, $t \ge 0$. The *polar cone* to C is defined by

$$C^\circ = \{\xi \in \mathbf{R}^n \mid \langle \xi, x \rangle \le 0 \ \forall\, x \in C\}.$$

Lemma A.13. *For any nonempty closed convex cone $C \subset \mathbf{R}^n$ it holds that*

$$x = \pi_C(x) + \pi_{C^\circ}(x) \quad \forall\, x \in \mathbf{R}^n,$$

and in particular,

$$C^\circ = \{x \in \mathbf{R}^n \mid \pi_C(x) = 0\},$$
$$\pi_C(x - \pi_C(x)) = 0 \quad \forall\, x \in \mathbf{R}^n.$$

An important property concerns separation of (convex) sets by hyperplanes. The following separation theorem can be found in [235, Corollary 11.4.2].

Theorem A.14. *Let $S_1, S_2 \subset \mathbf{R}^n$ be nonempty closed convex sets, with at least one of them being also bounded (hence, compact).*

Then $S_1 \cap S_2 = \emptyset$ if and only if there exist $\xi \in \mathbf{R}^n \setminus \{0\}$ and $t \in \mathbf{R}$ such that

$$\langle \xi, x^1 \rangle < t < \langle \xi, x^2 \rangle \quad \forall\, x^1 \in S_1, x^2 \in S_2.$$

Given a convex set $S \subset \mathbf{R}^n$, a function $f : S \to \mathbf{R}$ is said to be *convex* (on the set S) if

$$f(tx^1 + (1-t)x^2) \le tf(x^1) + (1-t)f(x^2) \quad \forall\, x^1, x^2 \in S, \ \forall\, t \in [0, 1].$$

Equivalently, f is convex if its *epigraph* $\{(x, t) \in S \times \mathbf{R} \mid f(x) \le t\}$ is a convex set. It is immediate that a linear combination of convex functions with nonnegative coefficients is a convex function, and the maximum over a finite family of convex functions is a convex function.

Furthermore, f is said to be *strongly convex* (on S) if there exists $\gamma > 0$ such that

$$f(tx^1 + (1-t)x^2) \le tf(x^1) + (1-t)f(x^2) - \gamma t(1-t)\|x^1 - x^2\|^2$$
$$\forall\, x^1, x^2 \in S, \ \forall\, t \in [0, 1].$$

The sum of a convex function and a strongly convex function is evidently strongly convex. The following are characterizations of convexity for smooth functions.

Proposition A.15. *Let $O \subset \mathbf{R}^n$ be a nonempty open convex set, and let $f : O \to \mathbf{R}$ be differentiable in O.*

Then the following items are equivalent:

(a) The function f is convex on O.
(b) $f(x^1) \geq f(x^2) + \langle f'(x^2),\, x^1 - x^2\rangle$ for all $x^1, x^2 \in O$.
(c) $\langle f'(x^1) - f'(x^2),\, x^1 - x^2\rangle \geq 0$ for all $x^1, x^2 \in O$.

If f is twice differentiable in O, then the properties above are further equivalent to

(d) The Hessian $f''(x)$ is positive semidefinite for all $x \in O$.

It is clear that a quadratic function is convex if and only if its (constant) Hessian is a positive semidefinite matrix (see item (d) in Proposition A.15). Moreover, a quadratic function is strongly convex if and only if its Hessian is positive definite.

Given a convex function $f : \mathbf{R}^n \to \mathbf{R}$, an element $a \in \mathbf{R}^n$ is called a *subgradient* of f at a point $x \in \mathbf{R}^n$ if

$$f(y) \geq f(x) + \langle a,\, y - x\rangle \quad \forall\, y \in \mathbf{R}^n.$$

The set of all the elements $a \in \mathbf{R}^n$ with this property is called the *subdifferential* of f at x, denoted by $\partial f(x)$.

Proposition A.16. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be convex on $\mathbf{R}^n$.*

Then for each $x \in \mathbf{R}^n$ the subdifferential $\partial f(x)$ is a nonempty compact convex set. Moreover, f is continuous and directionally differentiable at every $x \in \mathbf{R}^n$ in every direction $\xi \in \mathbf{R}^n$, and it holds that

$$f'(x;\, \xi) = \max_{y \in \partial f(x)} \langle y,\, \xi\rangle.$$

For some further calculus rules for subdifferentials see Sect. 1.4.1, where they are presented in a more general (not necessarily convex) setting. Section 1.4.1 provides all the necessary material for the convex calculus in this book.

We complete this section by some definitions and facts concerned with the notion of monotonicity for (multi)functions. For a (generally) set-valued mapping Ψ from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$, define its domain

$$\operatorname{dom} \Psi = \{x \in \mathbf{R}^n \mid \Psi(x) \neq \emptyset\}.$$

Then Ψ is said to be *monotone* if

$$\langle y^1 - y^2,\, x^1 - x^2\rangle \geq 0 \quad \forall\, y^1 \in \Psi(x^1),\ \forall\, y^2 \in \Psi(x^2),\ \forall\, x^1,\, x^2 \in \operatorname{dom} \Psi,$$

and *maximal monotone* if, in addition, its *graph* $\{(x, y) \in \mathbf{R}^n \times \mathbf{R}^n \mid y \in \Psi(x)\}$ is not contained in the graph of any other monotone set-valued mapping.

Some examples of maximal monotone mappings are: a continuous monotone function $F : \mathbf{R} \to \mathbf{R}$, the subdifferential multifunction $\partial f(\cdot)$ of a convex function $f : \mathbf{R}^n \to \mathbf{R}$, and the normal cone multifunction $N_S(\cdot)$ for a closed convex set $S \subset \mathbf{R}^n$.

The sum of two monotone mappings is monotone, and the sum of two maximal monotone mappings is maximal monotone if the domain of one intersects the interior of the domain of the other.

Furthermore, Ψ is said to be *strongly monotone* if there exists $\gamma > 0$ such that $\Psi - \gamma I$ is monotone, which is equivalent to the property

$$\langle y^1 - y^2,\, x^1 - x^2 \rangle \geq \gamma \|x^1 - x^2\|^2 \quad \forall\, y^1 \in \Psi(x^1),\, \forall\, y^2 \in \Psi(x^2),\, \forall\, x^1,\, x^2 \in \operatorname{dom} \Psi.$$

In particular, the identity mapping I is strongly monotone, and the sum of a monotone mapping and a strongly monotone mapping is strongly monotone.

The following characterization of monotonicity for smooth mappings can be found, e.g., in [239, Proposition 12.3].

Proposition A.17. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable on $\mathbf{R}^n$.*

Then F is monotone if and only if $F'(x)$ is positive semidefinite for all $x \in \mathbf{R}^n$.

We refer to [17, 30] and [239, Chap. 12] for other details on (maximal) monotone mappings and related issues.