CHAPTER 4

Trust-Region Methods

Line search methods and trust-region methods both generate steps with the help of a quadratic model of the objective function, but they use this model in different ways. Line search methods use it to generate a search direction, and then focus their efforts on finding a suitable step length α along this direction. Trust-region methods define a region around the current iterate within which they *trust* the model to be an adequate representation of the objective function, and then choose the step to be the approximate minimizer of the model in this region. In effect, they choose the direction and length of the step simultaneously. If a step is not acceptable, they reduce the size of the region and find a new minimizer. In general, the direction of the step changes whenever the size of the trust region is altered.

The size of the trust region is critical to the effectiveness of each step. If the region is too small, the algorithm misses an opportunity to take a substantial step that will move it much closer to the minimizer of the objective function. If too large, the minimizer of the model may be far from the minimizer of the objective function in the region, so we may have to reduce the size of the region and try again. In practical algorithms, we choose the size of the region according to the performance of the algorithm during previous iterations. If the model is consistently reliable, producing good steps and accurately predicting the behavior of the objective function along these steps, the size of the trust region may be increased to allow longer, more ambitious, steps to be taken. A failed step is an indication that our model is an inadequate representation of the objective function over the current trust region. After such a step, we reduce the size of the region and try again.

Figure 4.1 illustrates the trust-region approach on a function f of two variables in which the current point x_k and the minimizer x^* lie at opposite ends of a curved valley. The quadratic model function m_k, whose elliptical contours are shown as dashed lines, is constructed from function and derivative information at x_k and possibly also on information accumulated from previous iterations and steps. A line search method based on this model searches along the step to the minimizer of m_k (shown), but this direction will yield at most a small reduction in f, even if the optimal steplength is used. The trust-region method steps to the minimizer of m_k within the dotted circle (shown), yielding a more significant reduction in f and better progress toward the solution.

In this chapter, we will assume that the model function m_k that is used at each iterate x_k is quadratic. Moreover, m_k is based on the Taylor-series expansion of f around

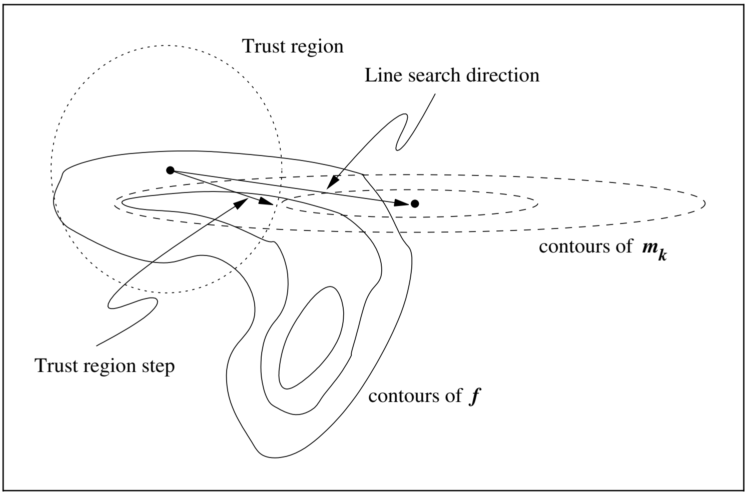

Figure 4.1 Trust-region and line search steps.

x_k, which is

$$f(x_k + p) = f_k + g_k^T p + \tfrac{1}{2} p^T \nabla^2 f(x_k + tp) p, \tag{4.1}$$

where $f_k = f(x_k)$ and $g_k = \nabla f(x_k)$, and t is some scalar in the interval $(0, 1)$. By using an approximation B_k to the Hessian in the second-order term, m_k is defined as follows:

$$m_k(p) = f_k + g_k^T p + \tfrac{1}{2} p^T B_k p, \tag{4.2}$$

where B_k is some symmetric matrix. The difference between $m_k(p)$ and $f(x_k + p)$ is $O\left(\|p\|^2\right)$, which is small when p is small.

When B_k is equal to the true Hessian $\nabla^2 f(x_k)$, the approximation error in the model function m_k is $O\left(\|p\|^3\right)$, so this model is especially accurate when $\|p\|$ is small. This choice $B_k = \nabla^2 f(x_k)$ leads to the trust-region Newton method, and will be discussed further in Section 4.4. In other sections of this chapter, we emphasize the generality of the trust-region approach by assuming little about B_k except symmetry and uniform boundedness.

To obtain each step, we seek a solution of the subproblem

$$\min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \tfrac{1}{2} p^T B_k p \qquad \text{s.t. } \|p\| \le \Delta_k, \tag{4.3}$$

where $\Delta_k > 0$ is the trust-region radius. In most of our discussions, we define $\| \cdot \|$ to be the Euclidean norm, so that the solution p_k^* of (4.3) is the minimizer of m_k in the ball of radius Δ_k. Thus, the trust-region approach requires us to solve a sequence of subproblems (4.3) in which the objective function and constraint (which can be written as $p^T p \le \Delta_k^2$) are both quadratic. When B_k is positive definite and $\|B_k^{-1} g_k\| \le \Delta_k$, the solution of (4.3) is easy to identify—it is simply the unconstrained minimum $p_k^{\mathrm{B}} = -B_k^{-1} g_k$ of the quadratic $m_k(p)$. In this case, we call p_k^{B} the *full step*. The solution of (4.3) is not so obvious in other cases, but it can usually be found without too much computational expense. In any case, as described below, we need only an *approximate* solution to obtain convergence and good practical behavior.

OUTLINE OF THE TRUST-REGION APPROACH

One of the key ingredients in a trust-region algorithm is the strategy for choosing the trust-region radius Δ_k at each iteration. We base this choice on the agreement between the model function m_k and the objective function f at previous iterations. Given a step p_k we define the ratio

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}; \tag{4.4}$$

the numerator is called the *actual reduction*, and the denominator is the *predicted reduction* (that is, the reduction in f predicted by the model function). Note that since the step p_k

is obtained by minimizing the model m_k over a region that includes $p = 0$, the predicted reduction will always be nonnegative. Hence, if ρ_k is negative, the new objective value $f(x_k + p_k)$ is greater than the current value $f(x_k)$, so the step must be rejected. On the other hand, if ρ_k is close to 1, there is good agreement between the model m_k and the function f over this step, so it is safe to expand the trust region for the next iteration. If ρ_k is positive but significantly smaller than 1, we do not alter the trust region, but if it is close to zero or negative, we shrink the trust region by reducing Δ_k at the next iteration.

The following algorithm describes the process.

Algorithm 4.1 (Trust Region).

```
Given $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$, and $\eta \in \left[0, \frac{1}{4}\right)$:
for $k = 0, 1, 2, \ldots$
    Obtain $p_k$ by (approximately) solving (4.3);
    Evaluate $\rho_k$ from (4.4);
    if $\rho_k < \frac{1}{4}$
        $\Delta_{k+1} = \frac{1}{4}\Delta_k$
    else
        if $\rho_k > \frac{3}{4}$ and $\|p_k\| = \Delta_k$
            $\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$
        else
            $\Delta_{k+1} = \Delta_k$;
    if $\rho_k > \eta$
        $x_{k+1} = x_k + p_k$
    else
        $x_{k+1} = x_k$;
end (for).
```

Here $\hat{\Delta}$ is an overall bound on the step lengths. Note that the radius is increased only if $\|p_k\|$ actually reaches the boundary of the trust region. If the step stays strictly inside the region, we infer that the current value of Δ_k is not interfering with the progress of the algorithm, so we leave its value unchanged for the next iteration.

To turn Algorithm 4.1 into a practical algorithm, we need to focus on solving the trust-region subproblem (4.3). In discussing this matter, we sometimes drop the iteration subscript k and restate the problem (4.3) as follows:

$$\min_{p \in \mathbb{R}^n} m(p) \overset{\text{def}}{=} f + g^T p + \tfrac{1}{2} p^T B p \qquad \text{s.t. } \|p\| \le \Delta. \tag{4.5}$$

A first step to characterizing exact solutions of (4.5) is given by the following theorem (due to Moré and Sorensen [214]), which shows that the solution p^* of (4.5) satisfies

$$(B + \lambda I)p^* = -g \tag{4.6}$$

for some $\lambda \ge 0$.

Theorem 4.1.

The vector p^* is a global solution of the trust-region problem

$$\min_{p\in\mathbb{R}^n} m(p) = f + g^T p + \tfrac{1}{2} p^T B p, \qquad \text{s.t. } \|p\| \le \Delta, \tag{4.7}$$

if and only if p^* is feasible and there is a scalar $\lambda \ge 0$ such that the following conditions are satisfied:

$$(B + \lambda I)p^* = -g, \tag{4.8a}$$

$$\lambda(\Delta - ||p^*||) = 0, \tag{4.8b}$$

$$(B + \lambda I) \quad \text{is positive semidefinite.} \tag{4.8c}$$

We delay the proof of this result until Section 4.3, and instead discuss just its key features here with the help of Figure 4.2. The condition (4.8b) is a complementarity condition that states that at least one of the nonnegative quantities λ and $(\Delta - \|p^*\|)$ must be zero. Hence, when the solution lies strictly inside the trust region (as it does when $\Delta = \Delta_1$ in Figure 4.2), we must have $\lambda = 0$ and so $Bp^* = -g$ with B positive semidefinite, from (4.8a) and (4.8c), respectively. In the other cases $\Delta = \Delta_2$ and $\Delta = \Delta_3$, we have $\|p^*\| = \Delta$, and so λ is allowed to take a positive value. Note from (4.8a) that

$$\lambda p^* = -Bp^* - g = -\nabla m(p^*).$$

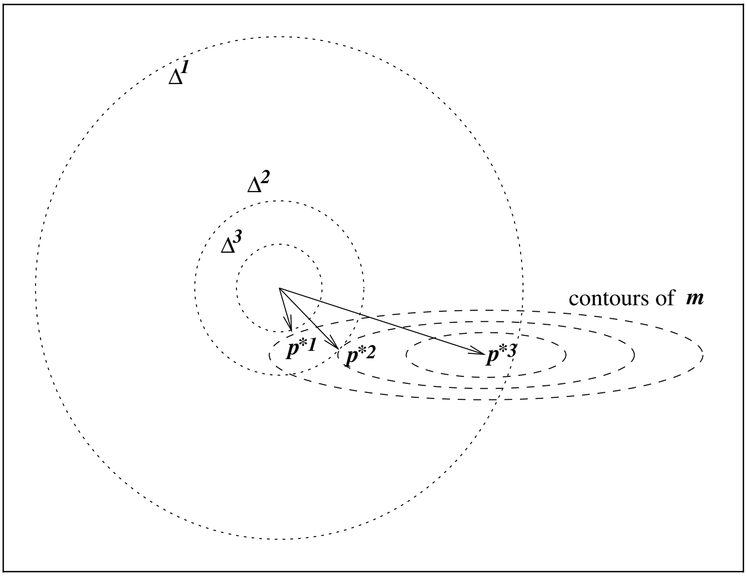

Figure 4.2 Solution of trust-region subproblem for different radii Δ^1, Δ^2, Δ^3.

Thus, when $\lambda > 0$, the solution p^* is collinear with the negative gradient of m and normal to its contours. These properties can be seen in Figure 4.2.

In Section 4.1, we describe two strategies for finding *approximate* solutions of the subproblem (4.3), which achieve at least as much reduction in m_k as the reduction achieved by the so-called *Cauchy point*. This point is simply the minimizer of m_k along the steepest descent direction $-g_k$. subject to the trust-region bound. The first approximate strategy is the *dogleg method*, which is appropriate when the model Hessian B_k is positive definite. The second strategy, known as *two-dimensional subspace minimization*, can be applied when B_k is indefinite, though it requires an estimate of the most negative eigenvalue of this matrix. A third strategy, described in Section 7.1, uses an approach based on the conjugate gradient method to minimize m_k, and can therefore be applied when B is large and sparse.

Section 4.3 is devoted to a strategy in which an iterative method is used to identify the value of λ for which (4.6) is satisfied by the solution of the subproblem. We prove global convergence results in Section 4.2. Section 4.4 discusses the trust-region Newton method, in which the Hessian B_k of the model function is equal to the Hessian $\nabla^2 f(x_k)$ of the objective function. The key result of this section is that, when the trust-region Newton algorithm converges to a point x^* satisfying second-order sufficient conditions, it converges superlinearly.

4.1 ALGORITHMS BASED ON THE CAUCHY POINT

THE CAUCHY POINT

As we saw in Chapter 3, line search methods can be globally convergent even when the optimal step length is not used at each iteration. In fact, the step length α_k need only satisfy fairly loose criteria. A similar situation applies in trust-region methods. Although in principle we seek the optimal solution of the subproblem (4.3), it is enough for purposes of global convergence to find an approximate solution p_k that lies within the trust region and gives a *sufficient reduction* in the model. The sufficient reduction can be quantified in terms of the Cauchy point, which we denote by p_k^{C} and define in terms of the following simple procedure.

Algorithm 4.2 (Cauchy Point Calculation).

Find the vector p_k^{S} that solves a linear version of (4.3), that is,

$$p_k^{\mathrm{S}} = \arg\min_{p \in \mathbb{R}^n} f_k + g_k^T p \qquad \text{s.t. } \|p\| \le \Delta_k; \tag{4.9}$$

Calculate the scalar $\tau_k > 0$ that minimizes $m_k(\tau p_k^{\mathrm{S}})$ subject to satisfying the trust-region bound, that is,

$$\tau_k = \arg\min_{\tau \ge 0} m_k(\tau p_k^{\mathrm{S}}) \qquad \text{s.t. } \|\tau p_k^{\mathrm{S}}\| \le \Delta_k; \tag{4.10}$$

Set $p_k^{\mathrm{C}} = \tau_k p_k^{\mathrm{S}}$.

It is easy to write down a closed-form definition of the Cauchy point. For a start, the solution of (4.9) is simply

$$p_k^{\text{s}} = -\frac{\Delta_k}{\|g_k\|} g_k.$$

To obtain τ_k explicitly, we consider the cases of $g_k^T B_k g_k \leq 0$ and $g_k^T B_k g_k > 0$ separately. For the former case, the function $m_k(\tau p_k^{\text{s}})$ decreases monotonically with τ whenever $g_k \neq 0$, so τ_k is simply the largest value that satisfies the trust-region bound, namely, $\tau_k = 1$. For the case $g_k^T B_k g_k > 0$, $m_k(\tau p_k^{\text{s}})$ is a convex quadratic in τ, so τ_k is either the unconstrained minimizer of this quadratic, $\|g_k\|^3/(\Delta_k g_k^T B_k g_k)$, or the boundary value 1, whichever comes first. In summary, we have

$$p_k^{\text{C}} = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k, \tag{4.11}$$

where

$$\tau_k = \begin{cases} 1 & \text{if } g_k^T B_k g_k \leq 0; \\ \min\left(\|g_k\|^3/(\Delta_k g_k^T B_k g_k), 1\right) & \text{otherwise.} \end{cases} \tag{4.12}$$

Figure 4.3 illustrates the Cauchy point for a subproblem in which B_k is positive definite. In this example, p_k^{C} lies strictly inside the trust region.

The Cauchy step p_k^{C} is inexpensive to calculate—no matrix factorizations are required—and is of crucial importance in deciding if an approximate solution of the trust-region subproblem is acceptable. Specifically, a trust-region method will be globally

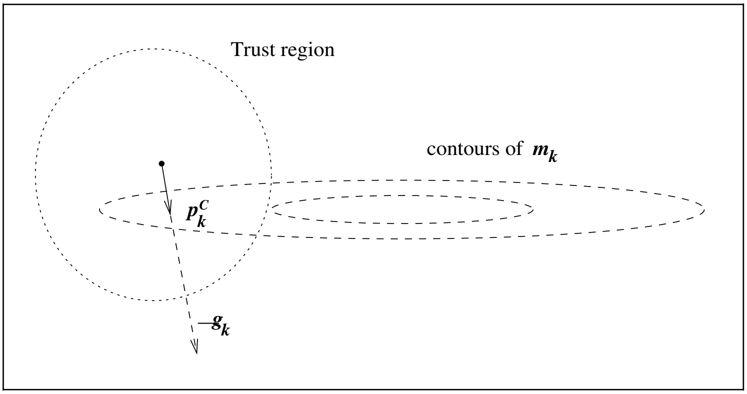

Figure 4.3 The Cauchy point.

convergent if its steps p_k give a reduction in the model m_k that is at least some fixed positive multiple of the decrease attained by the Cauchy step.

IMPROVING ON THE CAUCHY POINT

Since the Cauchy point p_k^{C} provides sufficient reduction in the model function m_k to yield global convergence, and since the cost of calculating it is so small, why should we look any further for a better approximate solution of (4.3)? The reason is that by always taking the Cauchy point as our step, we are simply implementing the steepest descent method with a particular choice of step length. As we have seen in Chapter 3, steepest descent performs poorly even if an *optimal* step length is used at each iteration.

The Cauchy point does not depend very strongly on the matrix B_k, which is used only in the calculation of the step length. Rapid convergence can be expected only if B_k plays a role in determining the *direction* of the step as well as its length, and if B_k contains valid curvature information about the function.

A number of trust-region algorithms compute the Cauchy point and then try to improve on it. The improvement strategy is often designed so that the full step $p_k^{\text{B}} = -B_k^{-1} g_k$ is chosen whenever B_k is positive definite and $\|p_k^{\text{B}}\| \leq \Delta_k$. When B_k is the exact Hessian $\nabla^2 f(x_k)$ or a quasi-Newton approximation, this strategy can be expected to yield superlinear convergence.

We now consider three methods for finding approximate solutions to (4.3) that have the features just described. Throughout this section we will be focusing on the internal workings of a single iteration, so we simplify the notation by dropping the subscript "k" from the quantities Δ_k, p_k, m_k, and g_k and refer to the formulation (4.5) of the subproblem. In this section, we denote the solution of (4.5) by $p^*(\Delta)$, to emphasize the dependence on Δ.

THE DOGLEG METHOD

The first approach we discuss goes by the descriptive title of the *dogleg method*. It can be used when B is positive definite.

To motivate this method, we start by examining the effect of the trust-region radius Δ on the solution $p^*(\Delta)$ of the subproblem (4.5). When B is positive definite, we have already noted that the unconstrained minimizer of m is $p^{\text{B}} = -B^{-1}g$. When this point is feasible for (4.5), it is obviously a solution, so we have

$$p^*(\Delta) = p^{\text{B}}, \qquad \text{when } \Delta \geq \|p^{\text{B}}\|. \tag{4.13}$$

When Δ is small relative to p^{B}, the restriction $\|p\| \leq \Delta$ ensures that the quadratic term in m has little effect on the solution of (4.5). For such Δ, we can get an approximation to $p(\Delta)$

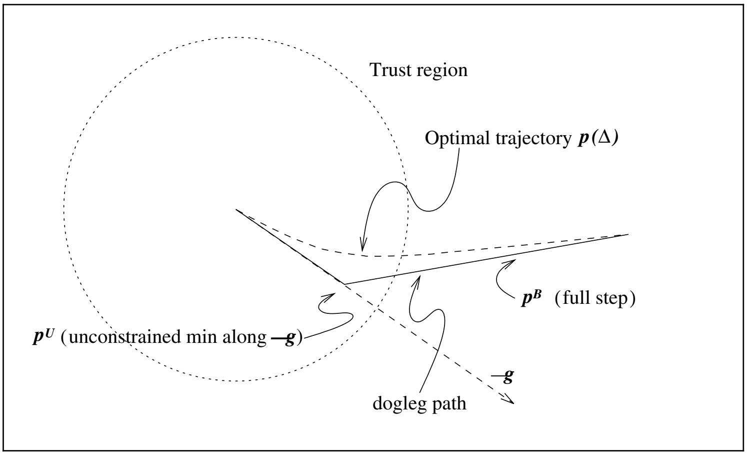
Trust region
Optimal trajectory $p(\Delta)$
p^B (full step)
p^U (unconstrained min along $-g$)
dogleg path
$-g$

Figure 4.4 Exact trajectory and dogleg approximation.

by simply omitting the quadratic term from (4.5) and writing

$$p^*(\Delta) \approx -\Delta \frac{g}{\|g\|}, \qquad \text{when } \Delta \text{ is small.} \tag{4.14}$$

For intermediate values of Δ, the solution $p^*(\Delta)$ typically follows a curved trajectory like the one in Figure 4.4.

The dogleg method finds an approximate solution by replacing the curved trajectory for $p^*(\Delta)$ with a path consisting of two line segments. The first line segment runs from the origin to the minimizer of m along the steepest descent direction, which is

$$p^{\mathrm{U}} = -\frac{g^T g}{g^T B g} g, \tag{4.15}$$

while the second line segment runs from p^{U} to p^{B} (see Figure 4.4). Formally, we denote this trajectory by $\tilde{p}(\tau)$ for $\tau \in [0, 2]$, where

$$\tilde{p}(\tau) = \begin{cases} \tau p^{\mathrm{U}}, & 0 \le \tau \le 1, \\ p^{\mathrm{U}} + (\tau - 1)(p^{\mathrm{B}} - p^{\mathrm{U}}), & 1 \le \tau \le 2. \end{cases} \tag{4.16}$$

The dogleg method chooses p to minimize the model m along this path, subject to the trust-region bound. The following lemma shows that the minimum along the dogleg path can be found easily.

Lemma 4.2.

Let B be positive definite. Then

(i) $\|\tilde{p}(\tau)\|$ *is an increasing function of* τ*, and*

(ii) $m(\tilde{p}(\tau))$ *is a decreasing function of* τ*.*

Proof. It is easy to show that (i) and (ii) both hold for $\tau \in [0, 1]$, so we restrict our attention to the case of $\tau \in [1, 2]$. For (i), define $h(\alpha)$ by

$$\begin{aligned} h(\alpha) &= \tfrac{1}{2}\|\tilde{p}(1+\alpha)\|^2 \\ &= \tfrac{1}{2}\|p^{\mathrm{U}} + \alpha(p^{\mathrm{B}} - p^{\mathrm{U}})\|^2 \\ &= \tfrac{1}{2}\|p^{\mathrm{U}}\|^2 + \alpha(p^{\mathrm{U}})^T(p^{\mathrm{B}} - p^{\mathrm{U}}) + \tfrac{1}{2}\alpha^2\|p^{\mathrm{B}} - p^{\mathrm{U}}\|^2. \end{aligned}$$

Our result is proved if we can show that $h'(\alpha) \geq 0$ for $\alpha \in (0, 1)$. Now,

$$\begin{aligned} h'(\alpha) &= -(p^{\mathrm{U}})^T(p^{\mathrm{U}} - p^{\mathrm{B}}) + \alpha\|p^{\mathrm{U}} - p^{\mathrm{B}}\|^2 \\ &\geq -(p^{\mathrm{U}})^T(p^{\mathrm{U}} - p^{\mathrm{B}}) \\ &= \frac{g^Tg}{g^TBg}g^T\left(-\frac{g^Tg}{g^TBg}g + B^{-1}g\right) \\ &= g^Tg\frac{gB^{-1}g}{g^TBg}\left[1 - \frac{(g^Tg)^2}{(g^TBg)(g^TB^{-1}g)}\right] \\ &\geq 0, \end{aligned}$$

where the final inequality is a consequence of the Cauchy-Schwarz inequality. (We leave the details as an exercise.)

For (ii), we define $\hat{h}(\alpha) = m(\tilde{p}(1+\alpha))$ and show that $\hat{h}'(\alpha) \leq 0$ for $\alpha \in (0, 1)$. Substitution of (4.16) into (4.5) and differentiation with respect to the argument leads to

$$\begin{aligned} \hat{h}'(\alpha) &= (p^{\mathrm{B}} - p^{\mathrm{U}})^T(g + Bp^{\mathrm{U}}) + \alpha(p^{\mathrm{B}} - p^{\mathrm{U}})^TB(p^{\mathrm{B}} - p^{\mathrm{U}}) \\ &\leq (p^{\mathrm{B}} - p^{\mathrm{U}})^T(g + Bp^{\mathrm{U}} + B(p^{\mathrm{B}} - p^{\mathrm{U}})) \\ &= (p^{\mathrm{B}} - p^{\mathrm{U}})^T(g + Bp^{\mathrm{B}}) = 0, \end{aligned}$$

giving the result. □

It follows from this lemma that the path $\tilde{p}(\tau)$ intersects the trust-region boundary $\|p\| = \Delta$ at exactly one point if $\|p^{\mathrm{B}}\| \geq \Delta$, and nowhere otherwise. Since m is decreasing along the path, the chosen value of p will be at p^{B} if $\|p^{\mathrm{B}}\| \leq \Delta$, otherwise at the point of intersection of the dogleg and the trust-region boundary. In the latter case, we compute the appropriate value of τ by solving the following scalar quadratic equation:

$$\|p^{\mathrm{U}} + (\tau - 1)(p^{\mathrm{B}} - p^{\mathrm{U}})\|^2 = \Delta^2.$$

Consider now the case in which the exact Hessian $\nabla^2 f(x_k)$ is available for use in the model problem (4.5). When $\nabla^2 f(x_k)$ is positive definite, we can simply set $B = \nabla^2 f(x_k)$ (that is, $p^{\text{B}} = (\nabla^2 f(x_k))^{-1} g_k$) and apply the procedure above to find the Newton-dogleg step. Otherwise, we can define p^{B} by choosing B to be one of the positive definite modified Hessians described in Section 3.4, then proceed as above to find the dogleg step. Near a solution satisfying second-order sufficient conditions (see Theorem 2.4), p^{B} will be set to the usual Newton step, allowing the possibility of rapid local convergence of Newton's method (see Section 4.4).

The use of a modified Hessian in the Newton-dogleg method is not completely satisfying from an intuitive viewpoint, however. A modified factorization perturbs the diagonals of $\nabla^2 f(x_k)$ in a somewhat arbitrary manner, and the benefits of the trust-region approach may not be realized. In fact, the modification introduced during the factorization of the Hessian is redundant in some sense because the trust-region strategy introduces its own modification. As we show in Section 4.3, the exact solution of the trust-region problem (4.3) with $B_k = \nabla^2 f(x_k)$ is $(\nabla^2 f(x_k) + \lambda I)^{-1} g_k$, where λ is chosen large enough to make $(\nabla^2 f(x_k) + \lambda I)$ positive definite, and its value depends on the trust-region radius Δ_k. We conclude that the Newton-dogleg method is most appropriate when the objective function is convex (that is, $\nabla^2 f(x_k)$ is always positive semidefinite). The techniques described below may be more suitable for the general case.

The dogleg strategy can be adapted to handle indefinite matrices B, but there is not much point in doing so because the full step p^{B} is not the unconstrained minimizer of m in this case. Instead, we now describe another strategy, which aims to include directions of negative curvature (that is, directions d for which $d^T B d < 0$) in the space of candidate trust-region steps.

TWO-DIMENSIONAL SUBSPACE MINIMIZATION

When B is positive definite, the dogleg method strategy can be made slightly more sophisticated by widening the search for p to the entire two-dimensional subspace spanned by p^{U} and p^{B} (equivalently, g and $-B^{-1}g$). The subproblem (4.5) is replaced by

$$\min_p m(p) = f + g^T p + \tfrac{1}{2} p^T B p \quad \text{s.t. } \|p\| \le \Delta, \; p \in \text{span}[g, B^{-1}g]. \tag{4.17}$$

This is a problem in two variables that is computationally inexpensive to solve. (After some algebraic manipulation it can be reduced to finding the roots of a fourth degree polynomial.) Clearly, the Cauchy point p^{C} is feasible for (4.17), so the optimal solution of this subproblem yields at least as much reduction in m as the Cauchy point, resulting in global convergence of the algorithm. The two-dimensional subspace minimization strategy is obviously an extension of the dogleg method as well, since the entire dogleg path lies in $\text{span}[g, B^{-1}g]$.

This strategy can be modified to handle the case of indefinite B in a way that is intuitive, practical, and theoretically sound. We mention just the salient points of the handling of the indefiniteness here, and refer the reader to papers by Byrd, Schnabel, and Schultz (see [54] and [279]) for details. When B has negative eigenvalues, the two-dimensional subspace in (4.17) is changed to

$$\text{span}[g, (B+\alpha I)^{-1}g], \qquad \text{for some } \alpha \in (-\lambda_1, -2\lambda_1], \tag{4.18}$$

where λ_1 denotes the most negative eigenvalue of B. (This choice of α ensures that $B+\alpha I$ is positive definite, and the flexibility in the choice of α allows us to use a numerical procedure such as the Lanczos method to compute it.) When $\|(B+\alpha I)^{-1}g\| \le \Delta$, we discard the subspace search of (4.17), (4.18) and instead define the step to be

$$p = -(B+\alpha I)^{-1}g + v, \tag{4.19}$$

where v is a vector that satisfies $v^T(B+\alpha I)^{-1}g \le 0$. (This condition ensures that $\|p\| \ge \|(B+\alpha I)^{-1}g\|$.) When B has zero eigenvalues but no negative eigenvalues, we define the step to be the Cauchy point $p = p^{\mathrm{C}}$.

When the exact Hessian is available, we can set $B = \nabla^2 f(x_k)$, and note that $B^{-1}g$ is the Newton step. Hence, when the Hessian is positive definite at the solution x^* and when x_k is close to x^* and Δ is sufficiently large, the subspace minimization problem (4.17) will be solved by the Newton step.

The reduction in model function m achieved by the two-dimensional subspace minimization strategy often is close to the reduction achieved by the exact solution of (4.5). Most of the computational effort lies in a single factorization of B or $B+\alpha I$ (estimation of α and solution of (4.17) are less significant), while strategies that find nearly exact solutions of (4.5) typically require two or three such factorizations (see Section 4.3).

4.2 GLOBAL CONVERGENCE

REDUCTION OBTAINED BY THE CAUCHY POINT

In the preceding discussion of algorithms for approximately solving the trust-region subproblem, we have repeatedly emphasized that global convergence depends on the approximate solution obtaining at least as much decrease in the model function m as the Cauchy point. (In fact, a fixed positive fraction of the Cauchy decrease suffices.) We start the global convergence analysis by obtaining an estimate of the decrease in m achieved by the Cauchy point. We then use this estimate to prove that the sequence of gradients $\{g_k\}$ generated by Algorithm 4.1 has an accumulation point at zero, and in fact converges to zero when η is strictly positive.

Our first main result is that the dogleg and two-dimensional subspace minimization algorithms and Steihaug's algorithm (Algorithm 7.2) produce approximate solutions p_k of the subproblem (4.3) that satisfy the following estimate of decrease in the model function:

$$m_k(0) - m_k(p_k) \ge c_1 \|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right), \tag{4.20}$$

for some constant $c_1 \in (0, 1]$. The usefulness of this estimate will become clear in the following two sections. For now, we note that when Δ_k is the minimum value in (4.20), the condition is slightly reminiscent of the first Wolfe condition: The desired reduction in the model is proportional to the gradient and the size of the step.

We show now that the Cauchy point p_k^{C} satisfies (4.20), with $c_1 = \frac{1}{2}$.

Lemma 4.3.

The Cauchy point p_k^{C} satisfies (4.20) with $c_1 = \frac{1}{2}$, that is,

$$m_k(0) - m_k(p_k^{\mathrm{C}}) \geq \tfrac{1}{2}\|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right). \tag{4.21}$$

PROOF. For simplicity, we drop the iteration index k in the proof.

We consider first the case $g^T Bg \leq 0$. Here, we have

$$\begin{aligned} m(p^{\mathrm{C}}) - m(0) &= m(-\Delta g/\|g\|) - f \\ &= -\frac{\Delta}{\|g\|}\|g\|^2 + \frac{1}{2}\frac{\Delta^2}{\|g\|^2} g^T Bg \\ &\leq -\Delta\|g\| \\ &\leq -\|g\| \min\left(\Delta, \frac{\|g\|}{\|B\|}\right), \end{aligned}$$

and so (4.21) certainly holds.

For the next case, consider $g^T Bg > 0$ and

$$\frac{\|g\|^3}{\Delta g^T Bg} \leq 1. \tag{4.22}$$

From (4.12), we have $\tau = \|g\|^3 / \left(\Delta g^T Bg\right)$, and so from (4.11) it follows that

$$\begin{aligned} m(p^{\mathrm{C}}) - m(0) &= -\frac{\|g\|^4}{g^T Bg} + \frac{1}{2} g^T Bg \frac{\|g\|^4}{(g^T Bg)^2} \\ &= -\frac{1}{2}\frac{\|g\|^4}{g^T Bg} \\ &\leq -\frac{1}{2}\frac{\|g\|^4}{\|B\|\|g\|^2} \\ &= -\frac{1}{2}\frac{\|g\|^2}{\|B\|} \\ &\leq -\frac{1}{2}\|g\| \min\left(\Delta, \frac{\|g\|}{\|B\|}\right), \end{aligned}$$

so (4.21) holds here too.

In the remaining case, (4.22) does not hold, and therefore

$$g^T Bg < \frac{\|g\|^3}{\Delta}. \tag{4.23}$$

From (4.12), we have $\tau = 1$, and using this fact together with (4.23), we obtain

$$\begin{aligned} m(p^{\mathrm{C}}) - m(0) &= -\frac{\Delta}{\|g\|}\|g\|^2 + \frac{1}{2}\frac{\Delta^2}{\|g\|^2} g^T B g \\ &\le -\Delta\|g\| + \frac{1}{2}\frac{\Delta^2}{\|g\|^2}\frac{\|g\|^3}{\Delta} \\ &= -\tfrac{1}{2}\Delta\|g\| \\ &\le -\tfrac{1}{2}\|g\| \min\left(\Delta, \frac{\|g\|}{\|B\|}\right), \end{aligned}$$

yielding the desired result (4.21) once again. □

To satisfy (4.20), our approximate solution p_k has only to achieve a reduction that is at least some fixed fraction c_2 of the reduction achieved by the Cauchy point. We state the observation formally as a theorem.

Theorem 4.4.

Let p_k be any vector such that $\|p_k\| \le \Delta_k$ and $m_k(0) - m_k(p_k) \ge c_2\left(m_k(0) - m_k(p_k^{\mathrm{C}})\right)$. Then p_k satisfies (4.20) with $c_1 = c_2/2$. In particular, if p_k is the exact solution p_k^* of (4.3), then it satisfies (4.20) with $c_1 = \frac{1}{2}$.

Proof. Since $\|p_k\| \le \Delta_k$, we have from Lemma 4.3 that

$$m_k(0) - m_k(p_k) \ge c_2\left(m_k(0) - m_k(p_k^{\mathrm{C}})\right) \ge \tfrac{1}{2}c_2\|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right),$$

giving the result. □

Note that the dogleg and two-dimensional subspace minimization algorithms both satisfy (4.20) with $c_1 = \frac{1}{2}$, because they all produce approximate solutions p_k for which $m_k(p_k) \le m_k(p_k^{\mathrm{C}})$.

Convergence to Stationary Points

Global convergence results for trust-region methods come in two varieties, depending on whether we set the parameter η in Algorithm 4.1 to zero or to some small positive value. When $\eta = 0$ (that is, the step is taken whenever it produces a lower value of f), we can show that the sequence of gradients $\{g_k\}$ has a limit point at zero. For the more stringent acceptance test with $\eta > 0$, which requires the actual decrease in f to be at least some small fraction of the predicted decrease, we have the stronger result that $g_k \to 0$.

In this section we prove the global convergence results for both cases. We assume throughout that the approximate Hessians B_k are uniformly bounded in norm, and that f

is bounded below on the level set

$$S \overset{\text{def}}{=} \{x \mid f(x) \le f(x_0)\}. \tag{4.24}$$

For later reference, we define an open neighborhood of this set by

$$S(R_0) \overset{\text{def}}{=} \{x \mid \|x - y\| < R_0 \text{ for some } y \in S\},$$

where R_0 is a positive constant.

To allow our results to be applied more generally, we also allow the length of the approximate solution p_k of (4.3) to exceed the trust-region bound, provided that it stays within some fixed multiple of the bound; that is,

$$\|p_k\| \le \gamma \Delta_k, \qquad \text{for some constant } \gamma \ge 1. \tag{4.25}$$

The first result deals with the case $\eta = 0$.

Theorem 4.5.

Let $\eta = 0$ in Algorithm 4.1. Suppose that $\|B_k\| \le \beta$ for some constant β, that f is bounded below on the level set S defined by (4.24) and Lipschitz continuously differentiable in the neighborhood $S(R_0)$ for some $R_0 > 0$, and that all approximate solutions of (4.3) satisfy the inequalities (4.20) and (4.25), for some positive constants c_1 and γ. We then have

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{4.26}$$

PROOF. By performing some technical manipulation with the ratio ρ_k from (4.4), we obtain

$$\begin{aligned} |\rho_k - 1| &= \left| \frac{(f(x_k) - f(x_k + p_k)) - (m_k(0) - m_k(p_k))}{m_k(0) - m_k(p_k)} \right| \\ &= \left| \frac{m_k(p_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)} \right|. \end{aligned}$$

Since from Taylor's theorem (Theorem 2.1) we have that

$$f(x_k + p_k) = f(x_k) + g(x_k)^T p_k + \int_0^1 [g(x_k + tp_k) - g(x_k)]^T p_k \, dt,$$

for some $t \in (0, 1)$, it follows from the definition (4.2) of m_k that

$$\begin{aligned} |m_k(p_k) - f(x_k + p_k)| &= \left| \tfrac{1}{2} p_k^T B_k p_k - \int_0^1 [g(x_k + tp_k) - g(x_k)]^T p_k \, dt \right| \\ &\le (\beta/2)\|p_k\|^2 + \beta_1 \|p_k\|^2, \end{aligned} \tag{4.27}$$

where we have used β_1 to denote the Lipschitz constant for g on the set $S(R_0)$, and assumed that $\|p_k\| \le R_0$ to ensure that x_k and $x_k + tp_k$ both lie in the set $S(R_0)$.

Suppose for contradiction that there is $\epsilon > 0$ and a positive index K such that

$$\|g_k\| \ge \epsilon, \qquad \text{for all } k \ge K. \tag{4.28}$$

From (4.20), we have for $k \ge K$ that

$$m_k(0) - m_k(p_k) \ge c_1 \|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right) \ge c_1 \epsilon \min\left(\Delta_k, \frac{\epsilon}{\beta}\right). \tag{4.29}$$

Using (4.29), (4.27), and the bound (4.25), we have

$$|\rho_k - 1| \le \frac{\gamma^2 \Delta_k^2 (\beta/2 + \beta_1)}{c_1 \epsilon \min(\Delta_k, \epsilon/\beta)}. \tag{4.30}$$

We now derive a bound on the right-hand-side that holds for all sufficiently small values of Δ_k, that is, for all $\Delta_k \le \bar{\Delta}$, where $\bar{\Delta}$ is defined as follows:

$$\bar{\Delta} = \min\left(\frac{1}{2} \frac{c_1 \epsilon}{\gamma^2 (\beta/2 + \beta_1)}, \frac{R_0}{\gamma}\right). \tag{4.31}$$

The R_0/γ term in this definition ensures that the bound (4.27) is valid (because $\|p_k\| \le \gamma \Delta_k \le \gamma \bar{\Delta} \le R_0$). Note that since $c_1 \le 1$ and $\gamma \ge 1$, we have $\bar{\Delta} \le \epsilon/\beta$. The latter condition implies that for all $\Delta_k \in [0, \bar{\Delta}]$, we have $\min(\Delta_k, \epsilon/\beta) = \Delta_k$, so from (4.30) and (4.31), we have

$$|\rho_k - 1| \le \frac{\gamma^2 \Delta_k^2 (\beta/2 + \beta_1)}{c_1 \epsilon \Delta_k} = \frac{\gamma^2 \Delta_k (\beta/2 + \beta_1)}{c_1 \epsilon} \le \frac{\gamma^2 \bar{\Delta} (\beta/2 + \beta_1)}{c_1 \epsilon} \le \frac{1}{2}.$$

Therefore, $\rho_k > \frac{1}{4}$, and so by the workings of Algorithm 4.1, we have $\Delta_{k+1} \ge \Delta_k$ whenever Δ_k falls below the threshold $\bar{\Delta}$. It follows that reduction of Δ_k $\left(\text{by a factor of } \frac{1}{4}\right)$ can occur in our algorithm only if

$$\Delta_k \ge \bar{\Delta},$$

and therefore we conclude that

$$\Delta_k \ge \min\left(\Delta_K, \bar{\Delta}/4\right) \qquad \text{for all } k \ge K. \tag{4.32}$$

Suppose now that there is an infinite subsequence $\mathcal{K}$ such that $\rho_k \ge \frac{1}{4}$ for $k \in \mathcal{K}$. For

$k \in \mathcal{K}$ and $k \geq K$, we have from (4.29) that

$$\begin{aligned} f(x_k) - f(x_{k+1}) &= f(x_k) - f(x_k + p_k) \\ &\geq \tfrac{1}{4}\left[m_k(0) - m_k(p_k)\right] \\ &\geq \tfrac{1}{4} c_1 \epsilon \min(\Delta_k, \epsilon/\beta). \end{aligned}$$

Since f is bounded below, it follows from this inequality that

$$\lim_{k \in \mathcal{K},\, k \to \infty} \Delta_k = 0,$$

contradicting (4.32). Hence no such infinite subsequence $\mathcal{K}$ can exist, and we must have $\rho_k < \frac{1}{4}$ for all k sufficiently large. In this case, Δ_k will eventually be multiplied by $\frac{1}{4}$ at every iteration, and we have $\lim_{k\to\infty} \Delta_k = 0$, which again contradicts (4.32). Hence, our original assertion (4.28) must be false, giving (4.26). □

Our second global convergence result, for the case $\eta > 0$, borrows much of the analysis from the proof above. Our approach here follows that of Schultz, Schnabel, and Byrd [279].

Theorem 4.6.

Let $\eta \in \left(0, \frac{1}{4}\right)$ in Algorithm 4.1. Suppose that $\|B_k\| \leq \beta$ for some constant β, that f is bounded below on the level set S (4.24) and Lipschitz continuously differentiable in $S(R_0)$ for some $R_0 > 0$, and that all approximate solutions p_k of (4.3) satisfy the inequalities (4.20) and (4.25) for some positive constants c_1 and γ. We then have

$$\lim_{k \to \infty} g_k = 0. \tag{4.33}$$

PROOF. We consider a particular positive index m with $g_m \neq 0$. Using β_1 again to denote the Lipschitz constant for g on the set $S(R_0)$, we have

$$\|g(x) - g_m\| \leq \beta_1 \|x - x_m\|,$$

for all $x \in S(R_0)$. We now define the scalars ϵ and R to satisfy

$$\epsilon = \tfrac{1}{2}\|g_m\|, \qquad R = \min\left(\frac{\epsilon}{\beta_1}, R_0\right).$$

Note that the ball

$$\mathcal{B}(x_m, R) = \{x \mid \|x - x_m\| \leq R\}$$

is contained in $S(R_0)$, so Lipschitz continuity of g holds inside $\mathcal{B}(x_m, R)$. We have

$$x \in \mathcal{B}(x_m, R) \;\Rightarrow\; \|g(x)\| \geq \|g_m\| - \|g(x) - g_m\| \geq \tfrac{1}{2}\|g_m\| = \epsilon.$$

If the entire sequence $\{x_k\}_{k \geq m}$ stays inside the ball $\mathcal{B}(x_m, R)$, we would have $\|g_k\| \geq \epsilon > 0$

for all $k \geq m$. The reasoning in the proof of Theorem 4.5 can be used to show that this scenario does not occur. Therefore, the sequence $\{x_k\}_{k \geq m}$ eventually leaves $\mathcal{B}(x_m, R)$.

Let the index $l \geq m$ be such that x_{l+1} is the first iterate after x_m outside $\mathcal{B}(x_m, R)$. Since $\|g_k\| \geq \epsilon$ for $k = m, m+1, \ldots, l$, we can use (4.29) to write

$$\begin{aligned} f(x_m) - f(x_{l+1}) &= \sum_{k=m}^{l} f(x_k) - f(x_{k+1}) \\ &\geq \sum\nolimits_{k=m, x_k \neq x_{k+1}}^{l} \eta[m_k(0) - m_k(p_k)] \\ &\geq \sum\nolimits_{k=m, x_k \neq x_{k+1}}^{l} \eta c_1 \epsilon \min\left(\Delta_k, \frac{\epsilon}{\beta}\right), \end{aligned}$$

where we have limited the sum to the iterations k for which $x_k \neq x_{k+1}$, that is, those iterations on which a step was actually taken. If $\Delta_k \leq \epsilon/\beta$ for all $k = m, m+1, \ldots, l$, we have

$$f(x_m) - f(x_{l+1}) \geq \eta c_1 \epsilon \sum_{k=m, x_k \neq x_{k+1}}^{l} \Delta_k \geq \eta c_1 \epsilon R = \eta c_1 \epsilon \min\left(\frac{\epsilon}{\beta_1}, R_0\right). \tag{4.34}$$

Otherwise, we have $\Delta_k > \epsilon/\beta$ for some $k = m, m+1, \ldots, l$, and so

$$f(x_m) - f(x_{l+1}) \geq \eta c_1 \epsilon \frac{\epsilon}{\beta}. \tag{4.35}$$

Since the sequence $\{f(x_k)\}_{k=0}^{\infty}$ is decreasing and bounded below, we have that

$$f(x_k) \downarrow f^* \tag{4.36}$$

for some $f^* > -\infty$. Therefore, using (4.34) and (4.35), we can write

$$\begin{aligned} f(x_m) - f^* &\geq f(x_m) - f(x_{l+1}) \\ &\geq \eta c_1 \epsilon \min\left(\frac{\epsilon}{\beta}, \frac{\epsilon}{\beta_1}, R_0\right) \\ &= \frac{1}{2} \eta c_1 \|g_m\| \min\left(\frac{\|g_m\|}{2\beta}, \frac{\|g_m\|}{2\beta_1}, R_0\right) > 0. \end{aligned}$$

Since $f(x_m) - f^* \downarrow 0$, we must have $g_m \to 0$, giving the result. □

4.3 ITERATIVE SOLUTION OF THE SUBPROBLEM

In this section, we describe a technique that uses the characterization (4.6) of the subproblem solution, applying Newton's method to find the value of λ which matches the given trust-region radius Δ in (4.5). We also prove the key result Theorem 4.1 concerning the characterization of solutions of (4.3).

The methods of Section 4.1 make no serious attempt to find the exact solution of the subproblem (4.5). They do, however, make some use of the information in the model Hessian B_k, and they have advantages of reasonable implementation cost and nice global convergence properties.

When the problem is relatively small (that is, n is not too large), it may be worthwhile to exploit the model more fully by looking for a closer approximation to the solution of the subproblem. In this section, we describe an approach for finding a good approximation at the cost of a few factorizations of the matrix B (typically three factorization), as compared with a single factorization for the dogleg and two-dimensional subspace minimization methods. This approach is based on the characterization of the exact solution given in Theorem 4.1, together with an ingenious application of Newton's method in one variable. Essentially, the algorithm tries to identify the value of λ for which (4.6) is satisfied by the solution of (4.5).

The characterization of Theorem 4.1 suggests an algorithm for finding the solution p of (4.7). Either $\lambda = 0$ satisfies (4.8a) and (4.8c) with $\|p\| \le \Delta$, or else we define

$$p(\lambda) = -(B + \lambda I)^{-1} g$$

for λ sufficiently large that $B + \lambda I$ is positive definite and seek a value $\lambda > 0$ such that

$$\|p(\lambda)\| = \Delta. \tag{4.37}$$

This problem is a one-dimensional root-finding problem in the variable λ.

To see that a value of λ with all the desired properties exists, we appeal to the eigendecomposition of B and use it to study the properties of $\|p(\lambda)\|$. Since B is symmetric, there is an orthogonal matrix Q and a diagonal matrix Λ such that $B = Q \Lambda Q^T$, where

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

and $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ are the eigenvalues of B; see (A.16). Clearly, $B + \lambda I = Q(\Lambda + \lambda I) Q^T$, and for $\lambda \ne \lambda_j$, we have

$$p(\lambda) = -Q(\Lambda + \lambda I)^{-1} Q^T g = -\sum_{j=1}^{n} \frac{q_j^T g}{\lambda_j + \lambda} q_j, \tag{4.38}$$

where q_j denotes the jth column of Q. Therefore, by orthonormality of $q_1, q_2, \ldots, q_n$, we have

$$\|p(\lambda)\|^2 = \sum_{j=1}^{n} \frac{\left(q_j^T g\right)^2}{(\lambda_j + \lambda)^2}. \tag{4.39}$$

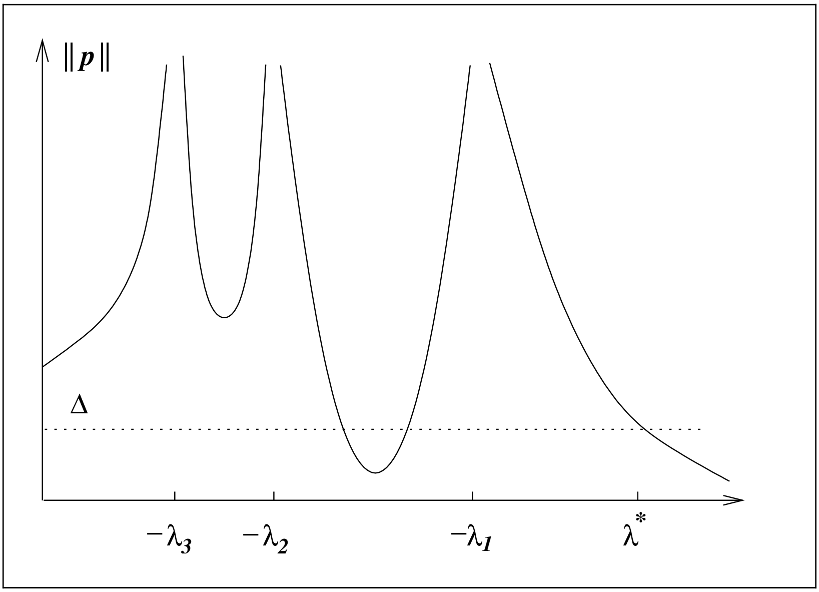

Figure 4.5 $\|p(\lambda)\|$ as a function of λ.

This expression tells us a lot about $\|p(\lambda)\|$. If $\lambda > -\lambda_1$, we have $\lambda_j + \lambda > 0$ for all $j = 1, 2, \ldots, n$, and so $\|p(\lambda)\|$ is a continuous, nonincreasing function of λ on the interval $(-\lambda_1, \infty)$. In fact, we have that

$$\lim_{\lambda \to \infty} \|p(\lambda)\| = 0. \tag{4.40}$$

Moreover, we have when $q_j^T g \neq 0$ that

$$\lim_{\lambda \to -\lambda_j} \|p(\lambda)\| = \infty. \tag{4.41}$$

Figure 4.5 plots $\|p(\lambda)\|$ against λ in a case in whcih $q_1^T g, q_2^T g$, and $q_3^T g$ are all nonzero. Note that the properties (4.40) and (4.41) hold and that $\|p(\lambda)\|$ is a nonincreasing function of λ on $(-\lambda_1, \infty)$. In particular, as is always the case when $q_1^T g \neq 0$, that there is a unique value $\lambda^* \in (-\lambda_1, \infty)$ such that $\|p(\lambda^*)\| = \Delta$. (There may be other, smaller values of λ for which $\|p(\lambda)\| = \Delta$, but these will fail to satisfy (4.8c).)

We now sketch a procedure for identifying the $\lambda^* \in (-\lambda_1, \infty)$ for which $\|p(\lambda^*)\| = \Delta$, which works when $q_1^T g \neq 0$. (We discuss the case of $q_1^T g = 0$ later.) First, note that when B positive definite and $\|B^{-1}g\| \leq \Delta$, the value $\lambda = 0$ satisfies (4.8), so the procedure can be terminated immediately with $\lambda^* = 0$. Otherwise, we could use the root-finding Newton's method (see the Appendix) to find the value of $\lambda > -\lambda_1$ that solves

$$\phi_1(\lambda) = \|p(\lambda)\| - \Delta = 0. \tag{4.42}$$

The disadvantage of this approach can be seen by considering the form of $\|p(\lambda)\|$ when λ is greater than, but close to, $-\lambda_1$. For such λ, we can approximate ϕ_1 by a rational function, as follows:

$$\phi_1(\lambda) \approx \frac{C_1}{\lambda + \lambda_1} + C_2,$$

where $C_1 > 0$ and C_2 are constants. Clearly this approximation (and hence ϕ_1) is highly nonlinear, so the root-finding Newton's method will be unreliable or slow. Better results will be obtained if we reformulate the problem (4.42) so that it is nearly linear near the optimal λ. By defining

$$\phi_2(\lambda) = \frac{1}{\Delta} - \frac{1}{\|p(\lambda)\|},$$

it can be shown using (4.39) that for λ slightly greater than $-\lambda_1$, we have

$$\phi_2(\lambda) \approx \frac{1}{\Delta} - \frac{\lambda + \lambda_1}{C_3}$$

for some $C_3 > 0$. Hence, ϕ_2 is nearly linear near $-\lambda_1$ (see Figure 4.6), and the root-finding

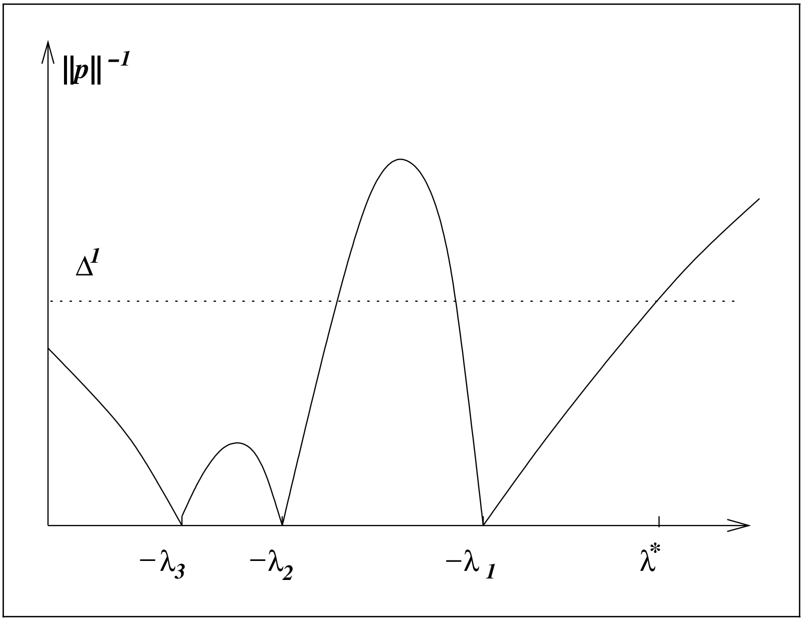

Figure 4.6 $1/\|p(\lambda)\|$ as a function of λ.

Newton's method will perform well, provided that it maintains $\lambda > -\lambda_1$. The root-finding Newton's method applied to ϕ_2 generates a sequence of iterates $\lambda^{(\ell)}$ by setting

$$\lambda^{(\ell+1)} = \lambda^{(\ell)} - \frac{\phi_2\left(\lambda^{(\ell)}\right)}{\phi_2'\left(\lambda^{(\ell)}\right)}. \tag{4.43}$$

After some elementary manipulation, this updating formula can be implemented in the following practical way.

Algorithm 4.3 (Trust Region Subproblem).

```
Given $\lambda^{(0)}, \Delta > 0$:
for $\ell = 0, 1, 2, \ldots$
        Factor $B + \lambda^{(\ell)} I = R^T R$;
        Solve $R^T R p_\ell = -g$, $R^T q_\ell = p_\ell$;

        Set
            $$\lambda^{(\ell+1)} = \lambda^{(\ell)} + \left(\frac{\|p_\ell\|}{\|q_\ell\|}\right)^2 \left(\frac{\|p_\ell\| - \Delta}{\Delta}\right);  \tag{4.44}$$

end (for).
```

Safeguards must be added to this algorithm to make it practical; for instance, when $\lambda^{(\ell)} < -\lambda_1$, the Cholesky factorization $B + \lambda^{(\ell)} I = R^T R$ will not exist. A slightly enhanced version of this algorithm does, however, converge to a solution of (4.37) in most cases.

The main work in each iteration of this method is, of course, the Cholesky factorization of $B + \lambda^{(\ell)} I$. Practical versions of this algorithm do not iterate until convergence to the optimal λ is obtained with high accuracy, but are content with an approximate solution that can be obtained in two or three iterations.

THE HARD CASE

Recall that in the discussion above, we assumed that $q_1^T g \neq 0$. In fact, the approach described above can be applied even when the most negative eigenvalue is a multiple eigenvalue (that is, $0 > \lambda_1 = \lambda_2 = \cdots$), provided that $Q_1^T g \neq 0$, where Q_1 is the matrix whose columns span the subspace corresponding to the eigenvalue λ_1. When this condition does not hold, the situation becomes a little complicated, because the limit (4.41) does not hold for $\lambda_j = \lambda_1$ and so there may not be a value $\lambda \in (-\lambda_1, \infty)$ such that $\|p(\lambda)\| = \Delta$ (see Figure 4.7). Moré and Sorensen [214] refer to this case as the *hard case*. At first glance, it is not clear how p and λ can be chosen to satisfy (4.8) in the hard case. Clearly, our root-finding technique will not work, since there is no solution for λ in the open interval $(-\lambda_1, \infty)$. But Theorem 4.1 assures us that the right value of λ lies in the interval $[-\lambda_1, \infty)$, so there is only

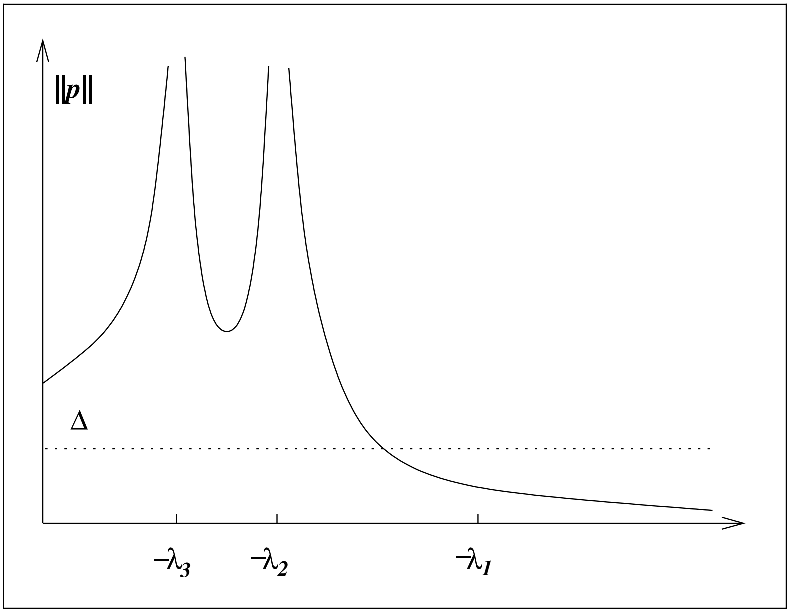

Figure 4.7 The hard case: $\|p(\lambda)\| < \Delta$ for all $\lambda \in (-\lambda_1, \infty)$.

one possibility: $\lambda = -\lambda_1$. To find p, it is not enough to delete the terms for which $\lambda_j = \lambda_1$ from the formula (4.38) and set

$$p = \sum_{j:\lambda_j \neq \lambda_1} \frac{q_j^T g}{\lambda_j + \lambda} q_j.$$

Instead, we note that $(B - \lambda_1 I)$ is singular, so there is a vector z such that $\|z\| = 1$ and $(B - \lambda_1 I)z = 0$. In fact, z is an eigenvector of B corresponding to the eigenvalue λ_1, so by orthogonality of Q we have $q_j^T z = 0$ for $\lambda_j \neq \lambda_1$. It follows from this property that if we set

$$p = \sum_{j:\lambda_j \neq \lambda_1} \frac{q_j^T g}{\lambda_j + \lambda} q_j + \tau z \tag{4.45}$$

for any scalar τ, we have

$$\|p\|^2 = \sum_{j:\lambda_j \neq \lambda_1} \frac{\left(q_j^T g\right)^2}{(\lambda_j + \lambda)^2} + \tau^2,$$

so it is always possible to choose τ to ensure that $\|p\| = \Delta$. It is easy to check that the conditions (4.8) holds for this choice of p and $\lambda = -\lambda_1$.

PROOF OF THEOREM 4.1

We now give a formal proof of Theorem 4.1, the result that characterizes the exact solution of (4.5). The proof relies on the following technical lemma, which deals with the unconstrained minimizers of quadratics and is particularly interesting in the case where the Hessian is positive semidefinite.

Lemma 4.7.

Let m be the quadratic function defined by

$$m(p) = g^T p + \tfrac{1}{2} p^T B p, \tag{4.46}$$

where B is any symmetric matrix. Then the following statements are true.

(i) *m attains a minimum if and only if B is positive semidefinite and g is in the range of B. If B is positive semidefinite, then every p satisfying $Bp = -g$ is a global minimizer of m.*

(ii) *m has a unique minimizer if and only if B is positive definite.*

PROOF. We prove each of the three claims in turn.

(i) We start by proving the "if" part. Since g is in the range of B, there is a p with $Bp = -g$. For all $w \in R^n$, we have

$$\begin{aligned} m(p+w) &= g^T(p+w) + \tfrac{1}{2}(p+w)^T B(p+w) \\ &= (g^T p + \tfrac{1}{2} p^T B p) + g^T w + (Bp)^T w + \tfrac{1}{2} w^T B w \\ &= m(p) + \tfrac{1}{2} w^T B w \\ &\geq m(p), \end{aligned} \tag{4.47}$$

since B is positive semidefinite. Hence, p is a minimizer of m.

For the "only if" part, let p be a minimizer of m. Since $\nabla m(p) = Bp + g = 0$, we have that g is in the range of B. Also, we have $\nabla^2 m(p) = B$ positive semidefinite, giving the result.

(ii) For the "if" part, the same argument as in (i) suffices with the additional point that $w^T B w > 0$ whenever $w \neq 0$. For the "only if" part, we proceed as in (i) to deduce that B is positive semidefinite. If B is not positive definite, there is a vector $w \neq 0$ such that $Bw = 0$. Hence, from (4.47), we have $m(p+w) = m(p)$, so the minimizer is not unique, giving a contradiction. □

To illustrate case (i), suppose that

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

which has eigenvalues 0, 1, 2 and is therefore singular. If g is any vector whose second component is zero, then g will be in the range of B, and the quadratic will attain a minimum. But if the second element in g is nonzero, we can decrease $m(\cdot)$ indefinitely by moving along the direction $(0, -g_2, 0)^T$.

We are now in a position to take account of the trust-region bound $\|p\| \leq \Delta$ and hence prove Theorem 4.1.

PROOF. (Theorem 4.1)

Assume first that there is $\lambda \geq 0$ such that the conditions (4.8) are satisfied. Lemma 4.7(i) implies that p^* is a global minimum of the quadratic function

$$\hat{m}(p) = g^T p + \tfrac{1}{2} p^T (B + \lambda I) p = m(p) + \frac{\lambda}{2} p^T p. \tag{4.48}$$

Since $\hat{m}(p) \geq \hat{m}(p^*)$, we have

$$m(p) \geq m(p^*) + \frac{\lambda}{2}((p^*)^T p^* - p^T p). \tag{4.49}$$

Because $\lambda(\Delta - \|p^*\|) = 0$ and therefore $\lambda(\Delta^2 - (p^*)^T p^*) = 0$, we have

$$m(p) \geq m(p^*) + \frac{\lambda}{2}(\Delta^2 - p^T p).$$

Hence, from $\lambda \geq 0$, we have $m(p) \geq m(p^*)$ for all p with $\|p\| \leq \Delta$. Therefore, p^* is a global minimizer of (4.7).

For the converse, we assume that p^* is a global solution of (4.7) and show that there is a $\lambda \geq 0$ that satisfies (4.8).

In the case $\|p^*\| < \Delta$, p^* is an unconstrained minimizer of m, and so

$$\nabla m(p^*) = Bp^* + g = 0, \qquad \nabla^2 m(p^*) = B \text{ positive semidefinite},$$

and so the properties (4.8) hold for $\lambda = 0$.

Assume for the remainder of the proof that $\|p^*\| = \Delta$. Then (4.8b) is immediately satisfied, and p^* also solves the constrained problem

$$\min m(p) \quad \text{subject to } \|p\| = \Delta.$$

By applying optimality conditions for constrained optimization to this problem (see (12.34)), we find that there is a λ such that the Lagrangian function defined by

$$\mathcal{L}(p, \lambda) = m(p) + \frac{\lambda}{2}(p^T p - \Delta^2)$$

has a stationary point at p^*. By setting $\nabla_p \mathcal{L}(p^*, \lambda)$ to zero, we obtain

$$Bp^* + g + \lambda p^* = 0 \quad \Rightarrow (B + \lambda I)p^* = -g, \tag{4.50}$$

so that (4.8a) holds. Since $m(p) \geq m(p^*)$ for any p with $p^T p = (p^*)^T p^* = \Delta^2$, we have for such vectors p that

$$m(p) \geq m(p^*) + \frac{\lambda}{2}\left((p^*)^T p^* - p^T p\right).$$

If we substitute the expression for g from (4.50) into this expression, we obtain after some rearrangement that

$$\tfrac{1}{2}(p - p^*)^T (B + \lambda I)(p - p^*) \geq 0. \tag{4.51}$$

Since the set of directions

$$\left\{ w \,:\, w = \pm \frac{p - p^*}{\|p - p^*\|}, \text{ for some } p \text{ with } \|p\| = \Delta \right\}$$

is dense on the unit sphere, (4.51) suffices to prove (4.8c).

It remains to show that $\lambda \geq 0$. Because (4.8a) and (4.8c) are satisfied by p^*, we have from Lemma 4.7(i) that p^* minimizes $\hat{m}$, so (4.49) holds. Suppose that there are only negative values of λ that satisfy (4.8a) and (4.8c). Then we have from (4.49) that $m(p) \geq m(p^*)$ whenever $\|p\| \geq \|p^*\| = \Delta$. Since we already know that p^* minimizes m for $\|p\| \leq \Delta$, it follows that m is in fact a global, unconstrained minimizer of m. From Lemma 4.7(i) it follows that $Bp = -g$ and B is positive semidefinite. Therefore conditions (4.8a) and (4.8c) are satisfied by $\lambda = 0$, which contradicts our assumption that only negative values of λ can satisfy the conditions. We conclude that $\lambda \geq 0$, completing the proof. □

CONVERGENCE OF ALGORITHMS BASED ON NEARLY EXACT SOLUTIONS

As we noted in the discussion of Algorithm 4.3, the loop to determine the optimal values of λ and p for the subproblem (4.5) does not iterate until high accuracy is achieved. Instead, it is terminated after two or three iterations with a fairly loose approximation to the true solution. The inexactness in this approximate solution is measured in a different way from the dogleg and subspace minimization algorithms. We can add safeguards to the root-finding Newton method to ensure that the key assumptions of Theorems 4.5 and 4.6 are satisfied by the approximate solution. Specifically, we require that

$$m(0) - m(p) \geq c_1(m(0) - m(p^*)), \tag{4.52a}$$

$$\|p\| \leq \gamma \Delta \tag{4.52b}$$

(where p^* is the exact solution of (4.3)), for some constants $c_1 \in (0, 1]$ and $\gamma > 0$. The condition (4.52a) ensures that the approximate solution achieves a significant fraction of the maximum decrease possible in the model function m. (It is not necessary to know p^*; there are practical termination criteria that imply (4.52a).) One major difference between (4.52) and the earlier criterion (4.20) is that (4.52) makes better use of the second-order part of $m(\cdot)$, that is, the $p^T Bp$ term. This difference is illustrated by the case in which $g = 0$ while B has negative eigenvalues, indicating that the current iterate x_k is a saddle point. Here, the right-hand-side of (4.20) is zero (indeed, the algorithms we described earlier would terminate at such a point). The right-hand-side of (4.52) is positive, indicating that decrease in the model function is still possible, so it forces the algorithm to move away from x_k.

The close attention that near-exact algorithms pay to the second-order term is warranted only if this term closely reflects the actual behavior of the function f—in fact, the trust-region Newton method, for which $B = \nabla^2 f(x)$, is the only case that has been treated in the literature. For purposes of global convergence analysis, the use of the exact Hessian allows us to say more about the limit points of the algorithm than merely that they are stationary points. The following result shows that second-order necessary conditions (Theorem 2.3) are satisfied at the limit points.

Theorem 4.8.

Suppose that the assumptions of Theorem 4.6 are satisfied and in addition that f is twice continuously differentiable in the level set S. Suppose that $B_k = \nabla^2 f(x_k)$ for all k, and that the approximate solution p_k of (4.3) at each iteration satisfies (4.52) for some fixed $\gamma > 0$. Then $\lim_{k\to\infty} \|g_k\| = 0$.

If, in addition, the level set S of (4.24) is compact, then either the algorithm terminates at a point x_k at which the second-order necessary conditions (Theorem 2.3) for a local solution hold, or else $\{x_k\}$ has a limit point x^* in S at which the second-order necessary conditions hold.

We omit the proof, which can be found in Moré and Sorensen [214, Section 4].

4.4 LOCAL CONVERGENCE OF TRUST-REGION NEWTON METHODS

Since global convergence of trust-region methods that use exact Hessians $\nabla^2 f(x)$ is established above, we turn our attention now to local convergence issues. The key to attaining the fast rate of convergence usually associated with Newton's method is to show that the trust-region bound eventually does not interfere as we approach a solution. Specifically, we hope that near the solution, the (approximate) solution of the trust-region subproblem is well inside the trust region and becomes closer and closer to the true Newton step. Steps that satisfy the latter property are said to be *asymptotically similar* to Newton steps.

We first prove a general result that applies to any algorithm of the form of Algorithm 4.1 (see Chapter 4) that generates steps that are asymptotically similar to Newton steps whenever the Newton steps easily satisfy the trust-region bound. It shows that the trust-region constraint eventually becomes inactive in algorithms with this property and that superlinear convergence can be attained. The result assumes that the exact Hessian $B_k = \nabla^2 f(x_k)$ is used in (4.3) when x_k is close to a solution x^* that satisfies second-order sufficient conditions (see Theorem 2.4). Moreover, it assumes that the algorithm uses an approximate solution p_k of (4.3) that achieves a similar decrease in the model function m_k as the Cauchy point.

Theorem 4.9.

Let f be twice Lipschitz continuously differentiable in a neighborhhod of a point x^* at which second-order sufficient conditions (Theorem 2.4) are satisfied. Suppose the sequence $\{x_k\}$ converges to x^* and that for all k sufficiently large, the trust-region algorithm based on (4.3) with $B_k = \nabla^2 f(x_k)$ chooses steps p_k that satisfy the Cauchy-point-based model reduction criterion (4.20) and are asymptotically similar to Newton steps $p_k^{\rm N}$ whenever $\|p_k^{\rm N}\| \le \frac{1}{2}\Delta_k$, that is,

$$\|p_k - p_k^{\rm N}\| = o(\|p_k^{\rm N}\|). \tag{4.53}$$

Then the trust-region bound Δ_k becomes inactive for all k sufficiently large and the sequence $\{x_k\}$ converges superlinearly to x^*.

Proof. We show that $\|p_k^{\rm N}\| \le \frac{1}{2}\Delta_k$ and $\|p_k\| \le \Delta_k$, for all sufficiently large k, so the near-optimal step p_k in (4.53) will eventually always be taken.

We first seek a lower bound on the predicted reduction $m_k(0) - m_k(p_k)$ for all sufficiently large k. We assume that k is large enough that the $o(\|p_k^{\rm N}\|)$ term in (4.53) is less than $\|p_k^{\rm N}\|$. When $\|p_k^{\rm N}\| \le \frac{1}{2}\Delta_k$, we then have that $\|p_k\| \le \|p_k^{\rm N}\| + o(\|p_k^{\rm N}\|) \le 2\|p_k^{\rm N}\|$, while if $\|p_k^{\rm N}\| > \frac{1}{2}\Delta_k$, we have $\|p_k\| \le \Delta_k < 2\|p_k^{\rm N}\|$. In both cases, then, we have

$$\|p_k\| \le 2\|p_k^{\rm N}\| \le 2\left\|\nabla^2 f(x_k)^{-1}\right\| \|g_k\|,$$

and so $\|g_k\| \ge \frac{1}{2}\|p_k\| / \left\|\nabla^2 f(x_k)^{-1}\right\|$.

We have from the relation (4.20) that

$$\begin{aligned}
& m_k(0) - m_k(p_k) \\
& \ge c_1 \|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\left\|\nabla^2 f(x_k)\right\|}\right) \\
& \ge c_1 \frac{\|p_k\|}{2\left\|\nabla^2 f(x_k)^{-1}\right\|} \min\left(\|p_k\|, \frac{\|p_k\|}{2\left\|\nabla^2 f(x_k)\right\| \left\|\nabla^2 f(x_k)^{-1}\right\|}\right) \\
& = c_1 \frac{\|p_k\|^2}{4\left\|\nabla^2 f(x_k)^{-1}\right\|^2 \left\|\nabla^2 f(x_k)\right\|}.
\end{aligned}$$

Because $x_k \to x^*$, we use continuity of $\nabla^2 f(x)$ and positive definiteness of $\nabla^2 f(x^*)$, to deduce that the following bound holds for all k sufficiently large:

$$\frac{c_1}{4\left\|\nabla^2 f(x_k)^{-1}\right\|^2\left\|\nabla^2 f(x_k)\right\|} \geq \frac{c_1}{8\left\|\nabla^2 f(x^*)^{-1}\right\|^2\left\|\nabla^2 f(x^*)\right\|} \overset{\text{def}}{=} c_3,$$

where $c_3 > 0$. Hence, we hae

$$m_k(0) - m_k(p_k) \geq c_3\|p_k\|^2 \tag{4.54}$$

for all sufficiently large k. By Lipschitz continuity of $\nabla^2 f(x)$ near x^*, and using Taylor's theorem (Theorem 2.1), we have

$$\begin{aligned}
&|(f(x_k) - f(x_k + p_k)) - (m_k(0) - m_k(p_k))| \\
&= \left|\tfrac{1}{2}p_k^T\nabla^2 f(x_k)p_k - \tfrac{1}{2}\int_0^1 p_k^T\nabla^2 f(x_k + tp_k)p_k\,dt\right| \\
&\leq \frac{L}{4}\|p_k\|^3,
\end{aligned}$$

where $L > 0$ is the Lipschitz constant for $\nabla^2 f(\cdot)$. Hence, by definition (4.4) of ρ_k, we have for sufficiently large k that

$$|\rho_k - 1| \leq \frac{\|p_k\|^3(L/4)}{c_3\|p_k\|^2} = \frac{L}{4c_3}\|p_k\| \leq \frac{L}{4c_3}\Delta_k. \tag{4.55}$$

Now, the trust-region radius can be reduced only if $\rho_k < \frac{1}{4}$ (or some other fixed number less than 1), so it is clear from (4.55) that the sequence $\{\Delta_k\}$ is bounded away from zero. Since $x_k \to x^*$, we have $\|p_k^{\mathrm{N}}\| \to 0$ and therefore $\|p_k\| \to 0$ from (4.53). Hence, the trust-region bound is inactive for all k sufficiently large, and the bound $\|p_k^{\mathrm{N}}\| \leq \frac{1}{2}\Delta_k$ is eventually always satisfied.

To prove superlinear convergence, we use the quadratic convergence of Newton's method, proved in Theorem 3.5. In particular, we have from (3.33) that

$$\|x_k + p_k^{\mathrm{N}} - x^*\| = o\left(\|x_k - x^*\|^2\right),$$

which implies that $\|p_k^{\mathrm{N}}\| = O(\|x_k - x^*\|)$. Therefore, using (4.53), we have

$$\begin{aligned}
&\|x_k + p_k - x^*\| \\
&\leq \|x_k + p_k^{\mathrm{N}} - x^*\| + \|p_k^{\mathrm{N}} - p_k\| = o\left(\|x_k - x^*\|^2\right) + o(\|p_k^{\mathrm{N}}\|) = o\left(\|x_k - x^*\|\right),
\end{aligned}$$

thus proving superlinear convergence. □

It is immediate from Theorem 3.5 that if $p_k = p_k^{\mathrm{N}}$ for all k sufficiently large, we have quadratic convergence of $\{x_k\}$ to x^*.

Reasonable implementations of the dogleg, subspace minimization, and nearly-exact algorithm of Section 4.3 with $B_k = \nabla^2 f(x_k)$ eventually use the steps $p_k = p_k^{\mathrm{N}}$ under the conditions of Theorem 4.9, and therefore converge quadratically. In the case of the dogleg and two-dimensional subspace minimization methods, the exact step p_k^{N} is one of the candidates for p_k—it lies inside the trust region, along the dogleg path, and inside the two-dimensional subspace. Since under the assumptions of Theorem 4.9, p_k^{N} is the unconstrained minimizer of m_k for k sufficiently large, it is certainly the minimizer in the more restricted domains, so we have $p_k = p_k^{\mathrm{N}}$. For the approach of Section 4.3, if we follow the reasonable strategy of checking whether p_k^{N} is a solution of (4.3) prior to embarking on Algorithm 4.3, then eventually we will also have $p_k = p_k^{\mathrm{N}}$ also.

4.5 OTHER ENHANCEMENTS

SCALING

As we noted in Chapter 2, optimization problems are often posed with poor scaling—the objective function f is highly sensitive to small changes in certain components of the vector x and relatively insensitive to changes in other components. Topologically, a symptom of poor scaling is that the minimizer x^* lies in a narrow valley, so that the contours of the objective $f(\cdot)$ near x^* tend towards highly eccentric ellipses. Algorithms that fail to compensate for poor scaling can perform badly; see Figure 2.7 for an illustration of the poor performance of the steepest descent approach.

Recalling our definition of a trust region—a region around the current iterate within which the model $m_k(\cdot)$ is an adequate representation of the true objective $f(\cdot)$—it is easy to see that a *spherical* trust region may not be appropriate when f is poorly scaled. Even if the model Hessian B_k is exact, the rapid changes in f along certain directions probably will cause m_k to be a poor approximation to f along these directions. On the other hand, m_k may be a more reliable approximation to f along directions in which f is changing more slowly. Since the shape of our trust region should be such that our confidence in the model is more or less the same at all points on the boundary of the region, we are led naturally to consider *elliptical* trust regions in which the axes are short in the sensitive directions and longer in the less sensitive directions.

Elliptical trust regions can be defined by

$$\|Dp\| \le \Delta, \tag{4.56}$$

where D is a diagonal matrix with positive diagonal elements, yielding the following scaled trust-region subproblem:

$$\min_{p \in \mathbb{R}^n} m_k(p) \overset{\text{def}}{=} f_k + g_k^T p + \tfrac{1}{2} p^T B_k p \qquad \text{s.t. } \|Dp\| \le \Delta_k. \tag{4.57}$$

When $f(x)$ is highly sensitive to the value of the ith component x_i, we set the corresponding diagonal element d_{ii} of D to be large, while d_{ii} is smaller for less-sensitive components.

Information to construct the scaling matrix D may be derived from the second derivatives $\partial^2 f/\partial x_i^2$. We can allow D to change from iteration to iteration; most of the theory of this chapter will still apply with minor modifications provided that each d_{ii} stays within some predetermined range $[d_{\text{lo}}, d_{\text{hi}}]$, where $0 < d_{\text{lo}} \le d_{\text{hi}} < \infty$. Of course, we do not need D to be a *precise* reflection of the scaling of the problem, so it is not necessary to devise elaborate heuristics or to perform extensive computations to get it just right.

The following procedure shows how the Cauchy point calculation (Algorithm 4.2) changes when we use a scaled trust region,

Algorithm 4.4 (Generalized Cauchy Point Calculation).
Find the vector p_k^{s} that solves

$$p_k^{\text{s}} = \arg\min_{p\in\mathbb{R}^n} f_k + g_k^T p \qquad \text{s.t. } \|Dp\| \le \Delta_k; \tag{4.58}$$

Calculate the scalar $\tau_k > 0$ that minimizes $m_k(\tau p_k^{\text{s}})$ subject to satisfying the trust-region bound, that is,

$$\begin{aligned} \tau_k &= \arg\min_{\tau>0} m_k(\tau p_k^{\text{s}}) \qquad \text{s.t. } \|\tau D p_k^{\text{s}}\| \le \Delta_k; \\ p_k^{\text{C}} &= \tau_k p_k^{\text{s}}. \end{aligned} \tag{4.59}$$

For this scaled version, we find that

$$p_k^{\text{s}} = -\frac{\Delta_k}{\|D^{-1}g_k\|} D^{-2} g_k, \tag{4.60}$$

and that the step length τ_k is obtained from the following modification of (4.12):

$$\tau_k = \begin{cases} 1 & \text{if } g_k^T D^{-2} B_k D^{-2} g_k \le 0 \\ \min\left(\dfrac{\|D^{-1}g_k\|^3}{\Delta_k g_k^T D^{-2} B_k D^{-2} g_k}, 1\right) & \text{otherwise.} \end{cases} \tag{4.61}$$

(The details are left as an exercise.)

A simpler alternative for adjusting the definition of the Cauchy point and the various algorithms of this chapter to allow for the elliptical trust region is simply to rescale the variables p in the subproblem (4.57) so that the trust region is spherical in the scaled variables. By defining

$$\tilde{p} \overset{\text{def}}{=} Dp,$$

and by substituting into (4.57), we obtain

$$\min_{\tilde{p}\in \mathbb{R}^n} \tilde{m}_k(\tilde{p}) \overset{\text{def}}{=} f_k + g_k^T D^{-1}\tilde{p} + \tfrac{1}{2}\tilde{p}^T D^{-1} B_k D^{-1}\tilde{p} \qquad \text{s.t. } \|\tilde{p}\| \le \Delta_k.$$

The theory and algorithms can now be derived in the usual way by substituting $\tilde{p}$ for p, $D^{-1}g_k$ for g_k, $D^{-1}B_kD^{-1}$ for B_k, and so on.

TRUST REGIONS IN OTHER NORMS

Trust regions may also be defined in terms of norms other than the Euclidean norm. For instance, we may have

$$\|p\|_1 \le \Delta_k \qquad \text{or} \qquad \|p\|_\infty \le \Delta_k,$$

or their scaled counterparts

$$\|Dp\|_1 \le \Delta_k \qquad \text{or} \qquad \|Dp\|_\infty \le \Delta_k,$$

where D is a positive diagonal matrix as before. Norms such as these offer no obvious advantages for small-medium unconstrained problems, but they may be useful for constrained problems. For instance, for the bound-constrained problem

$$\min_{x\in \mathbb{R}^n} f(x), \qquad \text{subject to } x \ge 0,$$

the trust-region subproblem may take the form

$$\min_{p\in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \tfrac{1}{2}p^T B_k p \qquad \text{s.t. } x_k + p \ge 0, \|p\| \le \Delta_k. \tag{4.62}$$

When the trust region is defined by a Euclidean norm, the feasible region for (4.62) consists of the intersection of a sphere and the nonnegative orthant—an awkward object, geometrically speaking. When the ∞-norm is used, however, the feasible region is simply the rectangular box defined by

$$x_k + p \ge 0, \qquad p \ge -\Delta_k e, \qquad p \le \Delta_k e,$$

where $e = (1, 1, \ldots, 1)^T$, so the solution of the subproblem is easily calculated by using techniques for bound-constrained quadratic programming.

For large problems, in which factorization or formation the model Hessian B_k is not computationally desirable, the use of a trust region defined by $\|\cdot\|_\infty$ will also give rise to a bound-constrained subproblem, which may be more convenient to solve than the standard subproblem (4.3). To our knowledge, there has not been much research on the relative performance of methods that use trust regions of different shapes on large problems.

NOTES AND REFERENCES

One of the earliest works on trust-region methods is Winfield [307]. The influential paper of Powell [244] proves a result like Theorem 4.5 for the case of $\eta = 0$, where the algorithm takes a step whenever it decreases the function value. Powell uses a weaker assumption than ours on the matrices $\|B\|$, but his analysis is more complicated. Moré [211] summarizes developments in algorithms and software before 1982, paying particular attention to the importance of using a scaled trust-region norm.

Byrd, Schnabel, and Schultz [279], [54] provide a general theory for inexact trust-region methods; they introduce the idea of two-dimensional subspace minimization and also focus on proper handling of the case of indefinite B to ensure stronger local convergence results than Theorems 4.5 and 4.6. Dennis and Schnabel [93] survey trust-region methods as part of their overview of unconstrained optimization, providing pointers to many important developments in the literature.

The monograph of Conn, Gould, and Toint [74] is an exhaustive treatment of the state of the art in trust-region methods for both unconstrained and constrained optimization. It includes an comprehensive annotated bibliography of the literature in the area.

✐ Exercises

✐ **4.1** Let $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$. At $x = (0, -1)$ draw the contour lines of the quadratic model (4.2) assuming that B is the Hessian of f. Draw the family of solutions of (4.3) as the trust region radius varies from $\Delta = 0$ to $\Delta = 2$. Repeat this at $x = (0, 0.5)$.

✐ **4.2** Write a program that implements the dogleg method. Choose B_k to be the exact Hessian. Apply it to solve Rosenbrock's function (2.22). Experiment with the update rule for the trust region by changing the constants in Algorithm 4.1, or by designing your own rules.

✐ **4.3** Program the trust-region method based on Algorithm 7.2. Choose B_k to be the exact Hessian, and use it to minimize the function

$$\min\ f(x) = \sum_{i=1}^{n} \left[(1 - x_{2i-1})^2 + 10(x_{2i} - x_{2i-1}^2)^2\right]$$

with $n = 10$. Experiment with the starting point and the stopping test for the CG iteration. Repeat the computation with $n = 50$.

Your program should indicate, at every iteration, whether Algorithm 7.2 encountered negative curvature, reached the trust-region boundary, or met the stopping test.

4.4 Theorem 4.5 shows that the sequence $\{\|g\|\}$ has an accumulation point at zero. Show that if the iterates x stay in a bounded set $\mathcal{B}$, then there is a limit point x_∞ of the sequence $\{x_k\}$ such that $g(x_\infty) = 0$.

4.5 Show that τ_k defined by (4.12) does indeed identify the minimizer of m_k along the direction $-g_k$.

4.6 The Cauchy–Schwarz inequality states that for any vectors u and v, we have

$$|u^T v|^2 \le (u^T u)(v^T v),$$

with equality only when u and v are parallel. When B is positive definite, use this inequality to show that

$$\gamma \stackrel{\text{def}}{=} \frac{\|g\|^4}{(g^T Bg)(g^T B^{-1} g)} \le 1,$$

with equality only if g and Bg (and $B^{-1}g$) are parallel.

4.7 When B is positive definite, the *double-dogleg method* constructs a path with three line segments from the origin to the full step. The four points that define the path are

- the origin;
- the unconstrained Cauchy step $p^{\mathrm{C}} = -(g^T g)/(g^T Bg)g$;
- a fraction of the full step $\bar{\gamma} p^{\mathrm{B}} = -\bar{\gamma} B^{-1} g$, for some $\bar{\gamma} \in (\gamma, 1]$, where γ is defined in the previous question; and
- the full step $p^{\mathrm{B}} = -B^{-1} g$.

Show that $\|p\|$ increases monotonically along this path.

(Note: The double-dogleg method, as discussed in Dennis and Schnabel [92, Section 6.4.2], was for some time thought to be superior to the standard dogleg method, but later testing has not shown much difference in performance.)

4.8 Show that (4.43) and (4.44) are equivalent. Hints: Note that

$$\begin{aligned} \frac{d}{d\lambda}\left(\frac{1}{\|p(\lambda)\|}\right) &= \frac{d}{d\lambda}\left(\|p(\lambda)\|^2\right)^{-1/2} = -\frac{1}{2}\left(\|p(\lambda)\|^2\right)^{-3/2} \frac{d}{d\lambda}\|p(\lambda)\|^2, \\ \frac{d}{d\lambda}\|p(\lambda)\|^2 &= -2\sum_{j=1}^{n} \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^3} \end{aligned}$$

(from (4.39)), and

$$\|q\|^2 = \|R^{-T} p\|^2 = p^T (B + \lambda I)^{-1} p = \sum_{j=1}^{n} \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^3}.$$

✐ **4.9** Derive the solution of the two-dimensional subspace minimization problem in the case where B is positive definite.

✐ **4.10** Show that if B is any symmetric matrix, then there exists $\lambda \geq 0$ such that $B + \lambda I$ is positive definite.

✐ **4.11** Verify that the definitions (4.60) for p_k^s and (4.61) for τ_k are valid for the Cauchy point in the case of an elliptical trust region. (Hint: Using the theory of Chapter 12, we can show that the solution of (4.58) satisfies $g_k + \alpha D^2 p_k^s = 0$ for some scalar $\alpha \geq 0$.)

✐ **4.12** The following example shows that the reduction in the model function m achieved by the two-dimensional minimization strategy can be much smaller than that achieved by the exact solution of (4.5).

In (4.5), set

$$g = \left(-\frac{1}{\epsilon}, -1, -\epsilon^2 \right)^T,$$

where ϵ is a small positive number. Set

$$B = \operatorname{diag}\left(\frac{1}{\epsilon^3}, 1, \epsilon^3 \right), \qquad \Delta = 0.5.$$

Show that the solution of (4.5) has components $\left(O(\epsilon), \frac{1}{2} + O(\epsilon), O(\epsilon)\right)^T$ and that the reduction in the model m is $\frac{3}{8} + O(\epsilon)$. For the two-dimensional minimization strategy, show that the solution is a multiple of $B^{-1} g$ and that the reduction in m is $O(\epsilon)$.