

Notes for Part 3: Trust-region methods for unconstrained optimization

Nick Gould, CSED, RAL, Chilton, OX11 0QX, England (n.gould@rl.ac.uk)

January 4, 2006

3 Sketches of proofs for Part 3

3.1 Proof of Theorem 3.1

Firstly note that, for all $\alpha \geq 0$,

$$m_k(-\alpha g_k) = f_k - \alpha \|g_k\|_2^2 + \frac{1}{2} \alpha^2 g_k^T B_k g_k. \quad (3.1)$$

If g_k is zero, the result is immediate. So suppose otherwise. In this case, there are three possibilities:

- (i) the curvature $g_k^T B_k g_k$ is not strictly positive; in this case $m_k(-\alpha g_k)$ is unbounded from below as α increases, and hence the Cauchy point occurs on the trust-region boundary.
- (ii) the curvature $g_k^T B_k g_k > 0$ and the minimizer of $m_k(-\alpha g_k)$ occurs at or beyond the trust-region boundary; once again, the Cauchy point occurs on the trust-region boundary.
- (iii) the curvature $g_k^T B_k g_k > 0$ and the minimizer of $m_k(-\alpha g_k)$, and hence the Cauchy point, occurs before the trust-region is reached.

We consider each case in turn;

Case (i). In this case, since $g_k^T B_k g_k \leq 0$, (3.1) gives

$$m_k(-\alpha g_k) = f_k - \alpha \|g_k\|_2^2 + \frac{1}{2} \alpha^2 g_k^T B_k g_k \leq f_k - \alpha \|g_k\|_2^2 \quad (3.2)$$

for all $\alpha \geq 0$. Since the Cauchy point lies on the boundary of the trust region

$$\alpha_k^C = \frac{\Delta_k}{\|g_k\|}. \quad (3.3)$$

Substituting this value into (3.2) gives

$$f_k - m_k(s_k^C) \geq \|g_k\|_2^2 \frac{\Delta_k}{\|g_k\|} \geq \kappa_s \|g_k\|_2 \Delta_k \geq \frac{1}{2} \kappa_s \|g_k\|_2 \Delta_k \quad (3.4)$$

since $\|g_k\|_2 \geq \kappa_s \|g_k\|$.

Case (ii). In this case, let α_k^* be the unique minimizer of (3.1); elementary calculus reveals that

$$\alpha_k^* = \frac{\|g_k\|_2^2}{g_k^T B_k g_k}. \quad (3.5)$$

Since this minimizer lies on or beyond the trust-region boundary (3.3) and (3.5) together imply that

$$\alpha_k^C g_k^T B_k g_k \leq \|g_k\|_2^2.$$

Substituting this last inequality in (3.1), and using (3.3) and $\|g_k\|_2 \geq \kappa_s \|g_k\|$, it follows that

$$f_k - m_k(s_k^C) = \alpha_k^C \|g_k\|_2^2 - \frac{1}{2} \alpha_k^C{}^2 g_k^T B_k g_k \geq \frac{1}{2} \alpha_k^C \|g_k\|_2^2 = \frac{1}{2} \|g_k\|_2^2 \frac{\Delta_k}{\|g_k\|} \geq \frac{1}{2} \kappa_s \|g_k\|_2 \Delta_k.$$

Case (iii). In this case, $\alpha_k^C = \alpha_k^*$, and (3.1) becomes

$$f_k - m_k(s_k^C) = \frac{\|g_k\|_2^4}{g_k^T B_k g_k} - \frac{1}{2} \frac{\|g_k\|_2^4}{g_k^T B_k g_k} = \frac{1}{2} \frac{\|g_k\|_2^4}{g_k^T B_k g_k} \geq \frac{1}{2} \frac{\|g_k\|_2^4}{\|B_k\|_2},$$

where

$$|g_k^T B_k g_k| \leq \|g_k\|_2^2 \|B_k\|_2 \leq \|g_k\|_2^2 (1 + \|B_k\|_2)$$

because of the Cauchy-Schwarz inequality.

The result follows since it is true in each of the above three possible cases. Note that the “1+” is only needed to cover case where $B_k = 0$, and that in this case, the “min” in the theorem might actually be replaced by $\kappa_s \Delta_k$.

3.2 Proof of Corollary 3.2

Immediate from Theorem 3.1 and the requirement that $m_k(s_k) \leq m_k(s_k^C)$

3.3 Proof of Lemma 3.3

The mean value theorem gives that

$$f(x_k + s_k) - f(x_k) = s_k^T \nabla_x f(x_k) + \frac{1}{2} s_k^T \nabla_{xx} f(\xi_k) s_k$$

for some ξ_k in the segment $[x_k, x_k + s_k]$. Thus

$$\begin{aligned} |f(x_k + s_k) - m_k(s_k)| &= \left| \frac{1}{2} s_k^T H(\xi_k) s_k - s_k^T B_k s_k \right| \leq \frac{1}{2} \|s_k\|_2^2 \|H(\xi_k) - B_k\|_2 + \frac{1}{2} \|s_k\|_2^2 \|B_k\|_2 \\ &\leq \frac{1}{2} (\kappa_h + \kappa_b) \|s_k\|_2^2 \leq \frac{1}{2} \kappa_f^2 (\kappa_h + \kappa_b) \|s_k\|_2^2 \leq \kappa_d \Delta_k^2 \end{aligned}$$

using the triangle and Cauchy-Schwarz inequalities.

3.4 Proof of Lemma 3.4

By definition,

$$1 + \|B_k\|_2 \leq \kappa_h + \kappa_b,$$

and hence for any radius satisfying the given (first) bound,

$$\kappa_s \Delta_k \leq \frac{\|g_k\|_2}{\kappa_h + \kappa_b} \leq \frac{\|g_k\|_2}{1 + \|B_k\|_2}.$$

As a consequence, Corollary 3.2 gives that

$$f_k - m_k(s_k) \geq \frac{1}{2} \|g_k\|_2 \min \left[\frac{\|g_k\|_2}{1 + \|B_k\|_2}, \kappa_s \Delta_k \right] = \frac{1}{2} \kappa_s \|g_k\|_2 \Delta_k. \quad (3.6)$$

But then Lemma 3.3 and the assumed (second) bound on the radius gives that

$$|\rho_k - 1| = \left| \frac{f(x_k + s_k) - m_k(s_k)}{f_k - m_k(s_k)} \right| \leq \frac{\kappa_d \Delta_k^2}{\kappa_s \|g_k\|_2 \Delta_k} = \frac{2\kappa_d \Delta_k}{\kappa_s \|g_k\|_2} \leq 1 - \eta_c. \quad (3.7)$$

Therefore, $\rho_k \geq \eta_c$ and the iteration is very successful.

3.5 Proof of Lemma 3.5

Suppose otherwise that Δ_k can become arbitrarily small. In particular, assume that iteration k is the first such that

$$\Delta_{k+1} \leq \kappa_\epsilon. \quad (3.8)$$

Then since the radius for the previous iteration must have been larger, the iteration was unsuccessful, and thus $\gamma_d \Delta_k \leq \Delta_{k+1}$. Hence

$$\Delta_k \leq \epsilon \min \left(\frac{1}{\kappa_s(\kappa_h + \kappa_b)}, \frac{\kappa_s(1 - \eta_b)}{2\kappa_d} \right) \leq \|g_k\|_2 \min \left(\frac{1}{\kappa_s(\kappa_h + \kappa_b)}, \frac{\kappa_s(1 - \eta_b)}{2\kappa_d} \right)$$

But this contradicts the assertion of Lemma 3.4 that the k -th iteration must be very successful.

3.6 Proof of Lemma 3.6

The mechanism of the algorithm ensures that $x_* = x_{k_0+1} = x_{k_0+j}$ for all $j > 0$, where k_0 is the index of the last successful iterate. Moreover, since all iterations are unsuccessful for sufficiently large k , the sequence $\{\Delta_k\}$ converges to zero. If $\|g_{k_0+1}\| > 0$, Lemma 3.4 then implies that there must be a successful iteration of index larger than k_0 , which is impossible. Hence $\|g_{k_0+1}\| = 0$.

3.7 Proof of Theorem 3.7

Lemma 3.6 shows that the result is true when there are only a finite number of successful iterations. So it remains to consider the case where there are an infinite number of successful iterations. Let \mathcal{S} be the index set of successful iterations. Now suppose that

$$\|g_k\|_2 \geq \epsilon \quad (3.9)$$

for some $\epsilon > 0$ and all k , and consider a successful iteration of index k . The fact that k is successful, Corollary 3.2, Lemma 3.5, and the assumption (3.9) give that

$$f_k - f_{k+1} \geq \eta_s [f_k - m_k(s_k)] \geq \delta_\epsilon \stackrel{\text{def}}{=} \frac{1}{2} \eta_s \epsilon \min \left[\frac{\epsilon}{1 + \kappa_b}, \kappa_s \kappa_\epsilon \right]. \quad (3.10)$$

Summing now over all successful iterations from 0 to k , it follows that

$$f_0 - f_{k+1} = \sum_{\substack{j=0 \\ j \in \mathcal{S}}}^k [f_j - f_{j+1}] \geq \sigma_k \delta_\epsilon,$$

where σ_k is the number of successful iterations up to iteration k . But since there are infinitely many such iterations, it must be that

$$\lim_{k \rightarrow \infty} \sigma_k = +\infty.$$

Thus (3.9) can only be true if f_{k+1} is unbounded from below, and conversely, if f_{k+1} is bounded from below, (3.9) must be false, and there is a subsequence of the $\|g_k\|_2$ converging to zero.

3.8 Proof of Theorem 3.8

Suppose otherwise that f_n is bounded from below, and that there is a subsequence of successful iterates, indexed by $\{t_i\} \subseteq \mathcal{S}$, such that

$$\|g_{t_i}\|_2 \geq 2\epsilon > 0 \quad (3.11)$$

for some $\epsilon > 0$ and for all i . Theorem 3.7 ensures the existence, for each t_i , of a first successful iteration $\ell_i > t_i$ such that $\|g_{\ell_i}\|_2 < \epsilon$. That is to say that there is another subsequence of \mathcal{S} indexed by $\{\ell_i\}$ such that

$$\|g_{\ell_i}\|_2 \geq \epsilon \text{ for } t_i \leq \ell_i < \ell_{i+1} \text{ and } \|g_{\ell_i}\|_2 < \epsilon. \quad (3.12)$$

We now restrict our attention to the subsequence of successful iterations whose indices are in the set

$$\mathcal{K} \stackrel{\text{def}}{=} \{k \in \mathcal{S} \mid t_i \leq k < \ell_i\},$$

where t_i and ℓ_i belong to the two subsequences defined above.

The subsequences $\{t_i\}$, $\{\ell_i\}$ and \mathcal{K} are all illustrated in Figure 3.1, where, for simplicity, it is assumed that all iterations are successful. In this figure, we have marked position j in each of the subsequences represented in abscissa when j belongs to that subsequence. Note in this example that $\ell_0 = t_1 = \ell_2 = \ell_3 = \ell_4 = \ell_5 = 8$, which we indicated by arrows from $t_0 = 0$, $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$ and $t_5 = 7$ to $k = 9$, and so on.

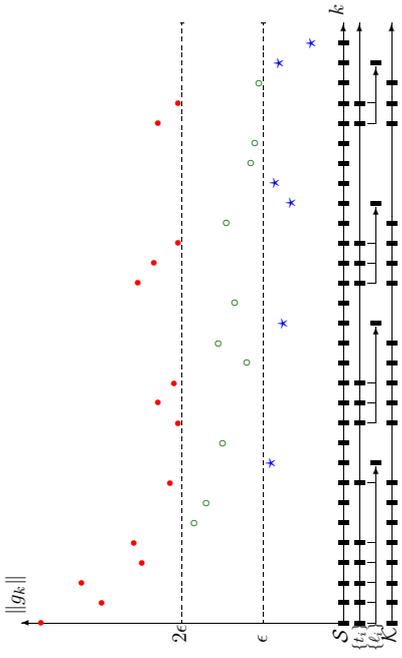


Figure 3.1: The subsequences of the proof of Theorem 3.8

As in the previous proof, it immediately follows that

$$f_k - f_{k+1} \geq \eta_s [f_k - m_k(s_k)] \geq \frac{1}{2} \eta_s \epsilon \min \left[\frac{\epsilon}{1 + \kappa_b}, \kappa_s \Delta_k \right] \quad (3.13)$$

holds for all $k \in \mathcal{K}$ because of (3.12). Hence, since $\{f_k\}$ is, by assumption, bounded from below, the left-hand side of (3.13) must tend to zero when k tends to infinity, and thus that

$$\lim_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} \Delta_k = 0.$$

As a consequence, the second term dominates in the minimum of (3.13) and it follows that, for $k \in K$ sufficiently large,

$$\Delta_k \leq \frac{2}{\epsilon \eta_{s^* K_s}} [f_k - f_{k+1}].$$

We then deduce from this bound that, for i sufficiently large,

$$\|x_{\ell_i} - x_{\ell_i}\| \leq \sum_{\substack{j=\ell_i \\ j \in K}}^{\ell_i-1} \|x_j - x_{j+1}\| \leq \sum_{\substack{j=\ell_i \\ j \in K}}^{\ell_i-1} \Delta_j \leq \frac{2}{\epsilon \eta_{s^* K_s}} [f_{\ell_i} - f_{\ell_i}]. \quad (3.14)$$

But, because $\{f_k\}$ is monotonic and, by assumption, bounded from below, the right-hand side of (3.14) must converge to zero. Thus $\|x_{\ell_i} - x_{\ell_i}\|$ tends to zero as i tends to infinity, and hence, by continuity, $\|g_{\ell_i} - g_{\ell_i}\|$ also tend to zero. However this is impossible because of the definitions of $\{\ell_i\}$ and $\{\ell_i\}$, which imply that $\|g_{\ell_i} - g_{\ell_i}\| \geq \epsilon$. Hence, no subsequence satisfying (3.11) can exist.

3.9 Proof of Theorem 3.9

The constraint $\|s\|_2 \leq \Delta$ is equivalent to

$$\frac{1}{2} \Delta^2 - \frac{1}{2} s^T s \geq 0. \quad (3.15)$$

Applying Theorem 1.9 to the problem of minimizing $q(s)$ subject to (3.15) gives

$$g + Bs_* = -\lambda_* s_* \quad (3.16)$$

for some Lagrange multiplier $\lambda_* \geq 0$ for which either $\lambda_* = 0$ or $\|s_*\|_2 = \Delta$ (or both). It remains to show that $B + \lambda_* I$ is positive semi-definite.

If s_* lies in the interior of the trust-region, necessarily $\lambda_* = 0$, and Theorem 1.10 implies that $B + \lambda_* I = B$ must be positive semi-definite. Likewise if $\|s_*\|_2 = \Delta$ and $\lambda_* = 0$, it follows from Theorem 1.10 that necessarily $v^T B v \geq 0$ for all $v \in \mathcal{N}_+ = \{v | s_*^T v \geq 0\}$. If $v \notin \mathcal{N}_+$, then $-v \in \mathcal{N}_+$, and thus $v^T B v \geq 0$ for all v . Thus the only outstanding case is where $\|s_*\|_2 = \Delta$ and $\lambda_* > 0$. In this case, Theorem 1.10 shows that $v^T (B + \lambda_* I) v \geq 0$ for all $v \in \mathcal{N}_+ = \{v | s_*^T v = 0\}$, so it remains to consider $v^T B v$ when $s_*^T v \neq 0$.

Let s be any point on the boundary of the trust-region, and let $w = s - s_*$. Then

$$-w^T s_* = (s_* - s)^T s_* = \frac{1}{2} (s_* - s)^T (s_* - s) = \frac{1}{2} w^T w \quad (3.17)$$

since $\|s\|_2 = \Delta = \|s_*\|_2$. Combining this with (3.16) gives

$$q(s) - q(s_*) = w^T (g + Bs_*) + \frac{1}{2} w^T B w = -\lambda_* w^T s_* + \frac{1}{2} w^T B w = \frac{1}{2} w^T (B + \lambda_* I) w, \quad (3.18)$$

and thus necessarily $w^T (B + \lambda_* I) w \geq 0$ since s_* is a global minimizer. It is easy to show that

$$s = s_* - 2 \frac{s_*^T v}{v^T v} v$$

lies on the trust-region boundary, and thus for this s , w is parallel to v from which it follows that $v^T (B + \lambda_* I) v \geq 0$.

When $B + \lambda_* I$ is positive definite, $s_* = -(B + \lambda_* I)^{-1} g$. If this point is on the trust-region boundary, while s is any value in the trust-region, (3.17) and (3.18) become $-w^T s_* \geq \frac{1}{2} w^T w$ and $q(s) \geq q(s_*) + \frac{1}{2} w^T (B + \lambda_* I) w$ respectively. Hence, $q(s) > q(s_*)$ for any $s \neq s_*$. If s_* is interior, $\lambda_* = 0$, B is positive definite, and thus s_* is the unique unconstrained minimizer of $q(s)$.

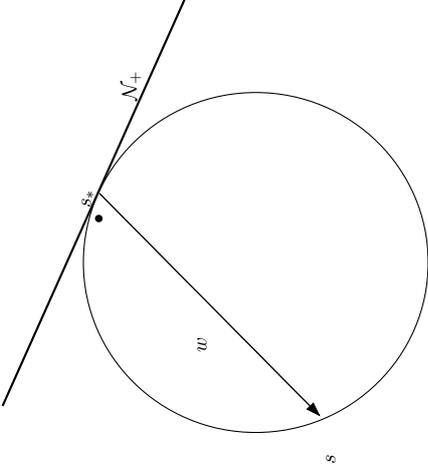


Figure 3.2: Construction of “missing” directions of positive curvature.

3.10 Newton’s method for the secular equation

Recall that the Newton correction at λ is $-\phi(\lambda)/\phi'(\lambda)$. Since

$$\phi(\lambda) = \frac{1}{\|s(\lambda)\|_2} - \frac{1}{\Delta} = \frac{1}{(s^T(\lambda)s(\lambda))^{\frac{1}{2}}} - \frac{1}{\Delta},$$

it follows, on differentiating, that

$$\phi'(\lambda) = -\frac{s^T(\lambda)\nabla_{\lambda}s(\lambda)}{(s^T(\lambda)s(\lambda))^{\frac{3}{2}}} - \frac{s^T(\lambda)\nabla_{\lambda}s(\lambda)}{\|s(\lambda)\|_2^3}.$$

In addition, on differentiating the defining equation

$$(B + \lambda I)s(\lambda) = -g,$$

it must be that

$$(B + \lambda I)\nabla_{\lambda}s(\lambda) + s(\lambda) = 0.$$

Notice that, rather than the value of $\nabla_{\lambda}s(\lambda)$, merely the numerator

$$s^T(\lambda)\nabla_{\lambda}s(\lambda) = -s^T(\lambda)(B + \lambda I)(\lambda)^{-1}s(\lambda)$$

is required in the expression for $\phi'(\lambda)$. Given the factorization $B + \lambda I = L(\lambda)L^T(\lambda)$, the simple relationship

$$s^T(\lambda)(B + \lambda I)^{-1}s(\lambda) = s^T(\lambda)L^{-T}(\lambda)L^{-1}(\lambda)s(\lambda) = (L^{-1}(\lambda)s(\lambda))^T(L^{-1}(\lambda)s(\lambda)) = \|w(\lambda)\|_2^2$$

where $L(\lambda)w(\lambda) = s(\lambda)$ then justifies the Newton step.

3.11 Proof of Theorem 3.10

We first show that

$$d^i T d^j = \frac{\|g^i\|_2^2 \|d^j\|_2^2}{\|g^j\|_2^2} > 0 \quad (3.19)$$

for all $0 \leq j \leq i \leq k$. For any i , (3.19) is trivially true for $j = i$. Suppose it is also true for all $i \leq l$. Then, the update for d^{l+1} gives

$$d^{l+1} = -g^{l+1} + \frac{\|g^{l+1}\|_2^2}{\|g^l\|_2^2} d^l.$$

Forming the inner product with d^j , and using the fact that $d^j T g^{l+1} = 0$ for all $j = 0, \dots, l$, and (3.19) when $j = l$, reveals

$$d^{l+1 T} d^j = -g^{l+1 T} d^j + \frac{\|g^{l+1}\|_2^2 \|g^l\|_2^2}{\|g^l\|_2^2 \|g^l\|_2^2} \|d^j\|_2^2 = \frac{\|g^{l+1}\|_2^2}{\|g^l\|_2^2} \|d^j\|_2^2 > 0.$$

Thus (3.19) is true for $i \leq l + 1$, and hence for all $0 \leq j \leq i \leq k$.

We now have from the algorithm that

$$s^i = s^0 + \sum_{j=0}^{i-1} \alpha^j d^j = \sum_{j=0}^{i-1} \alpha^j d^j$$

as, by assumption, $s^0 = 0$. Hence

$$s^i T d^i = \sum_{j=0}^{i-1} \alpha^j d^j T d^i = \sum_{j=0}^{i-1} \alpha^j d^j T d^i > 0 \quad (3.20)$$

as each $\alpha^j > 0$, which follows from the definition of α^j , since $d^j T H d^j > 0$, and from relationship (3.19). Hence

$$\begin{aligned} \|s^{i+1}\|_2^2 &= s^{i+1 T} s^{i+1} = (s^i + \alpha^i d^i)^T (s^i + \alpha^i d^i) \\ &= s^{i T} s^i + 2\alpha^i s^{i T} d^i + \alpha^{2i} d^{i T} d^i > s^{i T} s^i = \|s^i\|_2^2 \end{aligned}$$

follows directly from (3.20) and $\alpha^i > 0$ which is the required result.

3.12 Proof of Theorem 3.11

The proof is elementary but rather complicated. See

Y. Yuan, "On the truncated conjugate-gradient method", *Math. Programming*, **87** (2000) 561:573

for full details.