Notes for Part 2: Linesearch methods for unconstrained optimization

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

January 4, 2006

2 Sketches of proofs for Part 2

2.1 Proof of Theorem 2.1

From Taylor's theorem (Theorem 1.1), and using the bound

$$\alpha \leq \frac{2(\beta-1)g(x)^T p}{\gamma(x)\|p\|_2^2},$$

we have that

$$\begin{aligned} f(x+\alpha p) &\leq f(x) + \alpha g(x)^T p + \tfrac{1}{2}\gamma(x)\alpha^2\|p\|^2 \\ &\leq f(x) + \alpha g(x)^T p + \alpha(\beta-1)g(x)^T p \\ &= f(x) + \alpha\beta g(x)^T p \end{aligned}$$

2.2 Proof of Corollary 2.2

Theorem 2.1 shows that the linesearch will terminate as soon as $\alpha^{(l)} \leq \alpha_{\max}$. There are two cases to consider. Firstly, it may be that α_{init} satisfies the Armijo condition, in which case $\alpha_k = \alpha_{\text{init}}$. If not, there must be a last linesearch iteration, say the lth, for which $\alpha^{(l)} > \alpha_{\max}$ (if the linesearch has not already terminated). Then $\alpha_k \geq \alpha^{(l+1)} = \tau\alpha^{(l)} > \tau\alpha_{\max}$. Combining these two cases gives the required result.

2.3 Proof of Theorem 2.3

We shall suppose that $g_k \neq 0$ for all k and that

$$\lim_{k\to\infty} f_k > -\infty$$

From the Armijo condition, we have that

$$f_{k+1} - f_k \leq \alpha_k \beta p_k^T g_k$$

for all k, and hence summing over the first j iterations

$$f_{j+1} - f_0 \leq \sum_{k=0}^{j} \alpha_k \beta p_k^T g_k.$$

Since the left-hand side of this inequality is, by assumption, bounded below, so is the sum on right-hand-side As this sum is composed of negative terms, we deduce that

$$\lim_{k\to\infty} \alpha_k |p_k^T g_k| = 0.$$

Now define the two sets

$$\mathcal{K}_1 = \left\{ k \mid \alpha_{\text{init}} > \frac{2\tau(\beta-1)g_k^T p_k}{\gamma\|p_k\|_2^2} \right\}$$

and

$$\mathcal{K}_2 = \left\{ k \mid \alpha_{\text{init}} \leq \frac{2\tau(\beta-1)g_k^T p_k}{\gamma\|p_k\|_2^2} \right\},$$

where γ is the assumed uniform Lipschitz constant. For $k \in \mathcal{K}_1$,

$$\alpha_k \geq \frac{2\tau(\beta-1)g_k^T p_k}{\gamma\|p_k\|_2^2}$$

in which case

$$\alpha_k p_k^T g_k \leq \frac{2\tau(\beta-1)}{\gamma}\left(\frac{g_k^T p_k}{\|p_k\|}\right)^2 < 0.$$

Thus

$$\lim_{k\in\mathcal{K}_1\to\infty} \frac{|p_k^T g_k|}{\|p_k\|_2} = 0. \tag{2.1}$$

For $k \in \mathcal{K}_2$,

$$\alpha_k \geq \alpha_{\text{init}}$$

in which case

$$\lim_{k\in\mathcal{K}_2\to\infty} |p_k^T g_k| = 0. \tag{2.2}$$

Combining (2.1) and (2.2) gives the required result.

2.4 Proof of Theorem 2.4

Follows immediately from Theorem 2.3, since

$$\min\left(|p_k^T g_k|, |p_k^T g_k|/\|p_k\|_2\right) = \|g_k\|_2 \min(1, \|g_k\|_2)$$

and thus

$$\lim_{k\to\infty} \min\left(|p_k^T g_k|, |p_k^T g_k|/\|p_k\|_2\right) = 0$$

implies that $\lim_{k\to\infty} g_k = 0$.

2.5 Proof of Theorem 2.5

Let $\lambda_{\min}(B_k)$ and $\lambda_{\max}(B_k)$ be the smallest and largest eigenvalues of B_k. By assumption, there are bounds $\lambda_{\min} > 0$ and $\lambda_{\max}$ such that

$$\lambda_{\min} \leq \lambda_{\min}(B_k) \leq \frac{s^T B_k s}{\|s\|^2} \leq \lambda_{\max}(B_k) \leq \lambda_{\max}$$

and thus that

$$\lambda_{\max}^{-1} \le \lambda_{\max}^{-1}(B_k) = \lambda_{\min}(B_k^{-1}) \le \frac{s^T B_k^{-1} s}{\|s\|^2} \le \lambda_{\max}(B_k^{-1}) = \lambda_{\min}^{-1}(B_k) \le \lambda_{\min}^{-1}$$

for any nonzero vector s. Thus

$$|p_k^T g_k| = |g_k^T B_k^{-1} g_k| \ge \lambda_{\min}(B_k^{-1})\|g_k\|_2^2 \ge \lambda_{\max}^{-1}\|g_k\|_2^2$$

In addition

$$\|p_k\|_2^2 = g_k^T B_k^{-2} g_k \le \lambda_{\max}(B_k^{-2})\|g_k\|_2^2 \le \lambda_{\min}^{-2}\|g_k\|_2^2,$$

and hence

$$\|p_k\|_2 \le \lambda_{\min}^{-1}\|g_k\|_2$$

which leads to

$$\frac{|p_k^T g_k|}{\|p_k\|_2} \ge \frac{\lambda_{\min}}{\lambda_{\max}}\|g_k\|_2$$

Thus

$$\min\left(|p_k^T g_k|, |p_k^T g_k|/\|p_k\|_2\right) \ge \lambda_{\max}^{-1}\|g_k\|_2 \min\left(\lambda_{\min}, \|g_k\|_2\right).$$

and hence

$$\lim_{k\to\infty} \min\left(|p_k^T g_k|, |p_k^T g_k|/\|p_k\|_2\right) = 0$$

implies, as before, that $\lim_{k\to\infty} g_k = 0$.

2.6 Proof of Theorem 2.6

Consider the sequence of iterates x_k, $k \in \mathcal{K}$, whose limit is x_*. By continuity, H_k is positive definite for all such k sufficiently large. In particular, we have that there is a $k_0 \ge 0$ such that

$$p_k^T H_k p_k \ge \tfrac{1}{2}\lambda_{\min}(H_*)\|p_k\|_2^2$$

for all $k \in \mathcal{K} \ge k_0$, where $\lambda_{\min}(H_*)$ is the smallest eigenvalue of $H(x_*)$. We may then deduce that

$$|p_k^T g_k| = -p_k^T g_k = p_k^T H_k p_k \ge \tfrac{1}{2}\lambda_{\min}(H_*)\|p_k\|_2^2. \tag{2.3}$$

for all such k, and also that

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

since Theorem 2.5 implies that at least one of the left-hand sides of (2.3) and

$$\frac{|p_k^T g_k|}{\|p_k\|_2} = -\frac{p_k^T g_k}{\|p_k\|_2} \ge \tfrac{1}{2}\lambda_{\min}(H_*)\|p_k\|_2$$

converges to zero for such k.

From Taylor's theorem, there is a z_k between x_k and $x_k + p_k$ such that

$$f(x_k + p_k) = f_k + p_k^T g_k + \tfrac{1}{2}p_k^T H(z_k) p_k.$$

Thus, the Lipschitz continuity of H gives that

$$\begin{aligned} f(x_k + p_k) - f_k - \tfrac{1}{2}p_k^T g_k &= \tfrac{1}{2}(p_k^T g_k + p_k^T H(z_k) p_k) \\ &= \tfrac{1}{2}(p_k^T g_k + p_k^T H_k p_k) + \tfrac{1}{2}(p_k^T(H(z_k) - H_k)p_k) \\ &\le \tfrac{1}{2}\gamma\|z_k - x_k\|_2\|p_k\|_2^2 \le \tfrac{1}{2}\gamma\|p_k\|_2^3 \end{aligned} \tag{2.4}$$

since $H_k p_k + g_k = 0$. Now pick k sufficiently large so that

$$\gamma\|p_k\|_2 \le \lambda_{\min}(H_*)(1 - 2\beta).$$

In this case, (2.3) and (2.4) give that

$$f(x_k + p_k) - f_k \le \tfrac{1}{2}p_k^T g_k + \tfrac{1}{2}\lambda_{\min}(H_*)(1-2\beta)\|p_k\|_2^2 \le \tfrac{1}{2}(1 - (1-2\beta))p_k^T g_k = \beta p_k^T g_k,$$

and thus that a unit stepsize satisfies the Armijo condition, which proves (i).

To obtain the remaining results, note that $\|H_k^{-1}\|_2 \le 2/\lambda_{\min}(H_*)$ for all sufficiently large $k \in \mathcal{K}$. The iteration gives

$$x_{k+1} - x_* = x_k - x_* - H_k^{-1} g_k = x_k - x_* - H_k^{-1}(g_k - g(x_*)) = H_k^{-1}\left(g(x_*) - g_k - H_k(x_* - x_k)\right).$$

But Theorem 1.3 gives that

$$\|g(x_*) - g_k - H_k(x_* - x_k)\|_2 \le \gamma\|x_* - x_k\|_2^2.$$

Hence

$$\|x_{k+1} - x_*\|_2 \le \gamma\|H_k^{-1}\|_2\|x_* - x_k\|_2^2$$

which is (iii) when $\kappa = 2\gamma/\lambda_{\min}(H_*)$. Result (ii) follows since once an iterate becomes sufficiently close to x_*, (iii) implies that the next is even closer.

2.7 Conjugate Gradient methods

All of the results given here are easy to verify, and may be found in any of the books of suggested background reading material. The fact that any $p_k = p^i$ is a descent direction follows immediately since the identity

$$g^{i-1\,T} d^{i-1} = d^{i-1\,T}(g + Bp^{i-1}) = d^{i-1\,T} g + \sum_{j=0}^{i-2} \alpha_j d^{i-1\,T} B d^j = d^{i-1\,T} g$$

shows that if p^i minimizes $q(p)$ in $\mathcal{D}^i$ then

$$p^i = p^{i-1} - \frac{g^{i-1\,T} d^{i-1}}{d^{i-1\,T} B d^{i-1}} d^{i-1} = p^{i-1} - \frac{g^T d^{i-1}}{d^{i-1\,T} B d^{i-1}} d^{i-1}.$$

Thus

$$g^T p^i = g^T p^{i-1} - \frac{(g^T d^{i-1})^2}{d^{i-1\,T} B d^{i-1}},$$

from which it follows that $g^T p^i < g^T p^{i-1}$. The result then follows by induction, since

$$g^T p^1 = -\frac{\|g\|_2^4}{g^T B g} < 0.$$