SECTION C: CONTINUOUS OPTIMISATION
PROBLEM SET 3: SOLUTIONS

HONOUR SCHOOL OF MATHEMATICS, OXFORD UNIVERSITY
HILARY TERM 2006, DR RAPHAEL HAUSER

Solution to Problem 1: (i) Since $x_1 = x_0 + \alpha_0 d_0$ and $d_0 = -\nabla f(x_0)$, we have

$$\begin{aligned}\nabla f(x_1) &= 2Bx_1 + b = 2Bx_0 - 2\alpha_0 B\nabla f(x_0) + b\\ &= \nabla f(x_0) - 2\alpha_0 B\nabla f(x_0) \in \operatorname{span}\{\nabla f(x_0), B\nabla f(x_0)\}.\end{aligned}$$

(ii) We have shown this for $k = 0$, so we may assume it is true for $k \le l$, and then

$$\begin{aligned}\nabla f(x_{l+1}) &= 2Bx_{l+1} + b = 2Bx_l + 2\alpha_l B d_l + b\\ &= \nabla f(x_l) + 2\alpha_l B d_l. \end{aligned} \tag{0.1}$$

Because of

$$\operatorname{span}\{d_0, \dots, d_k\} = \operatorname{span}\{\nabla f(x_0), \dots, \nabla f(x_k)\}.$$

and the induction hypothesis we have $d_l \in \mathcal{K}_l$. Therefore, (0.1) shows

$$\nabla f(x_{l+1}) \in \operatorname{span}\big(\{\nabla f(x_l)\} \cup B\mathcal{K}_l\big) = \mathcal{K}_{l+1}.$$

(iii) This follows from the identity

$$(\mathrm{I} + A)^p = \mathrm{I} + \binom{p}{1}A + \binom{p}{2}A^2 + \cdots + \binom{p}{p-1}A^{p-1} + A^p$$

which is easily checked by induction on p.

(iv) Since $\operatorname{rank}(A) = r$, the image space of A is of dimension r. Therefore, at most r of the vectors $A\nabla f(x_0), \dots, A^k\nabla f(x_0)$ are linearly independent, and if $\nabla f(x_0)$ is linearly independent of the image space of A, then $\mathcal{K}_k$ is at most $r+1$ dimensional.

(v) In the proof of Lemma 2.3 we have shown that $\nabla f(x_j) \perp \nabla f(x_k)$ for all $j \neq k$. Since $\mathcal{K}_k$ is at most $r+1$ dimensional for all k, there are at most $r+1$ mutually orthogonal nonzero vectors in this space, which shows that it must be the case that $\nabla f(x_k) = 0$ for some $k \le r$. But since f is a strictly convex function, this is the exact characterisation of the global minimiser (see Lecture 2).

Solution to Problem 3: If $\Delta_k < \frac{\epsilon}{14\beta}$ for some k, then let

$$p := \max\{q \in \mathbb{N}_0 : \Delta_k \ge \Delta_{k-1} \ge \cdots \ge \Delta_{k-q}\}.$$

Since $\Delta_0 \ge \frac{\epsilon}{14\beta}$, it is the case that $k - p > 0$ and Δ_{k-p} was obtained by shrinking Δ_{k-p-1} via the relation

$$\Delta_{k-p} = \frac{1}{4}\Delta_{k-p-1}.$$

But now $\Delta_{k-p} \le \Delta_k < \frac{\epsilon}{14\beta}$ implies

$$\Delta_{k-p-1} < \frac{2\epsilon}{7\beta},$$

and then Lemma 3.4 and equation (1.5) of Lecture 6 show that $\Delta_{k-p} \ge \Delta_{k-p-1}$. This contradicts what we found above and proves our claim.

Solution to Problem 4: (i) If $\nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k) \le 0$ then

$$y_k^c = x_k - \Delta_k / \|\nabla f(x_k)\| \nabla f(x_k)$$

and

$$m_k(x_k) - m_k(y_k^c) \ge \nabla f(x_k)^{\mathrm{T}} \left(\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k) \right) = \Delta_k \|\nabla f(x_k)\| \ge \Delta_k \epsilon \overset{Prob.3}{\ge} \frac{\epsilon^2}{14\beta}.$$

(ii) Under these conditions we have

$$\begin{aligned} m_k(x_k) - m_k(y_k^c) &= \nabla f(x_k)^{\mathrm{T}} \left(\alpha_k^c \nabla f(x_k)\right) - \frac{1}{2} \left(\alpha_k^c\right)^2 \nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k) \\ &= \frac{1}{2} \frac{\|\nabla f(x_k)\|^4}{\nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k)} \ge \frac{1}{2} \frac{\|\nabla f(x_k)\|^2}{\beta} \ge \frac{\epsilon^2}{2\beta}, \end{aligned}$$

where we have used the bound $\|B_k\| \le \beta$.

(iii) Since $\nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k) > 0$, the line-search objective function

$$\phi(\alpha) = f(x_k) - \alpha \|\nabla f(x_k)\|^2 + \frac{\alpha^2}{2} \nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k)$$

is strictly convex. Therefore, the fact that the Cauchy point can be written as the convex combination

$$y_k^c = x_k + \frac{\alpha_k^c}{\alpha_k^u} (y_k^u - x_k)$$

implies

$$\phi(\alpha_k^c) < \left(1 - \frac{\alpha_k^c}{\alpha_k^u}\right) \phi(0) + \frac{\alpha_k^c}{\alpha_k^u} \phi(\alpha_k^u),$$

and hence,

$$\begin{aligned} m_k(x_k) - m_k(y_k^c) = \phi(0) - \phi(\alpha_k^c) &> \frac{\alpha_k^c}{\alpha_k^u} \left(m_k(x_k) - m_k(y_k^u)\right) \\ &\overset{(ii)}{\ge} \frac{1}{2} \frac{\|\nabla f(x_k)\|^4}{\nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k)} \frac{\Delta_k \nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k)}{\|\nabla f(x_k)\|^3} \\ &= \frac{\|\nabla f(x_k)\| \Delta_k}{2} = \frac{\epsilon \Delta_k}{2} \overset{Prob.3}{\ge} \frac{\epsilon^2}{28\beta} \end{aligned}$$

(iv) The list of cases we considered is exhaustive because of formula (2.1) from Lecture 6 and the formula for α_k^u preceding it.

(v) Because of the benchmarking of y_{k+1} against the Cauchy point and parts (i)–(iii), we have

$$m_k(x_k) - m_k(y_{k+1}) \geq m_k(x_k) - m_k(y_k^c) \geq \frac{\epsilon^2}{28\beta}.$$

On the other hand, since the step was accepted, formula (1.4) from Lecture 6 shows

$$f(x_k) - f(y_{k+1}) > \eta\big(m_k(x_k) - m_k(y_{k+1})\big) \geq \frac{\eta\epsilon^2}{28\beta}.$$