SECTION C: CONTINUOUS OPTIMISATION
PROBLEM SET 3

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

Instructions: Problem 2 is not mandatory, since it is not asterisked, but it is provided as an additional exercise for exam preparations.

***Problem 1.** Consider applying the conjugate gradient algorithm to the unconstrained minimisation problem $\min_{x\in\mathbb{R}^n} f(x)$, where $f(x) = x^{\mathrm{T}}Bx + b^{\mathrm{T}}x + a$ and B is a positive definite symmetric $n \times n$ matrix.

(i) Using the fact that $x_1 = x_0 + \alpha_0 d_0$, show that
$$\nabla f(x_1) \in \mathrm{span}\{\nabla f(x_0), B\nabla f(x_0)\}.$$

(ii) Generalise this result and show by induction that
$$\nabla f(x_k) \in \mathcal{K}_k := \mathrm{span}\{\nabla f(x_0), B\nabla f(x_0), \ldots, B^k\nabla f(x_0)\}$$
for $k = 0, \ldots, n$. The spaces $\mathcal{K}_k$ are called Krylov subspaces. Hint: in the proof of Lemma 2.3 of Lecture 5 we showed that
$$\mathrm{span}\{d_0, \ldots, d_k\} = \mathrm{span}\{\nabla f(x_0), \ldots, \nabla f(x_k)\}. \tag{0.1}$$

(iii) Now let $B = \mathrm{I} + A$ where $\mathrm{rank}(A) = r$. Show that
$$\mathcal{K}_k = \mathrm{span}\{\nabla f(x_0), A\nabla f(x_0), \ldots, A^k\nabla f(x_0)\}.$$

(iv) Show that $\mathcal{K}_k$ can be at most $r + 1$ dimensional.

(v) Using part (iv) of this exercise and an equation derived in the proof of Lemma 2.3 from Lecture 5, show that the conjugate gradient algorithm must terminate after at most $r + 1$ iterations at an iterate that corresponds to the minimiser of f.

Problem 2. Let $\|\cdot\|_N : \mathbb{R}^n \to \mathbb{R}$ be a norm, that is,
$$\begin{aligned}
\|x\|_N &\geq 0 \qquad \forall x \in \mathbb{R}^n,\\
\|x\|_N &= 0 \Leftrightarrow x = 0,\\
\|\lambda x\|_N &= |\lambda|\|x\|_N \qquad \forall x \in \mathbb{R}^n, \lambda \in \mathbb{R},\\
\|x + y\|_N &\leq \|x\|_N + \|y\|_N \qquad \forall x, y \in \mathbb{R}^n.
\end{aligned}$$

(i) Show that the unit ball $B_1 := \{x \in \mathbb{R}^n : \|x\|_N \leq 1\}$ is a convex set.

(ii) Consider a trust region method with trust region $R_k = x_k + \Delta_k B_1$, that is, R_k is the unit ball B_1 dilated or shrunk by a factor Δ_k and moved so that its centre lies at x_k. Now let
$$\phi(\alpha) = f(x_k) - \alpha\nabla f(x_k)^{\mathrm{T}}\nabla f(x_k) + \frac{\alpha^2}{2}\nabla f(x_k)^{\mathrm{T}}B_k\nabla f(x_k),$$

and consider the numbers $\alpha_k^u = \arg\min_{\alpha\geq 0} \phi(\alpha)$ and

$$\alpha_k^c = \arg\min_{\alpha\geq 0} \phi(\alpha) \quad \text{s.t.} \quad \|x_k - \alpha\nabla f(x_k)\|_N \leq \Delta_k.$$

We call $y_k^c = x_k - \alpha_k^c \nabla f(x_k)$ the Cauchy point, just as we did in the case of the Euclidean norm $\|\cdot\|_2$. Show that ϕ is strictly decreasing and $\alpha \mapsto \|x_k - \alpha\nabla f(x_k)\|_N$ strictly increasing over the interval $[0, \alpha_k^u]$.

(iii) Use the results of (ii) to derive a formula for the Cauchy point.

***Problem 3.** Under the assumptions of Theorem 1.2 of Lecture 6, prove that if $\Delta_k \geq \frac{\epsilon}{14\beta}$ for $k = 0$ then this inequality holds true for all k.

***Problem 4.** Prove Claim 2 in the proof of Theorem 1.2 from Lecture 6 under the additional assumption that $\Delta_0 \geq \frac{\epsilon}{14\beta}$. That is, in the case where $\|\nabla f(x_k)\| \geq \epsilon$ for all k, prove that whenever $x_{k+1} = y_{k+1}$ occurs, we have

$$f(x_{k+1}) - f(x_k) \leq -\frac{\eta\epsilon^2}{28\beta}.$$

We cut the proof into several stages as follows:

(i) Show that if $\nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k) \leq 0$, then

$$m_k(x_k) - m_k(y_k^c) \geq \frac{\epsilon^2}{14\beta}.$$

Hint: use the result of Problem 3 and formula (1.6) from Lecture 6.

(ii) Show that if $\nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k) > 0$ and $\alpha_k^c = \|\nabla f(x_k)\|^2 / \nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k)$ then

$$m_k(x_k) - m_k(y_k^c) \geq \frac{\epsilon^2}{2\beta}.$$

(iii) Show that if $\nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k) > 0$ and $\alpha_k^c = \Delta_k / \|\nabla f(x_k)\|$ then

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

Hint: Use the fact that in this case

$$0 < \alpha_k^c = \frac{\Delta_k}{\|\nabla f(x_k)\|} < \frac{\|\nabla f(x_k)\|^2}{\nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k)}$$

and exploit the convexity of the line-search objective function.

(iv) Why are these all the different cases we need to consider for bounds on $m_k(x_k) - m_k(y_k^c)$?

(v) Conclude that Claim 2 is true.