SECTION C: CONTINUOUS OPTIMISATION
PROBLEM SET 1: SOLUTIONS

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

Solution to Problem 1.

(i) The second order Taylor development of $t \mapsto f(x+tu)$ around $t=0$ is

$$f(x+tu) = f(x) + t\nabla f(x)^{\mathrm{T}} u + \frac{1}{2} t^2 u^{\mathrm{T}} H(x) u + o(t^2), \tag{0.1}$$

where $o(t^2)$ is a function such that

$$\lim_{t\to 0} \frac{o(t^2)}{t^2} = 0.$$

(ii) Theorem 2.4 (ii) shows

$$f(x+tu) \geq f(x) + t\nabla f(x)^{\mathrm{T}} u.$$

Therefore, (0.1) shows that

$$\frac{1}{2} t^2 u^{\mathrm{T}} H(x) u + o(t^2) = f(x+tu) - f(x) - t\nabla f(x)^{\mathrm{T}} u \geq 0.$$

Dividing by t^2 and taking limits, we obtain

$$u^{\mathrm{T}} H(x) u = u^{\mathrm{T}} H(x) u + 2 \lim_{t\to 0} \frac{o(t^2)}{t^2} = 2 \lim_{t\to 0} \frac{\frac{1}{2} t^2 u^{\mathrm{T}} H(x) u + o(t^2)}{t^2} \geq 0.$$

Since $u \in \mathbb{R}^n$ was arbitrary, this shows that $H(x)$ is positive semidefinite.

(iii) The fundamental theorem of differential and integral calculus shows

$$f(x+tu) = f(x) + \int_0^1 \nabla f(x+\vartheta t u)^{\mathrm{T}} t u \, d\vartheta \tag{0.2}$$

and

$$\nabla f(x+\vartheta t u)^{\mathrm{T}} u = \nabla f(x)^{\mathrm{T}} u + \int_0^1 \vartheta t u^{\mathrm{T}} H(x+\tau\vartheta t u) u \, d\tau. \tag{0.3}$$

The claim follows by substitution of (0.3) into (0.2).

(iv) The assumption $H(y) \succeq 0$ (pos. semidefinite) implies

$$u^{\mathrm{T}} H(x+\tau\vartheta t u) u \geq 0 \tag{0.4}$$

for all τ, ϑ, t such that $x + \tau\vartheta t u \in D$. In particular, this holds when $x + tu \in D$ and $\tau, \vartheta \in [0,1]$, because D is convex. Therefore,

$$\begin{aligned} f(x+tu) &= f(x) + t\nabla f(x)^{\mathrm{T}} u + t^2 \int_0^1 \int_0^1 \vartheta u^{\mathrm{T}} H(s+\tau\vartheta t u) u \, d\tau d\vartheta \\ &\geq f(x) + t\nabla f(x)^{\mathrm{T}} u, \end{aligned} \tag{0.5}$$

where the inequality holds because an integral of a nonnegative function is nonnegative. By virtue of Theorem 2.4 (ii), (0.5) shows that the function $t \mapsto f(x+tu)$ is convex in t. That is to say, for any y of the form $y = x + tu$ and any $\lambda \in [0,1]$ we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \tag{0.6}$$

But this holds true for all $u \in \mathbb{R}^n$, so in particular any $y \in D$ is of this form with $u = y - x$. Therefore, f is convex on D.

(v) The argument is the same as in (iv), except that the inequalities in (0.5) and (0.6) change to strict inequalities.

Solution to Problem 2.

(i) $f''(x) = C$ for all x, and since $C \succ 0$, part (v) of Problem 1 shows that f is strictly convex.

(ii) $\phi(\alpha) = (a + b^{\mathrm{T}} x_k + \frac{1}{2} x_k^{\mathrm{T}} C x_k) + \alpha \times (b^{\mathrm{T}} d_k + x_k^{\mathrm{T}} C d_k) + \alpha^2 \times \left(\frac{1}{2} d_k^{\mathrm{T}} C d_k\right)$ is a polynomial of degree 2 in α. Moreover, $\phi''(\alpha) = d_k^{\mathrm{T}} C d_k > 0$, since $C \succ 0$ and $d_k \neq 0$. Part (v) of Problem 1 therefore shows that ϕ is strictly convex.

(iii) $\phi'(\alpha) = (b^{\mathrm{T}} d_k + x_k^{\mathrm{T}} C d_k) + \alpha \times (d_k^{\mathrm{T}} C d_k)$. Since ϕ is strictly convex, it has a unique local minimiser α^* (which is also the global minimiser). The first order optimality condition $\phi'(\alpha^*) = 0$ implies

$$\alpha^* = -\frac{b^{\mathrm{T}} d_k + x_k^{\mathrm{T}} C d_k}{d_k^{\mathrm{T}} C d_k}.$$

(iv) We have $\alpha^* > 0$, because $d_k^{\mathrm{T}} C d_k > 0$ and

$$b^{\mathrm{T}} d_k + x_k^{\mathrm{T}} C d_k = \phi'(0) = \nabla f(x_k)^{\mathrm{T}} d_k < 0$$

by the assumption that d_k is a descent direction. Moreover,

$$\begin{aligned}
\phi(\alpha^*) &= (a + b^{\mathrm{T}} x_k + \frac{1}{2} x_k^{\mathrm{T}} C x_k) - \frac{(b^{\mathrm{T}} d_k + x_k^{\mathrm{T}} C d_k)^2}{2 d_k^{\mathrm{T}} C d_k} \\
&< (a + b^{\mathrm{T}} x_k + \frac{1}{2} x_k^{\mathrm{T}} C x_k) - c_1 \frac{(b^{\mathrm{T}} d_k + x_k^{\mathrm{T}} C d_k)^2}{d_k^{\mathrm{T}} C d_k} = \phi(0) + c_1 \alpha^* \phi'(0),
\end{aligned}$$

where the inequality follows from the assumption that $c_1 < 1/2$. This shows that the first Wolfe condition holds. The second Wolfe condition holds trivially, because $\phi'(0) = \nabla f(x_k) d_k < 0$, and hence,

$$\phi'(\alpha^*) = 0 > c_2 \phi'(0).$$

Solution to Problem 3.

(i) Since $\phi'(\alpha) < 0$ for all $\alpha \in [0, \alpha_k)$, it is true that

$$\begin{aligned}
f(x_{k+1}) &= f(x_k) + \int_0^\beta \phi'(\alpha) d\alpha + \int_\beta^{\alpha_k} \phi'(\alpha) d\alpha \\
&< f(x_k) + \int_0^\beta \phi'(\alpha) d\alpha \qquad \forall \beta \in [0, \alpha_k).
\end{aligned} \tag{0.7}$$

(ii) We have

$$\begin{aligned}
\phi'(\alpha) &= \nabla f(x_k + \alpha d_k)^{\mathrm{T}} d_k \\
&= \big(\nabla f(x_k) + \nabla f(x_k + \alpha d_k) - \nabla f(x_k)\big)^{\mathrm{T}} d_k \\
&\overset{\text{C.S.}}{\leq} \nabla f(x_k)^{\mathrm{T}} d_k + \|d_k\| \times \|\nabla f(x_k + \alpha d_k) - \nabla f(x_k)\| \\
&\overset{\text{Lip.}}{\leq} \nabla f(x_k)^{\mathrm{T}} d_k + \alpha \Lambda \|d_k\|^2.
\end{aligned} \tag{0.8}$$

(iii) It follows from (0.8) that

$$\nabla f(x_k)^{\mathrm{T}} d_k + \alpha \Lambda \|d_k\|^2 < 0 \tag{0.9}$$

implies $\phi'(\alpha) < 0$. But (0.9) is equivalent to

$$\alpha < \hat{\beta} := -\frac{\nabla f(x_k)^{\mathrm{T}} d_k}{\Lambda \|d_k\|^2}$$

Since therefore $\phi'(\alpha) < 0$ for all $\alpha < \hat{\beta}$, it must be true that $\hat{\beta} \in [0, \alpha_k]$. Using this in Equation (0.7), we find

$$\begin{aligned}
f(x_{k+1}) - f(x_k) &\leq \int_0^{\hat{\beta}} \big(\nabla f(x_k)^{\mathrm{T}} d_k + \alpha \Lambda \|d_k\|^2\big)\, d\alpha \\
&= \|d_k\| \times \|\nabla f(x_k)\| \cos\theta_k \times \hat{\beta} + \frac{\hat{\beta}^2}{2} \Lambda \|d_k\|^2 \\
&= \frac{\|d_k\|^2 \times \|\nabla f(x_k)\|^2 \cos^2\theta_k}{\Lambda \|d_k\|^2} \big(-1 + \frac{1}{2}\big),
\end{aligned}$$

which proves the required formula.