SECTION C: CONTINUOUS OPTIMISATION
PROBLEM SET 2: SOLUTIONS

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

Solution to Problem 1.

(i) We have

$$\begin{aligned}
&\big(B+UV^{\mathrm{T}}\big)\big(B^{-1}-B^{-1}U(\mathrm{I}+V^{\mathrm{T}}B^{-1}U)^{-1}V^{\mathrm{T}}B^{-1}\big)\\
&=\mathrm{I}-U(\mathrm{I}+V^{\mathrm{T}}B^{-1}U)^{-1}V^{\mathrm{T}}B^{-1}+UV^{\mathrm{T}}B^{-1}-U\big(V^{\mathrm{T}}B^{-1}U\big)(\mathrm{I}+V^{\mathrm{T}}B^{-1}U)^{-1}V^{\mathrm{T}}B^{-1}\\
&=\mathrm{I}-U\big(\mathrm{I}+V^{\mathrm{T}}B^{-1}U\big)(\mathrm{I}+V^{\mathrm{T}}B^{-1}U)^{-1}V^{\mathrm{T}}B^{-1}+UV^{\mathrm{T}}B^{-1}\\
&=\mathrm{I}-UV^{\mathrm{T}}B^{-1}+UV^{\mathrm{T}}B^{-1}=\mathrm{I}\,.
\end{aligned}$$

(ii) We observe that we never used the condition $m \leq n$ in part (i). Therefore, the SMW formula should be applicable to

$$\big(\mathrm{I}+V^{\mathrm{T}}(B^{-1}U)\big)^{-1}=\mathrm{I}-\mathrm{I}\,V^{\mathrm{T}}(\mathrm{I}+B^{-1}U\,\mathrm{I}\,V^{\mathrm{T}})^{-1}B^{-1}U\,\mathrm{I}=\mathrm{I}-V^{\mathrm{T}}(B+UV^{\mathrm{T}})^{-1}U,$$

which suggests that $\mathrm{I}+V^{\mathrm{T}}B^{-1}U$ is invertible. We can easily check this by multiplying the above result with $\mathrm{I}+V^{\mathrm{T}}B^{-1}U$:

$$\begin{aligned}
&\big(\mathrm{I}+V^{\mathrm{T}}(B^{-1}U)\big)\big(\mathrm{I}-V^{\mathrm{T}}(B+UV^{\mathrm{T}})^{-1}U\big)\\
&=\mathrm{I}-V^{\mathrm{T}}(B+UV^{\mathrm{T}})^{-1}U+V^{\mathrm{T}}B^{-1}U-V^{\mathrm{T}}B^{-1}UV^{\mathrm{T}}(B+UV^{\mathrm{T}})^{-1}U\\
&=\mathrm{I}-V^{\mathrm{T}}\big(\mathrm{I}+B^{-1}UV^{\mathrm{T}}\big)(B+UV^{\mathrm{T}})^{-1}U+V^{\mathrm{T}}B^{-1}U\\
&=\mathrm{I}-V^{\mathrm{T}}B^{-1}U+V^{\mathrm{T}}B^{-1}U=\mathrm{I}\,.
\end{aligned}$$

Solution to Problem 2.

(i) We have

$$q(x)=f(x_k)+\nabla f(x_k)^{\mathrm{T}}(x-x_k)+\frac{1}{2}(x-x_k)^{\mathrm{T}}\nabla^2 f(x_k)(x-x_k).$$

(ii) $\nabla q(x^*)=0$ implies

$$x^*=x_k-\nabla^2 f(x_k)^{-1}\nabla f(x_k).$$

Since $\nabla^2 q(x)=\nabla^2 f(x_k)\succ 0$ for all $x\in\mathbb{R}^n$, the second order sufficient optimality conditions hold at x^*, and x^* is the global minimiser of $q(x)$.

(iii) x^*-x_k is exactly equal to the Newton-Raphson step for f applied at x_k.

Solution to Problem 3.

(i) f is strictly convex because $\nabla^2 f(x) \equiv \mathrm{Diag}(\kappa, 1) \succ 0$. Moreover, $\nabla f(0) = \mathrm{Diag}(\kappa, 1)0 = 0$, thus $x^* = 0$ is a local minimiser because the 2nd order sufficient optimality conditions hold there. Finally, since f is strictly convex, this function has a unique local minimiser, which is also its global minimiser.

(ii) Arguing inductively, suppose $x_k = \tau(e, \kappa)^{\mathrm{T}}$, where $\tau > 0$ and $e = (-1)^k$. Then $\nabla f(x_k) \sim (e, 1)^{\mathrm{T}}$, so the search direction $d_k = (-e, -1)^{\mathrm{T}}$ is a positive scalar multiple of the steepest descent direction. If $\alpha_k = (2\kappa\tau)/(\kappa + 1)$ we find that

$$x_{k+1} = x_k + \alpha_k d_k = \tau \frac{\kappa - 1}{\kappa + 1} \begin{pmatrix} -e \\ \kappa \end{pmatrix} \tag{0.1}$$

and then $\nabla f(x_{k+1}) \sim (-e, 1)^{\mathrm{T}} \perp d_k$. Therefore, the step length α_k corresponds to an exact line search. Using the formula (0.1) inductively, we find

$$x_k = \left(\frac{\kappa - 1}{\kappa + 1}\right)^k \begin{pmatrix} (-1)^k \\ \kappa \end{pmatrix} \qquad \forall k \in \mathbb{N}_0.$$

(iii) The convergence is Q-linear because

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \rho = \frac{\kappa - 1}{\kappa + 1} < 1.$$

(iv) Expressed in the new coordinates, the objective function becomes

$$g(y) = f\big(x(y)\big) = \frac{1}{2}\left(\begin{pmatrix} \kappa^{-\frac{1}{2}} & 0 \\ 0 & 1 \end{pmatrix} y\right)^{\mathrm{T}} \begin{pmatrix} \kappa & 0 \\ 0 & 1 \end{pmatrix} \left(\begin{pmatrix} \kappa^{-\frac{1}{2}} & 0 \\ 0 & 1 \end{pmatrix} y\right) = \frac{1}{2} y^{\mathrm{T}} y,$$

and the starting point is $y_0 = \mathrm{Diag}(\kappa^{\frac{1}{2}}, 1)x_0 = (\kappa^{\frac{1}{2}}, \kappa)^{\mathrm{T}}$. The steepest descent direction at y_0 is $d_k = -\nabla g(y_0) = -y_0$. Therefore, the exact line search corresponds to the step length $\alpha_0 = 1$ and leads to the global minimiser $y^* = 0$ in one step.

(v) We find $z_0 = (\kappa + \kappa^2)^{-\frac{1}{2}}(1, 1)^{\mathrm{T}}$ and

$$h(z) := g(y(z)) = \frac{1}{2} z^{\mathrm{T}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} z.$$

This is the same situation as analysed in (ii), with $\kappa = 1$ and $\tau = (\kappa + \kappa^2)^{-\frac{1}{2}}$. Therefore, (0.1) shows that $z_1 = 0 = z^*$. Thus, again we have convergence in one step.

Solution to Problem 4.

(i) The secant equation is $B^{(K+1)}\delta = \gamma$. The validity of this equation is checked by direct calculation.

(ii) Let $f(x) = c + b^{\mathrm{T}}x + \frac{1}{2}x^{\mathrm{T}}Gx$. Then

$$\gamma = (b + Gx^{(k+1)}) - (b + Gx^{(k)}) = G(x^{(k+1)} - x^{(k)}) = G\delta,$$

and $\eta = (G - B^{(k)})\delta$. Therefore,

$$\begin{aligned}
B^{(k+1)} - G &= B^{(k)} - G + \frac{-(B^{(k)} - G)}{\delta^{\mathrm{T}}\delta}\delta\delta^{\mathrm{T}} - \delta\delta^{\mathrm{T}}\frac{B^{(k)} - G}{\delta^{\mathrm{T}}\delta} + \frac{\delta^{\mathrm{T}}(B^{(k)} - G)\delta}{(\delta^{\mathrm{T}}\delta)^2}\delta\delta^{\mathrm{T}} \\
&= \dots \\
&= \left(\mathrm{I} - \frac{\delta\delta^{\mathrm{T}}}{\delta^{\mathrm{T}}\delta}\right)\left(B^{(k)} - G\right)\left(\mathrm{I} - \frac{\delta\delta^{\mathrm{T}}}{\delta^{\mathrm{T}}\delta}\right).
\end{aligned}$$

(iii) We have

$$\begin{aligned}
\eta = \gamma - B^{(k)}\delta &= \begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}2 & 1\\1 & 1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} \\
&= \begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}-1\\0\end{pmatrix} = -\delta,
\end{aligned}$$

and

$$\begin{aligned}
B^{(k+1)} &= \begin{pmatrix}2 & 1\\1 & 1\end{pmatrix} + \frac{-2\delta\delta^{\mathrm{T}}}{\delta^{\mathrm{T}}\delta} + \frac{\delta^{\mathrm{T}}\delta}{(\delta^{\mathrm{T}}\delta)^2}\delta\delta^{\mathrm{T}} \\
&= \begin{pmatrix}2 & 1\\1 & 1\end{pmatrix} - \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix} = \begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}.
\end{aligned}$$

(iv) $B^{(k)}$ can become singular and the algorithm may not be able to compute a quasi-Newton step.