CNAc: Continuous Optimization — solutions to problem set 2

Honour School of Mathematics, Oxford University
Hilary Term 2006, Dr Nick Gould

Problem 1.
(i) Since $H(x) = \text{diag}(\kappa, 1)$ which is positive definite, any first-order critical point must be an isolated local minimizer. But $g(x) = \text{diag}(\kappa, 1)x$ which is zero if and only if $x = 0$, and thus $x_* = 0$ is the only local minimizer. Finally, as $f(x) \geq f(0) = 0$, x_* is the global minimizer.
(ii) Arguing inductively, suppose $x_k = \tau(e, \kappa)^T$, where $\tau > 0$ and $e = (-1)^k$. Then $g_k \equiv g(x_k) = \tau(\kappa e, \kappa)$. The steepest-descent method takes a step

$$x_k - \alpha_k g_k = \tau \begin{pmatrix} e(1-\alpha\kappa) \\ \kappa(1-\alpha) \end{pmatrix}$$

where the stepsize α_k is chosen to minimize $\phi(\alpha) = f(x_k - \alpha g_k)$. Since $\phi'(\alpha) = -\tau^2((1+e^2)\kappa^2 - \alpha\kappa^2(e^2\kappa+1)$, and as $e^2 = 1$, it follows that $\alpha_k = 2/(1+\kappa)$ and hence $x_{k+1} = \tau(\kappa-1)/(\kappa+1)(-e, \kappa)^T$.
(iii) The convergence is Q-linear because

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

The first component of x will be correct to one decimal place as soon as $\kappa^k < 0.1$, i.e., as soon as $k \geq \log 0.1 / \log \rho$. For $\kappa = 1000$, 1152 iterations are required, while for $\kappa = 10^6$, 1151296 iterations are needed!
(iv) Expressed in the new coordinates, the objective function becomes

$$\hat{f}(y) = f\big(x(y)\big) = \tfrac{1}{2} y^T y,$$

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

Problem 2.
(i) We have that $g(x) = g + Bx$ and thus $g(x+s) = g + Bx + Bs$. Subtracting these two identities yield the required result.
(ii) The secant condition and the requirement that the update be of rank one gives

$$B_{k+1}s_k = B_k s_k + \beta v^T s_k v = y_k$$

and hence that

$$(\beta v^T s_k)v = y_k - B_k s_k.$$

Thus v is parallel to $y_k - B_{k+1}s_k$ so long as

$$\beta v^T s_k \neq 0. \tag{1}$$

Assuming that (1), we may write $v = \gamma(y_k - B_{k+1}s_k)$ for some $\gamma \neq 0$, and hence that

$$\gamma = \beta v^T s_k = \beta\gamma s_k^T (y_k - B_{k+1}s_k)$$

from which we deduce that

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

so long as

$$s_k^T (y_k - B_{k+1} s_k) \neq 0.$$

(iii) The secant condition and the requirement that the update be of given form give

$$B_{k+1} s_k = B_k s_k + \theta y_k^T s_k y_k + \beta v^T s_k v = y_k$$

and hence that

$$\beta v^T s_k v = \phi y_k - B s_k, \;\text{ where }\; \phi = 1 - \theta y_k^T s_k.$$

Thus v is parallel to $\phi y_k - B_k s_k$ so long as

$$\beta v^T s_k \neq 0. \tag{2}$$

Assuming that (2), we may write $v = \gamma(\phi y_k - B_k s_k)$ for some $\gamma \neq 0$, and hence that

$$\gamma = \beta v^T s_k = \beta \gamma s_k^T (\phi y_k - B_k s_k)$$

from which we deduce that

$$B_{k+1} = B_k + (1-\phi) \frac{y_k y_k^T}{y_k^T s_k} + \frac{(\phi y_k - B_k s_k)(\phi y_k - B_k s_k)^T}{\phi y_k^T s_k - s_k^T B_k s_k}$$

so long as

$$\phi y_k^T s_k - s_k^T B_k s_k \neq 0.$$

(iv) The case $\phi = 0$ corresponds to the BFGS update, while $\phi = 1$ gives the symmetric rank-one update.

Problem 3.
If B_k is symmetric positive definite, we can write $B = LL^T$ for some non-singular matrix L.. Define $w = L^T s$ and let v be such that $lv = y$. Then

$$B^+ = L\left(I - \frac{ww^T}{w^T w} + \frac{vv^T}{v^T w}\right) L^T;$$

note that $P \stackrel{\text{def}}{=} I - ww^T / w^T w$ is a projection matrix, and $v^T w = s^T y > 0$. It thus remains to show that

$$M = P + \frac{vv^T}{v^T w}$$

is positive definite. Since vector u may be expressed as a linear combination of vectors in the range and null-spaces of P, and as the latter is simply vectors parallel to w, we write

$$u = \alpha w + Pt$$

for some $alpha$ and t (not both zero). We must show that $u^T M u > 0$. But a simple calculation (using the fact that $Pw = 0$) gives

$$u^T M u = \|Pt\|^2 + \beta^2 v^T w, \text{where} \beta = \alpha + \frac{v^T P t}{v^T w}.$$

Since by assumption not both t and α are zero and $v^T w > 0$, it then follows that $u^T M u > 0$.

Problem 4†.
(i) Since $x_1 = x_0 + \alpha_0 d_0$ and $d_0 = -g(x_0)$, we have

$$g(x_1) = Bx_1 + g = Bx_0 - \alpha_0 B g(x_0) + g = g(x_0) - \alpha_0 B g(x_0) \in \text{span}\{g(x_0), Bg(x_0)\}.$$

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

$$g(x_{l+1}) = Bx_{l+1} + g = Bx_l + \alpha_l Bd_l + g = g(x_l) + \alpha_l Bd_l. \tag{3}$$

Because

$$\text{span}\{d_0, \dots, d_k\} = \text{span}\{g(x_0), \dots, g(x_k)\}.$$

and the induction hypothesis we have $d_l \in \mathcal{K}_l$. Therefore, (3) shows

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

(iii) This follows from the identity

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

which is easily checked by induction on p.
(iv) Since $\text{rank}(A) = r$, the image space of A is of dimension r. Therefore, at most r of the vectors $Ag(x_0), \dots, A^k g(x_0)$ are linearly independent, and if $g(x_0)$ 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 $g(x_j) \perp g(x_k)$ 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 $g(x_k) = 0$ for some $k \leq r$. But since B is positive definite, f is bounded from below, and the only isolated local minimizer x_k is thus its global minimiser.

† Thanks to Raphael Hauser for this solution.