CNAc: Continuous Optimization
Solutions to problem set 4 — linearly constrained optimization and penalty methods

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

Problem 1.

(a) Suppose that x_* is a local but not global minimizer. Then there must be some $y \in \mathcal{C}$ for which $f(y) < f(x_*)$. But then by convexity $\alpha x_* + (1-\alpha)y \in \mathcal{C}$, and

$$f(\alpha x_* + (1-\alpha)y) \leq \alpha f(x_*) + (1-\alpha)f(y) < \alpha f(x_*) + (1-\alpha)f(x_*) = f(x_*)$$

for all $\alpha \in [0,1]$, Hence all points between x_* and y have a smaller value than $f(x_*)$, and thus x_* cannot be a local minimizer.

(b) We show that if x_* and y_* are distinct global minimizers of the convex function $f(x)$, then so is $\alpha x_* + (1-\alpha)y_*$ for all $\alpha \in [0,1]$. Since x_* and y_* are global minimizers, $f(x_*) = f(y_*)$. By convexity we thus have

$$f(\alpha x_* + (1-\alpha)y_*) \leq \alpha f(x_*) + (1-\alpha)f(y_*) = f(x_*)$$

But since x_* is a global minimizer

$$f(x_*) \leq f(\alpha x_* + (1-\alpha)y_*),$$

from which we deduce that $f(\alpha x_* + (1-\alpha)y_*) = f(x_*)$

(c) Suppose that x_* and y_* are distinct global minimizers, and thus $f(x_*) = f(y_*)$. Then since f is strictly convex,

$$f(\alpha x_* + (1-\alpha)y_*) < \alpha f(x_*) + (1-\alpha)f(y_*) = f(x_*)$$

for any $\alpha \in (0,1)$, which contradicts the global optimality of x_*. Thus there can only be a single global minimizer.

(d) Simple manipulation shows that for any distinct x and y,

$$f(\alpha x_* + (1-\alpha)y) - \alpha f(x_*) - (1-\alpha)f(y) = (\alpha^2 - \alpha)(x-y)^T H(x-y). \tag{0}$$

But $\alpha^2 - \alpha \leq 0$ for all $\alpha \in [0,1]$. Thus (0) shows that f is convex if H is positive semi-definite. If H is indefinite, f cannot be convex because (0) will be negative if we choose $x-y$ to be an eigenvector of H corresponding to a negative eigenvalue

(e) Since $\alpha^2 - \alpha < 0$ for all $\alpha \in (0,1)$, (0) shows that f is strictly convex if H is positive definite. If H has an eigenvalue $\lambda \leq 0$, f cannot be strictly convex because (0) will be less-than-or-equal-to zero if we choose $x-y$ to be an eigenvector of H corresponding to this eigenvalue.

Problem 2.

(a) We first need to check that $s^T Bs \geq 0$ when $As = 0$, as otherwise the solution lies at infinity. In all cases B is diagonal, so we write $B = \mathrm{diag}(b_{11} \;\; b_{22} \;\; b_{33})$. It is easy to see that the columns of the matrix

$$N = \begin{pmatrix} -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

form a basis for the null-space of A, so we need to check that

$$N^T BN = \begin{pmatrix} b_1 + b_2 & 0 \\ 0 & b_3 \end{pmatrix}$$

is positive semi-definite. For our first example $N^T BN$ has all its eigenvalues at 1, so the minimizer is finite. The minimizer satisfies

$$\begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}$$

which gives $x = (-2, 4, 1)$ and $y = 5$.

(b) In this case $N^T BN$ has eigenvalues 0 and 1, so there is a solution if and only if

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}$$

is consistent. The system gives $x_3 = 1$, but then the remaining equations lead to both $-x_2 + y = 1$ and $-x_2 + y = -1$. Thus the problem is unbounded from below.

(c) In this case $N^T BN$ has eigenvalues -1 and 1, so the problem is unbounded from below, and the solution lies at infinity.

Problem 3.
Iteration 0: there are no constraints active at the starting point $x_0 = (1/12, 25/28)$, so $\mathcal{W}_0 = \emptyset$. The unconstrained minimizer of $q(x)$ is at $x_0 + s_0 = (1, 1/2)$, so that $s_0 = (11/12, -11/28)$. Considering the line $x_0 + \alpha s_0$, we find that constraint

$$-[x]_1 - [x]_2 \geq -1 \tag{1}$$

is satisfied for all $0 \leq \alpha \leq 1/22$, constraint

$$-3[x]_1 - [x]_2 \geq -1.5 \tag{2}$$

is satisfied for all $0 \leq \alpha \leq 5/33$, constraint

$$[x]_1 \geq 0 \tag{3}$$

is satisfied for all $\alpha \geq 0$, and constraint

$$[x]_2 \geq 0 \tag{4}$$

is satisfied for all $0 \leq \alpha \leq 25/11 > 1$, so that the largest permitted stepsize is $\alpha_0 = 1/22$. Hence $x_1 = (1/8, 7/8)$, and $\mathcal{W}_1 = \mathcal{A}_1 = \{(1)\}$.

Iteration 1: we must find the minimizer s_1 of $q(x_1+s) = \frac{1}{2}([s]_1 - 7/8)^2 + \frac{1}{2}([s]_2 + 3/8)^2$ subject to $-[s]_1 - [s]_2 = 0$. Trivial calculation reveals that $s_1 = (5/8, -5/8)$. Considering the line $x_1 + \alpha s_1$, we find that constraint (2) is satisfied for $0 \le \alpha \le 1/5$, constraint (3) is satisfied for $\alpha \ge 0$, and constraint (4) is satisfied for $0 \le \alpha \le 7/5 > 1$—of course constraint (1) is always satisfied as $(1) \in \mathcal{W}_1$. Thus the largest permitted stepsize is $\alpha_1 = 1/5$, and hence $x_2 = (1/4, 3/4)$, and $\mathcal{W}_2 = \{(1), (2)\}$.
Iteration 2: now we need to find the minimizer s_2 of $q(x_2+s) = \frac{1}{2}([s]_1 - 3/4)^2 + \frac{1}{2}([s]_2 + 1/4)^2$ subject to $-[s]_1 - [s]_2 = 0$ and $-3[s]_1 - [s]_2 = 0$. The solution now is $s_2 = 0$, and the Lagrange multipliers satisfy the equation

$$\begin{pmatrix} -1 & -3 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} -[y_3]_1 \\ -[y_3]_2 \end{pmatrix} = \begin{pmatrix} 3/4 \\ -1/4 \end{pmatrix}.$$

Thus $x_3 = x_2 = (1/4, 3/4)$ and $y_3 = (-3/4, 1/2)$. Since $[y_3]_1 < 0$, progress can be made by deleting constraint (1), and hence $\mathcal{W}_3 = \{(2)\}$.
Iteration 3: we must find the minimizer s_3 of $q(x_3+s) = \frac{1}{2}([s]_1 - 3/4)^2 + \frac{1}{2}([s]_2 + 1/4)^2$ subject to $-3[s]_1 - [s]_2 = 0$. Again, trivial calculation reveals that $s_3 = (3/20, -9/20)$. Considering the line $x_3 + \alpha s_3$, we find that constraint (1) is satisfied for all $\alpha \ge 0$, constraint (3) is satisfied for $\alpha \ge 0$ and constraint (4) is satisfied for $0 \le \alpha \le 5/3 > 1$—of course now constraint (2) is always satisfied as $(2) \in \mathcal{W}_3$. Thus a stepsize $\alpha_4 = 1$ is allowed and $x_4 = (2/5, 3/10)$. Since the full step has been taken, we need to evaluate the Lagrange multipliers. We require

$$\begin{pmatrix} -3 \\ -1 \end{pmatrix} (-[y_4]_2) = \begin{pmatrix} 3/5 \\ 1/5 \end{pmatrix}$$

and thus $[y_4]_2 = 1/5 > 0$. Hence $x_* = x_4 = (2/5, 3/10)$ is the required minimizer.

Problem 4†.

(a) The KKT conditions are

$$\begin{aligned} -1 + 2y[x]_1 &= 0 && (5) \\ -1 + 2y[x]_2 &= 0 && (6) \\ 1 - [x]_1^2 - [x]_2^2 &= 0. && (7) \end{aligned}$$

If $y = 0$ then (5) and (6) are violated, so there are no solutions corresponding to this case. If $y \ne 0$ then $[x]_1 = [x]_2 = 1/(2y)$, thus (7) implies that the KKT points are (x_*, y_*) and $(-x_*, -y_*)$, where $y_* = [x_*]_1 = [x_*]_2 = 1/\sqrt{2}$.

(b) The quadratic penalty function is

$$\Phi(x, \mu) = -[x]_1 - [x]_2 + \frac{1}{2\mu}(1 - [x]_1^2 - [x]_2^2)^2.$$

The stationary points of $\Phi(x, \mu)$ satisfy

$$\nabla_x \Phi(x, \mu) = \begin{pmatrix} -1 - \frac{2[x]_1}{\mu}(1 - [x]_1^2 - [x]_2^2) \\ -1 - \frac{2[x]_2}{\mu}(1 - [x]_1^2 - [x]_2^2) \end{pmatrix}$$

which implies

$$\mu = -2[x]_1(1 - [x]_1^2 - [x]_2^2) = -2[x]_2(1 - [x]_1^2 - [x]_2^2). \tag{8}$$

Since $\mu > 0$, we have $1 - [x]_1^2 - [x]_2^2 \ne 0$, and hence (8) shows that $[x]_1 = [x]_2$. Substituting this back into (8), we find $2[x(\mu)]_1^3 - [x(\mu)]_1 - \mu/2 = 0$.

(c) We have

$$[x]_1(1 - 2[x]_1^2) + \mu/2 = 0. \tag{9}$$

So, as $\mu \to 0$, it must be the case that $[x]_1(1-2[x]_1^2) \to 0$. Since $[x]_1 \to 0$ would imply that $[x]_2 \to 0$, and hence the penalty term would blow up, this shows that

$$[x(\mu)]_1 \stackrel{\mu\to 0}{\longrightarrow} \frac{1}{\sqrt{2}} = [x_*]_1.$$

Assuming that $[x(\mu)]_1 = 1/\sqrt{2} + a\mu + O(\mu^2)$, (9) implies

$$\frac{1 - 2\left(1/\sqrt{2} + a\mu + O(\mu^2)\right)^2}{\mu} = -\frac{1}{2[x(\mu)]_1} \stackrel{\mu\to 0}{\longrightarrow} -\frac{1}{\sqrt{2}}.$$

Expanding the left hand side, we find

$$-\frac{1}{\sqrt{2}} = \lim_{\mu\to 0} \frac{-2\sqrt{2}a\mu + O(\mu^2)}{\mu} = -2\sqrt{2}a,$$

which shows that $a = 1/4$.

Problem 5.

The proof is identical to that of Theorem 5.1 so long as one uses the correct Lagrange multiplier estimates

$$[y_k]_i \stackrel{\text{def}}{=} -\frac{\|c(x_k)\|_2^2[c(x_k)]_i}{\mu_k}.$$

† Thanks to Raphael Hauser for this solution.