SECTION C: CONTINUOUS OPTIMISATION
PROBLEM SET 6: SOLUTIONS

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

Solution to Problem 1. (i) Let $z \in \mathbb{R}^n$ be such that $g_j(z) \geq g_j(x)$ and $Az = b$, and let $x(t) = x + t(z - x)$, $(t \in [0, 1])$. By concavity of g_j, we then have

$$g_j(x(t)) \geq (1 - t)g_j(x) + tg_j(z) \geq g_j(x)$$

for all $t \in [0, 1]$, and hence we have

$$\nabla g_j(x)^{\mathrm{T}}(z - x) = \frac{d}{dt}|_{t=0} g_j(x(t)) = \lim_{t \to 0+} \frac{g_j(x(t)) - g_j(x)}{t} \geq 0.$$

This shows

$$\{z \in \mathbb{R}^n : g_j(z) \geq g_j(x)\} \subseteq \{z \in \mathbb{R}^n : \nabla g_j(x)^{\mathrm{T}}(z - x) \geq 0\}.$$

The result now follows from the following relation:

$$\begin{aligned}
\{z \in \mathbb{R}^n : g_j(z) > g_j(x)\} &= \{z \in \mathbb{R}^n : g_j(z) \geq g_j(x)\}^\circ \\
&\subseteq \{z \in \mathbb{R}^n : \nabla g_j(x)^{\mathrm{T}}(z - x) \geq 0\}^\circ \\
&= \{z \in \mathbb{R}^n : \nabla g_j(x)^{\mathrm{T}}(z - x) > 0\}.
\end{aligned}$$

(ii) Let the Slater condition be satisfied and $z \in \mathbb{R}^n$ be such that $Az = b$ and $g_{\mathcal{I}}(z) > 0$. Since $g_{\mathcal{E}}(x) = Ax - b = 0$ are the equality constraints of (CP), the vectors $\{\nabla g_i(x) : i \in \mathcal{E}\}$ are exactly the row vectors of A, and these are linearly independent by the SCQ assumption.

Furthermore, let $x \in \mathcal{F}$, so that $A(z - x) = 0$, and let $x(t)$ be defined as above. Then for $j \in \mathcal{A}(x)$ we have $g_j(z) > 0 = g_j(x)$, and hence by part (i) we find

$$\nabla g_j(x)^{\mathrm{T}}(z - x) > 0.$$

This shows that $d := (z - x)$ satisfies

$$d^{\mathrm{T}} \nabla g_i(x) = 0, \qquad (i \in \mathcal{E}), \tag{0.1}$$

$$d^{\mathrm{T}} \nabla g_j(x) > 0, \qquad (j \in \mathcal{A}(x)), \tag{0.2}$$

which (together with the requirement that A has full row rank) form exactly the requirements of the MFCQ. This shows that the MFCQ is satisfied at all feasible points.

(iii) Let x^* be a feasible point where the MFCQ holds. Then, $\{\nabla g_i(x^*) : i \in \mathcal{E}\}$ is a linearly independent set of vectors, which shows that A has full row rank.

Furthermore, let $d \in \mathbb{R}^n$ be such that (0.1) and (0.2) are satisfied, and let $x(t) = x^* + td$ for $t \geq 0$. Then $Ax(t) = Ax + tAd = b$ for all t, and Taylor's theorem implies that $g_j(x(t)) > 0$ for all $j \in \mathcal{A}(x^*)$ for t sufficiently small. By continuity we also have

$g_j(x(t)) > 0$ for all $j \in \mathcal{I} \setminus \mathcal{A}(x^*)$ and t sufficiently small. Thus, there exists $t > 0$ such that $z = x(t)$ satisfies the conditions of the SCQ.

This shows that the SCQ holds true, and therefore by part (i) the MFCQ holds for all feasible points.

Solution to Problem 2. (i) we have to show that

$$\begin{aligned} &\min x_1 + \frac{2}{x_2} \\ \text{s.t.} \quad &- x_2 + \frac{1}{2} \le 0 \\ &- x_1 + x_2^2 \le 0 \\ &x_2 \ge 0. \end{aligned} \tag{0.3}$$

is a convex programming problem. Note that

$$D^2 f(x) = \begin{pmatrix} 0 & 0 \\ 0 & \frac{4}{x_2^3} \end{pmatrix} \succeq 0,$$

and it follows from the results of Lecture 1 that f is convex. Writing the problem in standard form we get

$$\begin{aligned} &\min f(x) = x_1 + \frac{2}{x_2} \\ \text{s.t.} \quad &g_1(x) = x_2 - \frac{1}{2} \ge 0, \\ &g_2(x) = x_1 - x_2^2 \ge 0, \\ &g_3(x) = x_2 \ge 0, \end{aligned}$$

g_1 and g_3 are affine functions and hence concave. g_2 is concave because $D^2 g_2(x) = \operatorname{diag}(0, -2)$ is negative semidefinite. Thus, (0.3) is a convex programming problem.

(ii) We are looking for local minimisers where $x_2 > 0$, that is, $3 \notin \mathcal{A}(x)$. The KKT conditions are the following,

$$\begin{aligned} \begin{pmatrix} 1 \\ -\frac{2}{x_2^{*2}} \end{pmatrix} &= \lambda_1^* \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda_2^* \begin{pmatrix} 1 \\ -2x_2^* \end{pmatrix} + \lambda_3^* \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ x_2^* - \frac{1}{2} &\ge 0, \\ x_1^* - x_2^{*2} &\ge 0, \\ x_2^* &\ge 0, \\ \lambda_1^*(x_2^* - \frac{1}{2}) &= 0, \\ \lambda_2^*(x_1^* - x_2^{*2}) &= 0, \\ \lambda_3^* x_2^* &= 0, \\ \lambda_1^*, \lambda_2^*, \lambda_3^* &\ge 0. \end{aligned}$$

Since we want $x_2^* > 0$ we must have $\lambda_3^* = 0$. Moreover, the first equation implies that $\lambda_2^* = 1$. Therefore,

$$x_1^* = x_2^{*2}, \tag{0.4}$$

and the first equation implies

$$\lambda_1^* = 2x_2^{*2} - \frac{2}{x_2^{*2}}. \tag{0.5}$$

If $\lambda_1^* \neq 0$ then $x_2^* = 1/2$. (0.4) implies $x_1^* = 1/\sqrt{2}$ and (0.5) implies $\lambda_1^* = -7$. This solution violates the last set of inequalities among the KKT conditions.

On the other hand, if $\lambda_1^* = 0$ then (0.5) implies $x_2^* = 1$. (0.4) shows that $x_1^* = 1$. Thus, $x^* = (1,1)$, $\lambda^* = (0,1,0)$ is the unique solution to the KKT conditions when $x_2^* > 0$.

The active set of at x^* is $\mathcal{A}(x^*) = \{2\}$. Moreover, the LICQ holds at this point. The KKT conditions are therefore both necessary and sufficient, proving that $x^* = (1,1)$ is the global minimiser.

Solution to Problem 3. (i) The Lagrangian of

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \; & c^{\mathrm{T}}x - \mu\sum_i \ln x_i \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{0.6}$$

is the following extended function,

$$\mathcal{L}(x,u,v) = \begin{cases} (c - A^{\mathrm{T}}v - u)^{\mathrm{T}}x + b^{\mathrm{T}}v - \mu\sum_i \ln x_i, & (u \geq 0, x > 0), \\ +\infty & (u \geq 0, x \not> 0), \\ -\infty & (u \not\geq 0). \end{cases}$$

(ii) The Lagrangian primal problem is $\min_x(\max_{(u,v)} \mathcal{L}(x,u,v)$. In order to show that this problem is equivalent to (0.6), we have to derive more explicit expressions for the objective function.

If $x \not> 0$ then $\max_{(u,v)} \mathcal{L}(x,u,v) = \mathcal{L}(x,0,0) = +\infty$.

Let us thus assume that $x > 0$. Then

$$\begin{aligned} \max_{(u,v)} \mathcal{L}(x,u,v) &= \max_{\{(u,v):u\geq 0\}} -u^{\mathrm{T}}x - (Ax-b)^{\mathrm{T}}v + (c^{\mathrm{T}}x - \mu\sum_i \ln x_i) && (0.7) \\ &= \max_v -(Ax-b)^{\mathrm{T}}v + (c^{\mathrm{T}}x - \mu\sum_i \ln x_i) && (0.8) \\ &= \begin{cases} +\infty & \text{if } Ax \neq b, \\ c^{\mathrm{T}}x - \mu\sum_i \ln x_i & \text{if } Ax = b, \end{cases} \end{aligned}$$

where (0.7) holds because when $u \not\geq 0$ then $\mathcal{L}(x,u,v) = -\infty$ is clearly not a maximum, and (0.8) holds because $u \geq 0$ and $x > 0$ imply $-u^{\mathrm{T}}x \leq 0$. In summary, we have

$$\max_{(u,v)} \mathcal{L}(x,u,v) = \begin{cases} +\infty & \text{if } x \not> 0 \text{ or } Ax \neq b, \\ c^{\mathrm{T}}x - \mu\sum_i \ln x_i & \text{otherwise.} \end{cases}$$

Therefore,

$$\text{(P)} \qquad \min_x \Big(\max_{(u,v)} \mathcal{L}(x,u,v) \Big)$$

is equivalent to

$$\begin{aligned} \min_{x\in\mathbb{R}^n}\ & c^{\mathrm{T}}x - \mu\sum_i \ln x_i \\ \text{s.t.}\quad & Ax = b \\ & x \geq 0, \end{aligned}$$

and this is equivalent to (0.6).

(iii) To work out the Lagrangian dual, we need explicit expressions for the objective function $\min_x \mathcal{L}(x,u,v)$.

If $u \ngeq 0$ then $\min_x \mathcal{L}(x,u,v) = -\infty$.

If $u \geq 0$ then

$$\begin{aligned} \min_x \mathcal{L}(x,u,v) &= \min_{x>0} \mathcal{L}(x,u,v) && (0.9) \\ &= \min_{x>0} (c - A^{\mathrm{T}}v - u)^{\mathrm{T}}x + b^{\mathrm{T}}v - \mu\sum_i \ln x_i \\ &= \begin{cases} -\infty & \text{if } c - A^{\mathrm{T}}v - u \not> 0, \\ (c - A^{\mathrm{T}}v - u)^{\mathrm{T}}x^* + b^{\mathrm{T}}v - \mu\sum_i \ln x_i^* & \text{otherwise,} \end{cases} && (0.10) \end{aligned}$$

where (0.9) holds because when $x \not> 0$, then $\mathcal{L}(x,u,v) = +\infty$ is clearly not optimal, (0.10) holds because if the i-th component of $c - A^{\mathrm{T}}v - u$ is $M_i \leq 0$ then $\lim_{x_i\to\infty} M_i x_i - \mu \ln x_i = -\infty$ ((0.10) follows by setting all other components of x to 1), and where x^* is determined by

$$\nabla x\Big(c - A^{\mathrm{T}}v - u)^{\mathrm{T}}x + b^{\mathrm{T}}v - \mu\sum_i \ln x_i\Big) = 0,$$

that is, $x_i^* = \mu/s_i$ $(i = 1,\dots,n)$, where $s = c - A^{\mathrm{T}}v - u$. Substituting into (0.10), we find

$$(c - A^{\mathrm{T}}v - u)^{\mathrm{T}}x^* + b^{\mathrm{T}}v - \mu\sum_i \ln x_i^* = b^{\mathrm{T}}v - \mu\sum_i \ln s_i + n(\mu - \ln\mu).$$

In summary, we have

$$\min_x \mathcal{L}(x,u,v) = \begin{cases} -\infty & \text{if } u \ngeq 0 \text{ or } s = c - A^{\mathrm{T}}v - u \not> 0, \\ b^{\mathrm{T}}v - \mu\sum_i \ln s_i + n(\mu - \ln\mu) & \text{if } u \geq 0, s = c - A^{\mathrm{T}}v - u > 0. \end{cases}$$

Therefore,

$$\text{(D)} \qquad \max_{(u,v)} \Big(\min_x \mathcal{L}(x,u,v) \Big)$$

is equivalent to

$$
\begin{aligned}
&\max_{(u,v)} b^{\mathrm{T}}v - \mu\sum_i \ln s_i + n(\mu - \ln\mu) \\
&\text{s.t.} \quad A^{\mathrm{T}}v + s = c - u \\
&\qquad\quad s > 0,
\end{aligned}
$$

and since $n(\mu - \ln\mu)$ is just a constant and $\ln s_i$ is maximised when s_i is maximised, that is, when $u_i = 0$ for all i, this problem is equivalent to

$$
\begin{aligned}
&\max_{(y,s)} b^{\mathrm{T}}y - \mu\sum_i \ln s_i \text{s.t.} \qquad\qquad A^{\mathrm{T}}v + s = c \\
&\quad s > 0.
\end{aligned}
$$

(iv) The Slater constraint qualification for (P) is $\exists\, x > 0$ such that $Ax = b$. The SCQ for (D) is $\exists\,(y, s)$ such that $s > 0$ and $A^{\mathrm{T}}y + s = c$.

(v) Since the SCQ holds, (P) has feasible solutions. If an optimal solution exists then it is unique, because the objective function is strictly convex. Furthermore, in this case the KKT conditions are both necessary and sufficient conditions for x to be a minimiser: $\exists\, u, v$ such that

$$
\begin{aligned}
c - \mu\left[\,{\scriptstyle x_1^{-1},\ldots,x_n^{-1}}\,\right]^{\mathrm{T}} - A^{\mathrm{T}}v - u &= 0 \\
Ax &= b \\
x &\geq 0 \\
u_i x_i &= 0.
\end{aligned}
$$

The conditions can only be satisfied when $x > 0$, and then the complementarity equation shows that $u = 0$. Thus, the optimal solution of (P) is characterised by the existence of (y, s) (where $y = v$) such that

$$
\begin{aligned}
Ax &= b \\
A^{\mathrm{T}}y + s &= c \\
Xe &= \mu S^{-1}e \\
x, s &> 0.
\end{aligned}
$$

But $A^{\mathrm{T}}y + s = c$, $s > 0$ is exactly the SCQ for (D).

(vi) Everything works the same way, because a reversal of the sign of the objective function turns (D) into a strictly convex problem. The set of equations characterising the optimal solution is exactly the same as for (P):

$$
\begin{aligned}
Ax &= b \\
A^{\mathrm{T}}y + s &= c \\
Se &= \mu X^{-1}e \\
x, s &> 0.
\end{aligned}
$$