CNAc: Continuous Optimization
Solutions to problem set 1 — optimality conditions

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

Problem 1.

(i) Differentiating $\theta(\alpha) = f(x+\alpha s)$ with respect to α, we obtain $\theta'(\alpha) = s^T g(x+\alpha s)$. Thus the Newton formula gives

$$f(x+s) - f(x) - s^T g(x) = \int_0^1 s^T (g(x+\alpha s) - g(x)) d\alpha.$$

Hence, using the Cauchy-Schwartz inequality and the Lipschitz continuity of $g(x)$, we have

$$|f(x+s) - f(x) - s^T g(x)| \leq \|s\| \int_0^1 \gamma^L(x) \|\alpha s\| d\alpha = \gamma^L(x)\|s\|^2 \int_0^1 \alpha d\alpha = \tfrac{1}{2}\gamma^L(x)\|s\|^2.$$

(ii) Differentiating the product $(1-\alpha)\theta'(\alpha)$ gives

$$\frac{d}{d\alpha}(1-\alpha)\theta'(\alpha) = -\theta'(\alpha) + (1-\alpha)\theta''(\alpha).$$

Integrating from 0 to 1 gives

$$-\theta'(0) = -\int_0^1 \theta'(\alpha) d\alpha + \int_0^1 (1-\alpha)\theta''(\alpha) d\alpha$$

which yields the required integration-by-parts formula using the Newton formula.

A second differentiation $\theta(\alpha)$ gives $\theta''(\alpha) = s^T H(x+\alpha s)s$, and trivially $\int_0^1 (1-\alpha) d\alpha = \frac{1}{2}$. Hence from the integration-by-parts formula, we have

$$f(x+s) - f(x) - s^T g(x) - \tfrac{1}{2} s^T H(x) s = \int_0^1 (1-\alpha) s^T (H(x+\alpha s) - H(x)) s d\alpha.$$

Thus, using the Cauchy-Schwartz inequality and the Lipschitz continuity of $H(x)$, it follows that

$$\begin{aligned} |f(x+s) - f(x) - s^T g(x) - \tfrac{1}{2} s^T H(x) s| \leq & \ \|s\|^2 \textstyle\int_0^1 (1-\alpha)\gamma^Q(x)\|\alpha s\| d\alpha \\ = & \ \gamma^Q(x)\|s\|^3 \textstyle\int_0^1 (1-\alpha)\alpha d\alpha = \tfrac{1}{6}\gamma^Q(x)\|s\|^3 \end{aligned}$$

as required

Problem 2.

(i) We may write ${a_i}^T s = 0$ for $i \in \mathcal{E}$ as ${a_i}^T s \geq 0$ for $i \in \mathcal{E}$ and $-{a_i}^T s \geq 0$ for $i \in \mathcal{E}$. Thus, using Farkas' lemma, $\mathcal{S} \notin \emptyset$ if and only if

$$\begin{aligned} g \in C = & \left\{ \sum_{i \in \mathcal{E}} u_i a_i - \sum_{i \in \mathcal{E}} v_i a_i + \sum_{i \in \mathcal{A}} y_i a_i \mid (u_i, v_i) \geq 0 \text{ for all } i \in \mathcal{E} \text{ and } y_i \geq 0 \text{ for all } i \in \mathcal{A} \right\} \\ = & \left\{ \sum_{i \in \mathcal{E}} z_i a_i + \sum_{i \in \mathcal{A}} y_i a_i \mid y_i \geq 0 \text{ for all } i \in \mathcal{A} \right\} \end{aligned}$$

where the sign of $z_i = u_i - v_i$ is unrestricted.

(ii) Just as in the proofs of Theorems 1.7 and 1.9, consider feasible perturbations $x(\alpha) = x_* + \alpha s + O(\alpha)$ about x_* for the equality ($i \in \mathcal{E}$) and active inequality ($i \in \mathcal{A}$) constraints such that $c_i(x(\alpha)) = 0$ for $i \in \mathcal{E}$ and $c_i(x(\alpha)) \geq 0$ for $i \in \mathcal{A}$. Then, just as in the previous theorems, necessarily

$$s^T a_i(x_*) = 0 \text{ for all } i \in \mathcal{E} \text{ and } s^T a_i(x_*) \geq 0 \text{ for all } i \in \mathcal{A}.$$

But for such perturbations, the objective function will decrease for small α if $s^T g < 0$, so x_* can only be a local minimizer if $\mathcal{S}$ is empty. Then Part (i) gives that

$$g(x_*) = \sum_{i \in \mathcal{E}} z_i a_i(x_*) + \sum_{i \in \mathcal{A}} y_i a_i(x_*) \text{ where } y_i \geq 0 \text{ for all } i \in \mathcal{A}$$

or equivalently

$$g(x_*) = \sum_{i \in \mathcal{E}} z_i a_i(x_*) + \sum_{i \in \mathcal{I}} y_i a_i(x_*) \text{ where } y_i \geq 0 \text{ and } y_i c_i(x_*) = 0 \text{ for all } i \in \mathcal{I}.$$

Of course, we also necessarily have that

$$c_i(x_*) = 0 \text{ for all } i \in \mathcal{E} \text{ and } c_i(x_*) \geq 0 \text{ for all } i \in \mathcal{I}.$$

Problem 3.

(i) The problem might be non-differentiable because small perturbations in x may cause different terms $f_i(x)$ to define the objective $f(x)$. For example, suppose $m = 2$, $f_1(x) = x + 1$ and $f_2(x) = -x + 1$. Then for $x \geq 0$, $f(x) = x + 1$ while for $x \leq 0$, $f(x) = -x + 1$, and there is a derivative discontinuity at $x = 0$. It might also be non-differentiable because of the $|\cdot|$ term. For instance if $m = 1$ and $f_1(x) = x$, $f(x)$ is non-differentiable at $x = 0$.

(ii) Clearly $|f_i(x)| \leq u$ is equivalent to $-u \leq f_i(x) \leq u$. Minimizing the largest $|f_i(x)|$ is equivalent to minimizing the largest upper bound on $|f_i(x)|$.

The constraints $-u \leq f_i(x) \leq u$ may be rewritten as $f_i(x) + u \geq 0$ and $u - f_i(x) \geq 0$. Let y_i^{L} and y_i^{U} (respectively) be Lagrange multipliers for these constraints, and let $A(x)$ be the Jacobian of the vector of f_i.

First-order necessary optimality conditions are that the y^{L} and y^{U} satisfy

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} A(x) \\ e^T \end{pmatrix} y^{\text{L}} - \begin{pmatrix} -A(x) \\ e^T \end{pmatrix} y^{\text{U}} = 0$$

and that

$$(f^{\max} + f_i(x)) y_i^{\text{L}} = 0 \text{ and } (f^{\max} - f_i(x)) y_i^{\text{U}} = 0,$$

where $f^{\max}$ is the optimal objective value. This is to say that

$$\begin{array}{rl} A(x)(y^{\text{L}} - y^{\text{U}}) & = 0 \\ e^T(y^{\text{L}} + y^{\text{U}}) & = 1 \text{ and } (y^{\text{L}}, y^{\text{U}}) \geq 0 \end{array}.$$

If $f^{\max} > 0$ only one of the pair $(y_i^{\text{L}}, y_u^{\text{L}})$ can be nonzero.

Problem 4†.
Clearly, the problem is equivalent to

$$\begin{array}{rl} \min\limits_{x \in \mathbb{R}^2} & f(x) = x_1^2 + x_2^2 \\ \text{such that} & c_1(x) = x_1^2 + (x_2 - 1)^2 - 1 \geq 0, \\ & c_2(x) = -x_1^2 - (x_2 - 2)^2 + 1 \geq 0. \end{array}$$

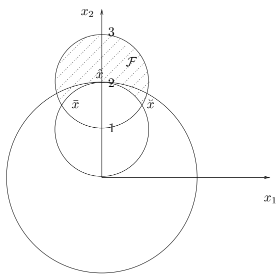

Figure 0.1: The feasible domain $\mathcal{F}$ is shaded.

We have

$$\begin{aligned}
\nabla f(x) &= \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}, \\
\nabla c_1(x) &= \begin{pmatrix} 2x_1 \\ 2(x_2-1) \end{pmatrix}, \\
\nabla c_2(x) &= \begin{pmatrix} -2x_1 \\ -2(x_2-2 \end{pmatrix}, \\
\ell(x,y) &= x_1^2 + x_2^2 - y_1\big(x_1^2 + (x_2-1)^2 - 1\big) - y_2\big(-x_1^2 - (x_2-2)^2 + 1\big), \\
\nabla_x \ell(x,y) &= \begin{pmatrix} 2x_1(1-y_1+y_2) \\ 2x_2 - 2y_1(x_2-1) + 2y_2(x_2-2) \end{pmatrix}.
\end{aligned}$$

The KKT conditions are as follows:

$$2x_1(1-y_1+y_2) = 0 \tag{1}$$
$$2x_2 - 2y_1(x_2-1) + 2y_2(x_2-2) = 0 \tag{2}$$
$$x_1^2 + (x_2-1)^2 - 1 \geq 0 \tag{3}$$
$$-x_1^2 - (x_2-2)^2 + 1 \geq 0 \tag{4}$$
$$y_1\big(x_1^2 + (x_2-1) - 1\big) = 0 \tag{5}$$
$$y_2\big(-x_1^2 - (x_2-2)^2 + 1\big) = 0 \tag{6}$$
$$y_1 \geq 0 \tag{7}$$
$$y_2 \geq 0. \tag{8}$$

Let us find all the KKT points. We need to distinguish four cases:

(a) If $\mathcal{A}(x) = \emptyset$ then (1),(2) imply $x = 0$, which violates (4). Thus, there are no KKT points that correspond to $\mathcal{A}(x) = \emptyset$.

(b) If $\mathcal{A} = \{2\}$ then $y_1 = 0$. (1),(2) and (4) imply

$$2x_1(1+y_2) = 0 \tag{9}$$
$$2x_2 + 2y_2(x_2-2) = 0 \tag{10}$$
$$x_1^2 + (x_2-2)^2 = 1. \tag{11}$$

(9) implies that either $x_1 = 0$ or $y_2 = -1$. The second case contradicts (8), so we may assume that the first case holds. But then (11) implies $x_2 \in \{1,3\}$. If $x_2 = 1$ then $\mathcal{A}(x) = \{1,2\}$, which contradicts our earlier assumption. Thus, we must have $x_2 = 3$. But then (10) implies $y_2 = -3$ which contradicts (8). Thus, there are no KKT points corresponding to $\mathcal{A} = \{2\}$.

(c) If $\mathcal{A}(x) = \{1\}$ then $y_2 = 0$. (1)– (3) become

$$\begin{aligned} 2x_1(1-y_1) &= 0, && (12) \\ 2x_2 - 2y_1(x_2-1) &= 0, && (13) \\ x_1^2 + (x_2-1)^2 &= 1. && (14) \end{aligned}$$

The unique solution of these equations is

$$\hat{x} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \qquad \hat{y} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

It is easily checked that $(\hat{x}, \hat{y})$ satisfies (1)–(8) and hence is a KKT point. Moreover, the LICQ holds at $\hat{x}$ because $\nabla c_1(\hat{x}) = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \neq 0$.

(d) If $\mathcal{A}(x) = \{1, 2\}$, then (3) and (4) must hold at equality, that is,

$$\begin{aligned} x_1^2 + (x_2-1)^2 - 1 &= 0, \\ x_1^2 + (x_2-2)^2 - 1 &= 0. \end{aligned}$$

This system of equations implies $x_2 = 3/2$, $x_1 = \pm\sqrt{3}/2$. Let us analyse the case $\breve{x} = \begin{pmatrix} 3/2 \\ \sqrt{3}/2 \end{pmatrix}$ only, as the two cases are similar. (1),(2) imply

$$\begin{aligned} \sqrt{3}(y_2 + 1 - y_1) &= 0, \\ 3 - y_1 - y_2 &= 0, \end{aligned}$$

which implies $\breve{y} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. It is easily checked that $(\breve{x}, \breve{y})$ satisfies (1)–(8) and hence is a KKT point. Likewise, $(\bar{x}, \bar{y})$ is a KKT point where $\bar{x} = \begin{pmatrix} 3/2 \\ \sqrt{3}/2 \end{pmatrix}$ and $\bar{y} = \breve{y}$. Furthermore, the LICQ holds at both points because $\nabla c_1(\breve{x}) = \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix}$ and $\nabla c_2(\breve{x}) = \begin{pmatrix} -\sqrt{3} \\ 1 \end{pmatrix}$ are linearly independent, and likewise for $\nabla c_1(\bar{x}) = \begin{pmatrix} -\sqrt{3} \\ 1 \end{pmatrix}$ and $\nabla c_2(\bar{x}) = \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix}$.

In summary, we have found three KKT points. It would be easy to evaluate f at all three points to find that $\breve{x}$ and $\bar{x}$ are global minimizers of our problem. It can also be seen by inspection that $\hat{x}$ is not a local minimizer. But as we now show, this information can also be derived from second order information:

Since the LICQ holds at $\hat{x}$, the first-order feasible directions from $\hat{x}$ are characterised by

$$\begin{aligned} s &\neq 0, \\ s^T \nabla c_1(\hat{x}) &\geq 0, \end{aligned}$$

which is equivalent to

$$\begin{aligned} s_2 &\geq 0, \\ s_1^2 + s_2^2 &> 0. \end{aligned}$$

If $s_2 > 0$ then $\hat{y}_1 s^T \nabla c_1(\hat{x}) = 4s_2 > 0$, so we don't need to check any additional conditions for such directions s. However, if $s_2 = 0$ then $\hat{y}_1 s^T \nabla c_1(\hat{x}) = 0$, and since $\mathcal{A}(\hat{x}) = \{1\}$, this shows that s is a first-order feasible direction for which the second order necessary optimality condition

$$s^T D_{xx}\ell(\hat{x}, \hat{y}) s \geq 0$$

has to be satisfied. But note that this condition is violated, because

$$s^T D_{xx}\ell(\hat{x},\hat{y})s = s_1^2 \frac{\partial^2}{\partial x_1^2}\ell(\hat{x},\hat{y}) = -2s_1^2 < 0.$$

Since $\hat{x}$ fails to satisfy the second order necessary optimality conditions, it cannot be a local minimizer of our problem.

Now for $\breve{x}$, where the LICQ holds, the set of first-order feasible directions is characterised by

$$\begin{aligned} s &\neq 0, \\ s^T\nabla c_1(\breve{x}) &\geq 0, \\ s^T\nabla c_2(\breve{x}) &\geq 0, \end{aligned}$$

which is equivalent to

$$\begin{aligned} |s_1| &\leq \tfrac{d_2}{\sqrt{3}}, \\ s_2 &> 0. \end{aligned}$$

But for any s that satisfies these conditions we have

$$\breve{y}_1 s^T\nabla c_1(\breve{x}) = 2(\sqrt{3}s_1 + s_2) > 0.$$

Thus, the set of feasible directions that lie in the set $\mathcal{S}$ in the notes for Part 1 of the course is the empty set. This shows that the sufficient optimality conditions are satisfied at $\breve{x}$, and that this must be a strict local minimizer. Likewise, one finds that the sufficient optimality conditions hold at $\bar{x}$.

Problem 5†.

(i) The objective function is unbounded along the line $x_2 = 0$, $x_1 \to \infty$. Thus, no global solution exists, but we can find a local minimum with the method of Lagrange multipliers.

(ii) We get

$$\nabla_x\ell(x,y) = \begin{pmatrix} -0.2(x_1-4) - 2yx_1 \\ 2x_2 - 2yx_2 \end{pmatrix}, \quad \nabla_{xx}\ell(x,y) = \begin{pmatrix} -0.2-2y & 0 \\ 0 & 2-2y \end{pmatrix}.$$

The KKT conditions are

$$\begin{aligned} -0.2(x_1-4) - 2yx_1 &= 0, \\ 2x_2 - 2yx_2 &= 0, \\ x_1^2 + x_2^2 - 1 &\geq 0, \\ y(x_1^2+x_2^2-1) &= 0, \\ y &\geq 0. \end{aligned}$$

(iii) For $\mathcal{A}(x) = \emptyset$ have $y = 0$ and hence, $x_1 = 4$, $x_2 = 0$. This satisfies the constraint, and thus $x^* = (4,0)$ and $y^* = 0$ is a KKT point.

If $\mathcal{A} = \{1\}$ then either $x_2 = 0$ or $y = 1$ (or both). In the former case the active constraint implies $x_1 = 1$, which corresponds to $y = 0.3$ or $x = -0.5$ which gives $y = -1$. The latter does not satisfy $y \geq 0$ but the former does, so $x^* = (1,0)$ and $y^* = 0.3$ is another KKT point. Finally, if $y = 1$, the first term of the gradient gives $x_1 = 4/11$, and thus $x_2 = \sqrt{105}/11$, giving a third KKT point $x^* = (4/11, \sqrt{105}/11)$ and $y^* = 1$.

(iv) The LICQ holds at x^* because $\nabla c_1(x^*) = 2\begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} \neq 0$ unless $x_* = 0$.

(v) For the first KKT point, the constraint is inactive and the Hessian of the Lagrangian is indefinite, so this point does not satisfy the 2nd-order necessary optimality conditions, and thus cannot be a local minimizer (or maximizer).

The set $\mathcal{S}$ of first-order feasible directions from the second KKT point x^* is

$$\{s \in \mathbb{R}^2 : s_1 = 0, s_2 \neq 0\}.$$

For any s from that set we have

$$s^T \nabla_{xx}\ell(x^*, y^*)s = \begin{pmatrix} 0 & s_2 \end{pmatrix} \begin{pmatrix} -0.4 & 0 \\ 0 & 1.4 \end{pmatrix} \begin{pmatrix} 0 \\ s_2 \end{pmatrix} = 1.4s_2^2 > 0.$$

Therefore, the sufficient optimality conditions are satisfied and x^* is a strict local minimizer.

Finally, for the third KKT point, the set of first-order feasible directions is

$$\{s \in \mathbb{R}^2 : 4/11s_1 + \sqrt{105}/11s_2 = 0\} = \{s \in \mathbb{R}^2 : s = \sigma(\sqrt{105}, -4) \text{ for all } \sigma\}$$

while the Hessian of the Lagrangian is

$$\nabla_{xx}\ell(x^*, y^*) = \begin{pmatrix} -2.2 & 0 \\ 0 & 0 \end{pmatrix}.$$

But then $s^T \nabla_{xx}\ell(x^*, y^*)s = -231\sigma^2 < 0$ and so this point does not satisfy the 2nd-order sufficient optimality conditions—it is a maximizer.

Problem 6†. The problem is of course trivial to solve directly, but we want to see how the Lagrange multiplier approach solves the problem "blindly". The problem is equivalent to solving

$$\min\ x^T x \ \text{ such that } \ g(x) = a^T x + b \geq 0.$$

We may assume that $a \neq 0$; otherwise the problem is trivial. The Lagrangian of this problem is

$$\ell(x, y) = x^T x - y(a^T x + b).$$

The gradient $\nabla g(x) = a$ is nonzero everywhere, and hence the LICQ holds at all feasible points. The KKT conditions are

$$\begin{aligned} x - ya &= 0, && (15) \\ y(a^T x + b) &= 0, && (16) \\ a^T x + b &\geq 0, && (17) \\ y &\geq 0. && (18) \end{aligned}$$

If $y = 0$ then $x = 0$, and then (17) implies $b \geq 0$. Either this is true, and then $(x, y) = (0, 0)$ satisfies the KKT conditions, or else $b < 0$ and then $y = 0$ is not a viable choice.
If $y > 0$, then $x = ya \neq 0$, $a^T x + b = y\|a\|^2 + b = 0$, and then $b < 0$, which is either true, in which case $(x, y) = \big((-b/\|a\|^2)a, -b/\|a\|^2\big)$ satisfies the KKT conditions, or else $y > 0$ is not a viable choice.
Thus, we have found that both in the case $b \geq 0$ and $b < 0$ there is exactly one point satisfying the KKT conditions, and since the KKT conditions must hold at the minimum of our optimisation problem, the resulting points must be local minimizers. Since the problem is convex these are also the global minimizers in both cases.

† Thanks to Raphael Hauser for these solutions.