SECTION C: CONTINUOUS OPTIMISATION
LECTURE 11: THE METHOD OF LAGRANGE MULTIPLIERS

HONOUR SCHOOL OF MATHEMATICS, OXFORD UNIVERSITY
HILARY TERM 2005, DR RAPHAEL HAUSER

1. Examples. In this lecture we will take a closer look at some examples and illustrate how optimality conditions are applied in the Method of Lagrange multipliers.

EXAMPLE 1.1. *Use the method of Lagrange multipliers to solve the problem*

$$\min_{x\in\mathbb{R}^2} \|x\| \tag{1.1}$$

$$\begin{aligned} s.t.\quad & \left\|x - \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right\| \geq 1, \\ & \left\|x - \left[\begin{smallmatrix}0\\2\end{smallmatrix}\right]\right\| \leq 1. \end{aligned} \tag{1.2}$$

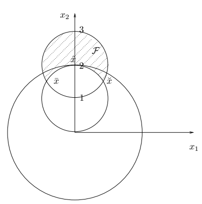

FIG. 1.1. *The feasible domain $\mathcal{F}$ is shaded.*

Clearly, the problem is equivalent to

$$\begin{aligned} &\min_{x\in\mathbb{R}^2} f(x) = x_1^2 + x_2^2 \\ &\text{s.t.}\quad g_1(x) = x_1^2 + (x_2-1)^2 - 1 \geq 0, \\ &\qquad\ g_2(x) = -x_1^2 - (x_2-2)^2 + 1 \geq 0. \end{aligned}$$

We have

$$\begin{aligned} &\nabla f(x) = \left[\begin{smallmatrix}2x_1\\2x_2\end{smallmatrix}\right], \\ &\nabla g_1(x) = \left[\begin{smallmatrix}2x_1\\2(x_2-1)\end{smallmatrix}\right], \\ &\nabla g_2(x) = \left[\begin{smallmatrix}-2x_1\\-2(x_2-2)\end{smallmatrix}\right], \\ &\mathcal{L}(x,\lambda) = x_1^2 + x_2^2 - \lambda_1\big(x_1^2 + (x_2-1)^2 - 1\big) - \lambda_2\big(-x_1^2 - (x_2-2)^2 + 1\big), \\ &\nabla_x \mathcal{L}(x,\lambda) = \begin{bmatrix} 2x_1(1-\lambda_1+\lambda_2) \\ 2x_2 - 2\lambda_1(x_2-1) + 2\lambda_2(x_2-2) \end{bmatrix}. \end{aligned}$$

The KKT conditions are the following:

$$2x_1(1-\lambda_1+\lambda_2) = 0 \tag{1.3}$$
$$2x_2 - 2\lambda_1(x_2-1) + 2\lambda_2(x_2-2) = 0 \tag{1.4}$$
$$x_1^2 + (x_2-1)^2 - 1 \geq 0 \tag{1.5}$$
$$-x_1^2 - (x_2-2)^2 + 1 \geq 0 \tag{1.6}$$
$$\lambda_1\big(x_1^2 + (x_2-1) - 1\big) = 0 \tag{1.7}$$
$$\lambda_2\big(-x_1^2 - (x_2-2)^2 + 1\big) = 0 \tag{1.8}$$
$$\lambda_1 \geq 0 \tag{1.9}$$
$$\lambda_2 \geq 0. \tag{1.10}$$

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

If $\mathcal{A}(x) = \emptyset$ then (1.3),(1.4) imply $x = 0$, which violates (1.6). Thus, there are no KKT points that correspond to $\mathcal{A}(x) = \emptyset$.

If $\mathcal{A} = \{2\}$ then $\lambda_1 = 0$. (1.3),(1.4) and (1.6) imply

$$2x_1(1+\lambda_2) = 0 \tag{1.11}$$
$$2x_2 + 2\lambda_2(x_2-2) = 0 \tag{1.12}$$
$$x_1^2 + (x_2-2)^2 = 1. \tag{1.13}$$

(1.11) implies that either $x_1 = 0$ or $\lambda_2 = -1$. The second case contradicts (1.10), so we may assume that the first case holds. But then (1.13) 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 (1.12) implies $\lambda_2 = -3$ which contradicts (1.10). Thus, there are no KKT points corresponding to $\mathcal{A} = \{2\}$.

If $\mathcal{A}(x) = \{1\}$ then $\lambda_2 = 0$. (1.3)– (1.5) become

$$2x_1(1-\lambda_1) = 0, \tag{1.14}$$
$$2x_2 - 2\lambda_1(x_2-1) = 0, \tag{1.15}$$
$$x_1^2 + (x_2-1)^2 = 1. \tag{1.16}$$

The unique solution of these equations is

$$\hat{x} = \left[\begin{smallmatrix}0\\2\end{smallmatrix}\right], \qquad \hat{\lambda} = \left[\begin{smallmatrix}2\\0\end{smallmatrix}\right].$$

It is easily checked that $(\hat{x}, \hat{\lambda})$ satisfies (1.3)–(1.10) and hence is a KKT point. Moreover, the LICQ holds at $\hat{x}$ because $\nabla g_1(\hat{x}) = \left[\begin{smallmatrix}0\\2\end{smallmatrix}\right] \neq 0$.

If $\mathcal{A}(x) = \{1,2\}$, then (1.5) and (1.6) 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 $\breve{x} = \left[\begin{smallmatrix}3/2\\ \sqrt{3}/2\end{smallmatrix}\right]$ only, as the two cases are similar. (1.3),(1.4) imply

$$\begin{aligned} \sqrt{3}(\lambda_2 + 1 - \lambda_1) = 0, \\ 3 - \lambda_1 - \lambda_2 = 0, \end{aligned}$$

which implies $\check{\lambda} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$. It is easily checked that $(\check{x}, \check{\lambda})$ satisfies (1.3)–(1.10) and hence is a KKT point. Likewise, $(\bar{x}, \bar{\lambda})$ is a KKT point where $\bar{x} = \left[\begin{smallmatrix} 3/2 \\ \sqrt{3}/2 \end{smallmatrix}\right]$ and $\bar{\lambda} = \check{\lambda}$. Furthermore, the LICQ holds at both points because $\nabla g_1(\check{x}) = \left[\begin{smallmatrix} \sqrt{3} \\ 1 \end{smallmatrix}\right]$ and $\nabla g_2(\check{x}) = \left[\begin{smallmatrix} -\sqrt{3} \\ 1 \end{smallmatrix}\right]$ are linearly independent, and likewise for $\nabla g_1(\bar{x}) = \left[\begin{smallmatrix} -\sqrt{3} \\ 1 \end{smallmatrix}\right]$ and $\nabla g_2(\bar{x}) = \left[\begin{smallmatrix} \sqrt{3} \\ 1 \end{smallmatrix}\right]$.

All in all we have found three KKT points. It would be easy to evaluate f at all three points to find that $\check{x}$ and $\bar{x}$ are global minimisers of (1.1). It can also be seen by inspection that $\hat{x}$ is not a local minimiser. Let us now show that this information can also be derived from second order information:

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

$$\begin{aligned} d &\neq 0, \\ d^{\mathrm{T}} \nabla g_1(\hat{x}) &\geq 0, \end{aligned}$$

which is equivalent to

$$\begin{aligned} d_2 &\geq 0, \\ d_1^2 + d_2^2 &> 0. \end{aligned}$$

If $d_2 > 0$ then $\hat{\lambda}_1 d^{\mathrm{T}} \nabla g_1(\hat{x}) = 4d_2 > 0$, so we don't need to check any additional conditions for such directions d. However, if $d_2 = 0$ then $\hat{\lambda}_1 d^{\mathrm{T}} \nabla g_1(\hat{x}) = 0$, and since $\mathcal{A}(\hat{x}) = \{1\}$, this shows that d is a feasible exit direction for which the second order necessary optimality condition

$$d^{\mathrm{T}} D_{xx} \mathcal{L}(\hat{x}, \hat{\lambda}) d \geq 0$$

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

$$d^{\mathrm{T}} D_{xx} \mathcal{L}(\hat{x}, \hat{\lambda}) d = d_1^2 \frac{\partial^2}{\partial x_1^2} \mathcal{L}(\hat{x}, \hat{\lambda}) = -2d_1^2 < 0.$$

Since $\hat{x}$ fails to satisfy the second order necessary optimality conditions, it cannot be a local minimiser of (1.1).

Now for $\check{x}$, where the LICQ holds, the set of feasible exit directions is characterised by

$$\begin{aligned} d &\neq 0, \\ d^{\mathrm{T}} \nabla g_1(\check{x}) &\geq 0, \\ d^{\mathrm{T}} \nabla g_2(\check{x}) &\geq 0, \end{aligned}$$

which is equivalent to

$$\begin{aligned} |d_1| &\leq \frac{d_2}{\sqrt{3}}, \\ d_2 &> 0. \end{aligned}$$

But for any d that satisfies these conditions we have

$$\check{\lambda}_1 d^{\mathrm{T}} \nabla g_1(\check{x}) = 2(\sqrt{3} d_1 + d_2) > 0.$$

Thus, the set of feasible exit directions that satisfy Condition (1.7) from Lecture 10 is the empty set. This shows that the sufficient optimality conditions are satisfied at $\check{x}$, and that this must be a strict local minimiser. Likewise, one finds that the sufficient optimality conditions hold at $\bar{x}$.

Example 1.2. *Consider the minimisation problem*

$$\begin{aligned} &\min \ -0.1(x_1 - 4)^2 + x_2^2 \\ &\textit{s.t.} \quad x_1^2 + x_2^2 - 1 \geq 0. \end{aligned}$$

(i) Does this problem have a global minimiser?
(ii) Set up the KKT conditions for this problem.
(iii) Find x^* and a vector λ^* of Lagrange multipliers so that (x^*, λ^*) satisfy the KKT conditions.
(iv) Is the LICQ satisfied at x^*?
(v) Characterise the set of feasible exit directions from x^*.
(vi) Check that the sufficient optimality conditions hold at x^* to show that x^* is a local minimiser.

(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

$$\nabla_x \mathcal{L}(x, \lambda) = \begin{pmatrix} -0.2(x_1-4)-2\lambda x_1 \\ 2x_2-2\lambda x_2 \end{pmatrix}, \qquad \nabla_{xx} \mathcal{L}(x, \lambda) = \begin{pmatrix} -0.2-2\lambda & 0 \\ 0 & 2-2\lambda \end{pmatrix}.$$

The KKT conditions are

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

(iii) For $\mathcal{A}(x) = \emptyset$ have $\lambda = 0$ and hence, $x_1 = 4$, $x_2 = 0$, which implies $\mathcal{A} = \{1\}$, contrary to our assumption. Thus, there are no KKT points corresponding to this case. If $\mathcal{A} = \{1\}$ then the unique solution is $x^* = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, $\lambda_1^* = 0.3$.

(iv) The LICQ holds at x^* because $\nabla g_1(x^*) = \left[\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right] \neq 0$.

(v) The set of feasible exit directions from x^* for which Condition (1.7) from Lecture 10 holds is

$$\{d \in \mathbb{R}^2 : \ d_1 = 0, d_2 \neq 0\}.$$

(vi) For any d from that set we have

$$d^{\mathrm{T}} \nabla_{xx} \mathcal{L}(x^*, \lambda^*) d = \begin{pmatrix} 0 & d_2 \end{pmatrix} \begin{pmatrix} -0.4 & 0 \\ 0 & 1.4 \end{pmatrix} \begin{pmatrix} 0 \\ d_2 \end{pmatrix} = 1.4 d_2^2 > 0.$$

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

EXAMPLE 1.3. *Consider the half space defined by* $H = \{x \in \mathbb{R}^n : a^{\mathrm{T}}x + b \geq 0\}$ *where* $a \in \mathbb{R}^n$ *and* $b \in \mathbb{R}$ *are given. Formulate and solve the optimisation problem of finding the point* x *in* H *that has the smallest Euclidean norm.*

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

$$\begin{aligned} &\min x^{\mathrm{T}}x \\ &\text{s.t.} \quad g(x) = a^{\mathrm{T}}x + b \geq 0. \end{aligned}$$

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

$$\mathcal{L}(x, \lambda) = x^{\mathrm{T}}x - \lambda(a^{\mathrm{T}}x + b).$$

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

$$x - \lambda a = 0, \tag{1.17}$$

$$\lambda(a^{\mathrm{T}}x + b) = 0, \tag{1.18}$$

$$a^{\mathrm{T}}x + b \geq 0, \tag{1.19}$$

$$\lambda \geq 0. \tag{1.20}$$

If $\lambda = 0$ then $x = 0$, and then (1.19) implies $b \geq 0$. Either this is true, and then $(x, \lambda) = (0, 0)$ satisfies the KKT conditions, or else $b < 0$ and then $\lambda = 0$ is not a viable choice.

If $\lambda > 0$, then $x = \lambda a \neq 0$, $a^{\mathrm{T}}x + b = \lambda\|a\|^2 + b = 0$, and then $b < 0$, which is either true, in which case $(x, \lambda) = \big((-b/\|a\|^2)a, -b/\|a\|^2\big)$ satisfies the KKT conditions, or else $\lambda > 0$ 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 minimisers. Since the problem is convex these are also the global minimisers in both cases.

Our last example illustrates that even in the case with only equality constraints the method of Lagrange multipliers is the right tool to solve constrained optimisation problems: eliminating some of the variables can lead to the wrong result.

EXAMPLE 1.4. *Solve the problem*

$$\begin{aligned} &\min_{x \in \mathbb{R}^2} x_1 + x_2 \\ &s.t. \quad x_1^2 + x_2^2 = 1 \end{aligned}$$

by eliminating the variable x_2. *Show that the choice of sign for a square root operation during the elimination process is critical; the wrong choice leads to an incorrect*

answer.

A sketch reveals that $(-1/\sqrt{2}, -1/\sqrt{2})$ is the unique local minimiser. Let us eliminate $x_2 = -\sqrt{1 - x_1^2}$ and rename x_1 as x. The problem becomes $\min_{x \in \mathbb{R}} f(x) = x - \sqrt{1 - x^2}$. Then $f'(x) = 1 - x/\sqrt{1 - x^2}$, and $f'(x) = 0$ for $x = \sqrt{1 - x^2}$, which is satisfied for $x = 1/\sqrt{2}$, leading to the solution $(x_1, x_2) = (1/\sqrt{2}, -1/\sqrt{2})$, which is neither a maximum nor a minimum. The elimination $x_2 = \sqrt{1 - x_1^2}$ however leads to the solutions $(x_1, x_2) = (-1/\sqrt{2}, -1/\sqrt{2})$ and $(x_1, x_2) = (1/\sqrt{2}, 1/\sqrt{2})$, both of which are true stationary points, one of them a minimiser, the other a maximiser.