EC9A0: Pre-sessional Advanced Mathematics Course

Constrained Optimisation: Equality Constraints

By Pablo F. Beker[1]

1 Introduction

Maintain the assumptions that $D \subseteq \mathbb{R}^K$, K finite, is open, and that $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}^J$, with $J \leq K$.

Suppose that one wants to solve the problem

$$\max_{x \in D} f(x) : g(x) = 0, \tag{1}$$

which means, in the previous notation, that one wants to find $\max_{\{x \in D | g(x)=0\}} f$. The method that is usually applied in economics consists of the following steps:
(1) Defining the Lagrangean function $\mathcal{L} : D \times \mathbb{R}^J \to \mathbb{R}$, by $\mathcal{L}(x, \lambda) = f(x) + \lambda \cdot g(x)$; and
(2) Finding $(x^*, \lambda^*) \in D \times \mathbb{R}^J$ such that $D\mathcal{L}(x^*, \lambda^*) = 0$.
That is, a recipe is applied as though there is a "Theorem" that states the following:

> *Let f and g be differentiable. $x^* \in D$ solves Problem (1) if and only if there exists $\lambda^* \in \mathbb{R}^J$ such that $Df(x^*) + \lambda^{*\top} Dg(x^*) = 0$.*

Unfortunately, though, *such a statement is not true*, as the following example shows:

Example 1. *Suppose $f(x_1, x_2) = x_1 x_2$ and $g(x_1, x_2) = (1 - x_1 - x_2)^3$. Clearly the set of solutions of $\max_{x \in \mathbb{R}^2} f(x)$ s.t. $g(x) = 0$ coincide with the set of solutions to $\max_{x \in \mathbb{R}^2_+} f(x)$ s.t. $g(x) = 0$. Since the second problem consists in maximising a continuous function on a nonempty compact set, it has a solution by Weierstrass Theorem. It is not difficult to see that the unique maximiser is $(x_1^*, x_2^*) = (\frac{1}{2}, \frac{1}{2})$. Then, according to the "theorem" one should be able to find λ^* such that $(x_1^*, x_2^*, \lambda^*)$ solves the following system of equations:*

$$
\begin{aligned}
&(a) \qquad \frac{\partial \mathcal{L}}{\partial x_1} = 0 \iff x_2 - 3\lambda(1 - x_1 - x_2)^2 = 0 \\
&(b) \qquad \frac{\partial \mathcal{L}}{\partial x_2} = 0 \iff x_2 - 3\lambda(1 - x_1 - x_2)^2 = 0 \\
&(c) \qquad \frac{\partial \mathcal{L}}{\partial \lambda} = 0 \iff (1 - x_1 - x_2)^3 = 0
\end{aligned}
$$

However, it is easy to see that a solution to this system of equations does not exist. Indeed, equation (c) implies that at any solution it must be the case that $x_1 + x_2 = 1$ but then equation (a) and (b) imply that both x_1 and x_2 are zero, a contradiction.

This example illustrates that the *only if* part of the previous "theorem" need not be true. That is, without further qualifications Lagrange multipliers may fail to exists even though the maximum is well defined. Now we study the reasons why the"theorem" fails.

[1]Based on notes by Andrés Carvajal

2 Lagrange Theorem

2.1 Intuititve Argument

For simplicity of presentation, suppose that $D = \mathbb{R}^2$ and $J = 1$, and denote the typical element of $\mathbb{R}^2$ by (x, y). So, given $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$, we want to find

$$\max_{(x,y)\in\mathbb{R}^2} f(x, y) : g(x, y) = 0.$$

Let us suppose that we do not know the Lagrangean method, but are quite familiar with unconstrained optimization. A "crude" method suggests the following:

(1) Suppose that we can solve from the equation $g(x, y) = 0$, to express y as a function of x: we find a function $y : \mathbb{R} \to \mathbb{R}$ such that $g(x, y) = 0$ if, and only if $y = h(x)$.

(2) With the function y at hand, we study the unconstrained problem $\max_{x\in\mathbb{R}} F(x)$, where $F : \mathbb{R} \to \mathbb{R}$ is defined by $F(x) = f(x, h(x))$.

(3) Since we want to use calculus, if f and g are differentiable, we need to figure out function h'. Now, if $g(x, h(x)) = 0$, then, differentiating both sides, we get that $\partial_x g(x, h(x)) + \partial_y g(x, h(x))h'(x) = 0$, from where

$$h'(x) = -\frac{\partial_x g(x, h(x))}{\partial_y g(x, h(x))}.$$

(4) Now, with F differentiable, we know that x^* solves $\max_{x\in\mathbb{R}} F(x)$ locally, only if $F'(x^*) = 0$. In our case, the last condition is simply that

$$\partial_x f(x^*, h(x^*)) + \partial_y f(x^*, h(x^*))h'(x^*) = 0,$$

or, equivalently,

$$\partial_x f(x^*, h(x^*)) - \partial_y f(x^*, h(x^*))\frac{\partial_x g(x^*, h(x^*))}{\partial_y g(x^*, h(x^*))} = 0.$$

So, if we define $y^* = h(x^*)$ and

$$\lambda^* = -\frac{\partial_y f(x^*, y^*)}{\partial_y g(x^*, y^*)} \in \mathbb{R},$$

we get that y^*

$$\partial_x f(x^*, y^*) + \lambda^* \partial_x g(x^*, y^*) = 0,$$

whereas

$$\partial_y f(x^*, y^*) + \lambda^* \partial_y g(x^*, y^*) = 0.$$

Then, our method has apparently shown that:

> *Let f and g be differentiable. $x^* \in D$* locally *solves the Problem* (**??**),[2] only if *there exists $\lambda^* \in \mathbb{R}^J$ such that $Df(x^*) + \lambda^{*\top} Dg(x^*) = 0$.*

[2] That is, there is $\varepsilon > 0$ such that $f(x) \leq f(x^*)$ for all $x \in B_\epsilon(x^*) \cap \{x \in D | g(x) = 0\}$.

The latter means that:

(i) The differential approach, as in the unrestricted case, only finds local extrema; and
(ii) The Lagrangean condition is only necessary and not sufficient by itself.

So, we need to determine under what conditions can we find the function $h(y)$ and, moreover, be sure that it is differentiable. Also, we need to be careful and study further conditions for sufficiency. This is what is covered in the next two sections.

2.2 Existence of $h(y)$

Notice that it has been crucial throughout our analysis that $\partial_y g(x^*, y^*) \neq 0$. Of course, even if the latter hadn't been true, but $\partial_x g(x^*, y^*) \neq 0$, our method would still have worked, mutatis mutandis. So, what we actually require is that $Dg(x^*, y^*)$ have rank 1, its maximum possible. The obvious question is: is this a general result, or does it only work in our simplified case?

To see that it is indeed a general result, we introduce without proof the following important result:

THEOREM 1 (The Implicit Function Theorem). *Let $D \subseteq \mathbb{R}^K$ and let $g : D \to \mathbb{R}^J \in \mathbf{C}^1$, with $J \leq K$. If $y^* \in \mathbb{R}^J$ and $(x^*, y^*) \in D$ is such that* $\text{rank}(D_y g(x^*, y^*)) = J$, *then there exist $\varepsilon, \delta > 0$ and $h : B_\varepsilon(x^*) \to B_\delta(y^*) \in \mathbf{C}^1$ such that:*

1. *for every $x \in B_\varepsilon(x^*)$, $(x, h(x)) \in D$;*
2. *for every $x \in B_\varepsilon(x^*)$, $g(x, y) = g(x^*, y^*)$ for $y \in B_\delta(y^*)$ if, and only if $y = h(x)$;*
3. *for every $x \in B_\varepsilon(x^*)$, $Dh(x) = -D_y g(x, h(x))^{-1} D_x g(x, h(x))$.*

This important theorem allows us to express y as a function of x and gives us the derivative of this function: exactly what we wanted! Of course, we need to satisfy the hypotheses of the theorem if we are to invoke it. In particular, the condition on the rank is known as "constraint qualification" and is crucial for the Lagrangean method to work (albeit it is oftentimes forgotten!).

So, in summary, we have argued that:

> *Suppose that $f, g \in \mathbf{C}^1$ and rank $D_y g(x) = J$. If $x^* \in D$ locally solves Problem* (**??**) then *there exists $\lambda^* \in \mathbb{R}^J$ such that $D\mathcal{L}(x^*, \lambda^*) = 0$.*

2.3 Necessary Second Order Conditions

For necessary second order conditions, we can again appeal to our crude method and use the results we inherit from unconstrained optimisation. Since we now need F to be twice differentiable,we must assume that so are f and g, and moreover, we need to know $h''(x)$. Since we already know $h'(x)$, by differentiation,

$$\begin{aligned} h''(x) &= -\frac{\partial}{\partial x}\left(\frac{\partial_x g(x, h(x))}{\partial_y g(x, h(x))}\right) \\ &= -\frac{1}{\partial_y g(x, h(x))}\begin{pmatrix} 1 & h'(x) \end{pmatrix} D^2 g(x, h(x)) \begin{pmatrix} 1 \\ h'(x) \end{pmatrix} \end{aligned}$$

Now, the condition that $F''(x^*) \leq 0$ is equivalent, by substitution,[3] to the requirement that

$$\left(1 \quad h'(x^*) \right) D^2_{(x,y)} \mathcal{L}(x^*, y^*, \lambda^*) \begin{pmatrix} 1 \\ h'(x^*) \end{pmatrix} \leq 0.$$

Obviously, this condition is satisfied if $D^2_{(x,y)} \mathcal{L}(x^*, y^*, \lambda^*)$ is negative semi-definite, but this would be overkill: notice that

$$\left(1 \quad h'(x^*) \right) \cdot Dg(x^*, y^*) = 0,$$

so it suffices that we guarantee that for every $\Delta \in \mathbb{R}^2 \setminus \{0\}$ such that $\Delta \cdot Dg(x^*, y^*) = 0$ we have that $\Delta^\top D^2_{(x,y)} \mathcal{L}(x^*, y^*, \lambda^*) \Delta \leq 0$.

So, in summary, we have argued that:

> *Suppose that $f, g \in \mathbf{C}^1$ and rank $D_y g(x) = J$. $x^* \in D$ locally solves Problem (1),* only if $\Delta^\top D^2_{(x,y)} \mathcal{L}(x^*, y^*, \lambda^*) \Delta \leq 0$ *for all* $\Delta \in \mathbb{R}^2 \setminus \{0\}$ *such that* $\Delta \cdot Dg(x^*, y^*) = 0$.

2.4 Theorem: Necessary Conditions

So, finally, we obtain the following Theorem:

Theorem 2 (Lagrange - FONC and SONC). *Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}^J$ be $\mathbf{C}^1$, with $J \leq K$. Let $x^* \in D$ be such that* $\operatorname{rank}(Dg(x^*)) = J$. *If x^* locally solves Problem (1), then there exists $\lambda^* \in \mathbb{R}^J$ such that*

1. $D\mathcal{L}(x^*, \lambda^*) = 0$.
2. $\Delta^\top D^2_{x,x} \mathcal{L}(x^*, \lambda^*) \Delta \leq 0$ *for all $\Delta \in \mathbb{R}^J \setminus \{0\}$ satisfying $\Delta \cdot Dg(x^*) = 0$;*

2.5 Theorem: Sufficient Conditions

Now we argue that without further qualifications the existence of $(x^*, \lambda^*) \in \mathbb{R}^K \times \mathbb{R}^J$ such that

$$Df(x^*) + \lambda^{*\top} Dg(x^*) = 0.$$

might not be sufficient for x^* to be a local maximiser of Problem (1).

Example 2. *Suppose $f(x_1, x_2) = -(\frac{1}{2} - x_1)^3$ and $g(x_1, x_2) = 1 - x_1 - x_2$. Then, $(x_1^*, x_2^*, \lambda^*) = (\frac{1}{2}, \frac{1}{2}, 0)$ satisfies the constraint qualification conditions, it solves*

$$\begin{array}{lccc}
(a) & \dfrac{\partial \mathcal{L}}{\partial x_1} = 0 & \iff & 3\left(\dfrac{1}{2} - x_1\right)^2 - \lambda = 0 \\
(b) & \dfrac{\partial \mathcal{L}}{\partial x_2} = 0 & \iff & -\lambda = 0 \\
(c) & \dfrac{\partial \mathcal{L}}{\partial \lambda} = 0 & \iff & 1 - x_1 - x_2 = 0
\end{array}$$

---

[3] Note that $F''(x)$ equals

$$\partial^2_{xx} f(x, h(x)) + \partial^2_{xy} f(x, h(x)) h'(x) + \partial^2_{yx} f(x, h(x)) h'(x) + \partial^2_{yy} f(x, h(x)) h'(x)^2 + \partial_y f(x, h(x)) h''(x)),$$

or, by substitution,

$$\left(1 \quad h'(x) \right) D^2 f(x, h(x)) \begin{pmatrix} 1 \\ h'(x) \end{pmatrix} - \frac{\partial_y f(x, h(x))}{\partial_y g(x, h(x))} \left(1 \quad h'(x) \right) D^2 g(x, h(x)) \begin{pmatrix} 1 \\ h'(x) \end{pmatrix}.$$

Substitution at x^* yields the expressions that follows, by definition of y^* and λ^*.

and satisfies the (necessary) second order condition since

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial x_i, x_i}(x_1^*, x_2^*, \lambda^*) &= 0, for\ i = 1, 2\\ \frac{\partial \mathcal{L}}{\partial x_i, x_j}(x_1^*, x_2^*, \lambda^*) &= 0, for\ i \neq j.\end{aligned}$$

However, clearly $(x_1^*, x_2^*) = (\frac{1}{2}, \frac{1}{2})$ *is not a local maximiser since* $f(\frac{1}{2}, \frac{1}{2}) = 0$ *but* $(\frac{1}{2} + \varepsilon, \frac{1}{2} - \varepsilon)$ *is also in the constrained set and* $f(\frac{1}{2} + \varepsilon, \frac{1}{2} - \varepsilon) > 0$ *for any* $\varepsilon > 0$

The following Theorem provides sufficient conditions:

THEOREM 3 (Lagrange - FOSC and SOSC).

Let $f : D \to \mathbb{R}$ *and* $g : D \to \mathbb{R}^J$ *be* $\mathbf{C}^2$, *with* $J \leq K$. *Suppose* $(x^*, \lambda^*) \in \mathbb{R}^K \times \mathbb{R}^J$ *satisfy:*
(i) $D\mathcal{L}(x^*, \lambda^*) = 0$ *and*
(ii) $\Delta^\top D^2_{x,x}\mathcal{L}(x^*, \lambda^*)\Delta < 0$ *for all* $\Delta \in \{\mathbb{R}^J \setminus \{0\} : \Delta \cdot Dg(x^*) = 0\}$.
Then, x^* *locally solves Problem* (1).