University of Warwick
Math for Economics, 2009–10

Lectures 6 and 7: Constrained Optimization

Suppose that $f : \mathbb{R} \to \mathbb{R}$ is differentiable, and suppose that for $a, b \in \mathbb{R}$, $a < b$, we want to find $x^* \in [a, b]$ such that $f(x) \leq f(x^*)$ at all $x \in [a, b]$. That is, we want to solve the problem

$$\max f(x) : x \geq a \text{ and } x \leq b.$$

If $x^* \in (a, b)$ solves the problem, then x^* is a local maximizer of f and $f'(x^*) = 0$. If, alternatively, $x^* = b$ solves the problem, then it must be that $f'(x^*) \geq 0$. Finally, if $x^* = a$ solves the problem, it follows that $f'(x^*) \leq 0$.

It is then straightforward that if x^* solves the problem, then there exist $\lambda_a^*, \lambda_b^* \in \mathbb{R}_+$ such that $f(x^*) - \lambda_b^* + \lambda_a^* = 0$, $\lambda_a^*(x^* - a) = 0$ and $\lambda_b^*(b - x^*) = 0$.[1] It is customary to define a function

$$\mathcal{L} : \mathbb{R}^3 \to \mathbb{R}; \mathcal{L}(x, \lambda_a, \lambda_b) = f(x) + \lambda_b(b - x) + \lambda_a(a - x),$$

which is called *the Lagrangean*, and with which the first condition can be re-written as

$$\frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda_a^*, \lambda_b^*) = 0.$$

In this section we show how these Lagrangean methods work, and emphasize when they fail.

1 Equality constraints

For this section, we 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 we want to solve the problem

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

which means, in our previous notation, that we want 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 "result" 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*, for reasons that we now study.

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

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

[1] The second and third condition simply express that (i) if $x^* \in (a, b)$, then $\lambda_a^* = 0$ and $\lambda_b^* = 0$; (ii) if $x^* = b$, then $\lambda_a^* = 0$ and $\lambda_b^* \geq 0$; and (iii) if $x^* = a$, then $\lambda_a^* \geq 0$ and $\lambda_b^* = 0$.

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

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

(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

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

or, equivalently,

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

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

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

we get that

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

whereas

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

Then, our method has apparently shown that:

> *Let f and g be differentiable. $x^* \in D$* locally *solves the Problem* (1),[2] only if *there exists $\lambda^* \in \mathbb{R}^J$ such that $Df(x^*) + \lambda^{*\top} Dg(x^*) = 0$.*

The latter means that: (i) as in the unrestricted case, the differential approach, at least in principle, only finds local extrema; and (ii) the Lagrangean condition is only necessary and not sufficient by itself. So, we need to be careful and study further conditions for sufficiency. Also, we need to determine under what conditions can we find the function y and, moreover, be sure that it is differentiable.

For sufficiency, we can again appeal to our crude method and use the sufficiency results we inherit from unconstrained optimization. Since we now need F to be differentiable twice, so as to make it possible that $F''(x^*) < 0$, we must assume that so are f and g, and moreover, we need to know $y''(x)$. Since we already know $y'(x)$, by differentiation,

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

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

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

$$\begin{pmatrix} 1 & y'(x^*) \end{pmatrix} D^2_{(x,y)}\mathcal{L}(x^*, y^*, \lambda^*) \begin{pmatrix} 1 \\ y'(x^*) \end{pmatrix} < 0.$$

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

$$\begin{pmatrix} 1 & y'(x^*) \end{pmatrix} \cdot Dg(x^*, y^*) = 0,$$

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 $\Delta^\top D^2_{(x,y)}\mathcal{L}(x^*, y^*, \lambda^*)\Delta < 0$.

So, in summary, we seem to have argued to following result:

> *Suppose that $f, g \in \mathbf{C}^1$. Then:*
> *(1) $x^* \in D$ locally solves Problem* (1), only if *there exists $\lambda^* \in \mathbb{R}^J$ such that $D\mathcal{L}(x^*, \lambda^*) = 0$.*
> *(2) If $f, g \in \mathbf{C}^2$ and there exists $\lambda^* \in \mathbb{R}^J$ such that (i) $D\mathcal{L}(x^*, \lambda^*) = 0$, and (ii) that for every $\Delta \in \mathbb{R}^2 \setminus \{0\}$ such that $\Delta \cdot Dg(x^*, y^*) = 0$, we have that $\Delta^\top D^2_{(x,y)}\mathcal{L}(x^*, y^*, \lambda^*)\Delta < 0$; then, $x^* \in D$ locally solves Problem* (1).

But we still need to argue that we can indeed solve y as a function of x . 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 $D \subseteq \mathbb{R}^{K+J}$ and let $g : D \to \mathbb{R}^J \in \mathbf{C}^1$. If $(x^*, y^*) \in D$ is such that* $\operatorname{rank}(D_y g(x^*, y^*)) = J$, *then there exist $\varepsilon, \delta > 0$ and $\gamma : B_\varepsilon(x^*) \to B_\delta(y^*) \in \mathbf{C}^1$ such that:*

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

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, finally, the following result is true:

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

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

or, by substitution,

$$\begin{pmatrix} 1 & y'(x) \end{pmatrix} D^2 f(x, y(x)) \begin{pmatrix} 1 \\ y'(x) \end{pmatrix} - \frac{\partial_y f(x, y(x))}{\partial_y g(x, y(x))} \begin{pmatrix} 1 & y'(x) \end{pmatrix} D^2 g(x, y(x)) \begin{pmatrix} 1 \\ y'(x) \end{pmatrix}.$$

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

Theorem 2 (Lagrange). *Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}^J$ be of class $\mathbf{C}^1$, with $J \leq K$. Let $x^* \in D$ be such that* $\text{rank}(Dg(x^*)) = J$. *Then,*

1. *If x^* locally solves Problem* (1), *then there exists $\lambda^* \in \mathbb{R}^J$ such that $D\mathcal{L}(x^*, \lambda^*) = 0$.*

2. *If there exists $\lambda^* \in \mathbb{R}^J$ such that (i) $D\mathcal{L}(x^*, \lambda^*) = 0$ and (ii) for every $\Delta \in \mathbb{R}^J \setminus \{0\}$ such that $\Delta \cdot Dg(x^*) = 0$, it is true that $\Delta^\top D^2_{x,x}\mathcal{L}(x^*, \lambda^*)\Delta < 0$; then, x^* locally solves Problem* (1).

2 Inequality constraints

2.1 Linear programming

Let A be an $m \times n$ matrix, and let $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$.[4] Consider the following problem:

$$v_p = \max_{x \in \mathbb{R}^n} c \cdot x : Ax \leq b.$$

This is one of several equivalent representations of linear programs: problems where a linear function is to be optimized over a polyhedron. There are several interesting results and well understood algorithms that solve this kind of problem. Here, we focus on two specific results.

Define the following "dual" problem,

$$v_d = \min_{y \in \mathbb{R}^m_+} b \cdot y : y^\top A = c^\top.$$

From now on, refer to the original problem as the "primal." Notice for any $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+$ such that $Ax \leq b$ and $y^\top A = c^\top$, it is true that $c \cdot x = y^\top Ax \leq y^\top b$, so it follows that, if they exist, $v_p \leq v_d$.

That is, for any $y \geq 0$ such that $y^\top A = c$, the number $b \cdot y$ is an upper bound to the solution of the primal problem; the dual problem finds the lowest such upper bound. Crucially, if both problems are feasible, then they both have solution and their solutions are the same!

Theorem 3 (The Duality Theorem). *Suppose that there exists $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ such that $A\bar{x} \leq b$ and $\bar{y}^\top A = c^\top$. Then, $v_p = v_d \in \mathbb{R}$.*

Proof: It suffices to show that there exists $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+$ such that $Ax \leq b$, $y^\top A = c^\top$ and $c \cdot x \geq b \cdot y$. By the Theorem of the Alternative (or Farkas's lemma), it suffices to show that for any $(\alpha, \beta, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}_+ \times \mathbb{R}^n$, if $\alpha^\top A - \beta c^\top = 0$ and $\beta b^\top + \mu^\top A^\top = 0$, then $\alpha^\top b + \mu^\top c \geq 0$. Now, to see that this true, consider the following two cases:
(1) If $\beta > 0$, then

$$\alpha^\top b = \frac{\beta}{\beta} b^\top \alpha = -\frac{1}{\beta}\mu^\top A^\top \alpha = -\frac{1}{\beta}\alpha A^\top \mu = -\frac{1}{\beta}\beta c^\top \mu.$$

(2) If $\beta = 0$, then

$$\alpha^\top b \geq \alpha^\top A\bar{x} = 0 = \mu^\top A^\top \bar{y} = \mu^\top c.$$

Q.E.D.

Theorem 4 (Complementary Slackness). *Suppose that $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ satisfies $A\bar{x} \leq b$ and $\bar{y}^\top A = c^\top$. The following statements are equivalent:*

[4] We will follow the convention that all vectors are taken as columns.

1. *$\bar{x}$ solves the primal problem and $\bar{y}$ solves the dual problem;*

2. $\bar{y}^\top(b - A\bar{x}) = 0$.

Proof: To see that 1 implies 2, notice that, by the Duality Theorem, $c \cdot \bar{x} = b \cdot \bar{y}$, while $\bar{y}^\top A = c^\top$.

To see that 2 implies 1, it suffices to show that $c \cdot \bar{x} = b \cdot \bar{y}$. But this is immediate, since $\bar{y}^\top A = c^\top$, if $\bar{y}^\top(b - A\bar{x}) = 0$. *Q.E.D.*

2.2 Non-linear programming

As before, let $f : D \to \mathbb{R} \in \mathbf{C}^1$ and $f : D \to \mathbb{R}^J \in \mathbf{C}^1$. Now suppose that we have to solve the problem

$$\max_{x \in D} f(x) : g(x) \geq 0. \tag{2}$$

Again, the "usual" method says that one should try to find $(x^*, \lambda^*) \in D \times \mathbb{R}^J_+$ such that $D_x\mathcal{L}(x^*, \lambda^*) = 0$, $g(x^*) \geq 0$ and $\lambda^* \cdot g(x^*) = 0$. It is as though there is a theorem that states:

> *If $x^* \in D$ locally solves Problem (2), then there exists $\lambda^* \in \mathbb{R}^J_+$ such that $D_x\mathcal{L}(x^*, \lambda^*) = 0$, $g(x^*) \geq 0$ and $\lambda^* \cdot g(x^*) = 0$.*

Now, even though in this statement we are recognizing the local character and (only) the necessity of the result, we still have to worry about constraint qualification. To see that this is the case, consider the following example:

Example 1. *Consider the problem*

$$\max_{(x,y) \in \mathbb{R}^2} -((x-3)^2 + y^2) : 0 \leq y \leq -(x-1)^3.$$

The Lagrangean of this problem can be written as

$$\mathcal{L}(x, y, \lambda_1, \lambda_2) = -(x-3)^2 - y^2 + \lambda_1(-(x-1)^3 - y) + \lambda_2 y.$$

Notice that, although $(1, 0)$ solves the problem, there is no solution $(x^*, \lambda_1^*, \lambda_2^*)$ to the following system:
(i) $-2(x^* - 3) + 3\lambda_1^*(x^* - 1)^2 = 0$ and $-2y^* - \lambda_1^* + \lambda_2^* = 0$;
(ii) $\lambda_1^* \geq 0$ and $\lambda_2^* \geq 0$;
(iii) $-(x^* - 1)^3 - y^* \geq 0$ and $y^* \geq 0$; and
(iv) $\lambda_1^*(-(x^* - 1)^3 - y^*) = 0$ and $\lambda_2^* y^* = 0$.

If the first order conditions were necessary even without the constraint qualification (i.e. if the statement were true) the system of equations in the previous example would necessarily have to have a solution. The point of the example is just that the theorem requires the constraint qualification condition: the following theorem is true.

Theorem 5 (Kühn - Tucker). *Let $f : D \to \mathbb{R} \in \mathbf{C}^1$ and $g : D \to \mathbb{R}^J \in \mathbf{C}^1$. Let $x^* \in D$ be such that $g(x^*) \geq 0$. Define the set $\mathcal{I} = \{j \in \{1, \ldots, J\} | g_j(x^*) = 0\}$, let $I = \#\mathcal{I}$, and suppose that $\operatorname{rank}(D\tilde{g}(x^*)) = I$ for $\tilde{g} : D \to \mathbb{R}^I$ defined by $\tilde{g}(x) = (g_j(x))_{j \in \mathcal{I}}$. Then,*

1. *If x^* is a local solution to Problem (2), then there exists $\lambda^* \in \mathbb{R}^J_+$ such that $D_x\mathcal{L}(x^*, \lambda^*) = 0$, $g(x^*) \geq 0$ and $\lambda^* \cdot g(x^*) = 0$.*

2. *Suppose that $f, g \in \mathbf{C}^2$ and there exists $\lambda^* \in \mathbb{R}^J_+$ such that:*
   (i) $D_x\mathcal{L}(x^*, \lambda^*) = 0$,
   (ii) $g(x^*) \geq 0$,
   (iii) $\lambda^* \cdot g(x^*) = 0$, and
   (iv) $\Delta^\top D^2_{x,x}\mathcal{L}(x^*, \lambda^*)\Delta < 0$ for all $\Delta \in \mathbb{R}^I \setminus \{0\}$ such that $\Delta \cdot D\tilde{g}(x^*) = 0$.
   Then, x^* is a local solution to Problem (2).

As before, it must be noticed that there is a gap between necessity and sufficiency, and that the theorem only gives local solutions. For the former problem, there is no solution. For the latter, one can study concavity of the objective function and convexity of the feasible set. Importantly, notice that with inequality constraint the sign of λ does matter: this is because of the geometry of the theorem: a local maximizer is attained when the feasible directions, as determined by the gradients of the binding constraints is exactly opposite to the desired direction, as determined by the gradient of the objective function. Obviously, locally only the binding constraints matter, which explains why the constraint qualification looks more complicated here than with equality constraints. Finally, it is crucial to notice that the process does not amount to maximizing $\mathcal{L}$: in general, $\mathcal{L}$ does not have a maximum; what one finds is a saddle point of $\mathcal{L}$.

The proof of the following result is left as an exercise

THEOREM 6. *Suppose that $f : \mathbb{R}^K \to \mathbb{R}$ and $g : \mathbb{R}^K \to \mathbb{R}^J$ are both of class $\mathbf{C}^1$.*

1. *Suppose that the set $F = \{x \in \mathbb{R}^n | g(x) \geq 0\}$ is compact, and that for every $x \in F$, if we denote $\mathcal{I}(x) = \{j \in \{1, \ldots, J\} | g_j(x) = 0\}$ and $I(x) = \#\mathcal{I}(x)$, we have that*

$$\operatorname{rank}([Dg_j(x)]_{j\in\mathcal{I}(x)}) = I(x).$$

   If there exists $x^* \in F$ such that
   (i) there is some $\lambda^* \in \mathbb{R}^m_+$ for which $D_x\mathcal{L}(x^*, \lambda^*) = 0$ and $\lambda^* \cdot g^*(x^*) = 0$; and
   (ii) for every $x \in F \setminus \{x^*\}$ and all $\lambda \in \mathbb{R}^m_+$, the equality $D_x\mathcal{L}(x, \lambda) = 0$ implies that $\lambda \cdot g(x) > 0$;
   then x^* uniquely solves Problem (2).

2. *Suppose that there exists no pair (x, λ) for which $D_x\mathcal{L}(x, \lambda) = 0$, $\lambda \geq 0$, $g(x) \geq 0$, and $\lambda \cdot g(x) = 0$. Then, x^* locally solves Problem (2) only if*

$$\operatorname{rank}((Dg_i(x))_{i\in\mathcal{I}}) < I,$$

   where $\mathcal{I} = \{i \in \{1, \ldots, J\} | g_i(x^*) = 0\}$ and $I = \#\mathcal{I}$.

3 Parametric programming

We now study how the solution of a problem depends on the parameters that define the problem.

3.1 Continuity

Let $\Omega \subseteq \mathbb{R}^M$ be nonempty, and let $D : \Omega \twoheadrightarrow \mathbb{R}^K$ be a correspondence from Ω into $\mathbb{R}^K$ The importance of the concept of continuity of correspondences is given by the following result

Theorem 7 (Theorem of the Maximum). *Let function $f : \mathbb{R}^K \times \Omega \to \mathbb{R}$ be continuous and let correspondence $D : \Omega \twoheadrightarrow \mathbb{R}^K$ be nonempty-, compact-valued and continuous. The correspondence $X : \Omega \twoheadrightarrow \mathbb{R}^K$ defined by*

$$X(\omega) = \operatorname{argmax}_{x \in D(\omega)} f(x, \omega)$$

is upper hemicontinuous (and nonempty- and compact-valued) and the ("value") function $v : \Omega \to \mathbb{R}$, defined by

$$v(\omega) = \max_{x \in D(\omega)} f(x, \omega),$$

is continuous.

3.2 Differentiability

Suppose now that both sets $D \subseteq \mathbb{R}^K$ and $\Omega \subseteq \mathbb{R}^M$, are open and finite-dimensional. Suppose that $f : D \times \Omega \to \mathbb{R}$ and $g : D \times \Omega \to \mathbb{R}^J$, and consider the following (simplified) parametric problem: given $\omega \in \Omega$, let

$$v(\omega) = \max_{x \in D} f(x, \omega) : g(x, \omega) = 0.$$

Suppose that the differentiability and second-order conditions are given, so that a point x^* solves this maximization problem if, and only if, there exists a $\lambda^* \in \mathbb{R}^J$ such that $D\mathcal{L}(x^*, \lambda^*, \omega) = 0$.

Suppose furthermore that we can define functions $x : \Omega \to D$ and $\lambda : \Omega \to \mathbb{R}^J$, given by the solution of the problem and the associated multiplier, for every ω. Then, it follows directly from the Implicit Function Theorem that if, for a given $\bar{\omega} \in \Omega$,

$$\operatorname{rank} \begin{pmatrix} 0_{J \times J} & D_x g(x^*, \bar{\omega}) \\ D_x g(x^*, \bar{\omega})^\top & D^2_{x,x} \mathcal{L}(x^*, \lambda^*, \bar{\omega}) \end{pmatrix} = J + K,$$

then there exists some $\epsilon > 0$ such that on $B_\epsilon(\bar{\omega})$ the functions x and λ are differentiable and

$$\begin{pmatrix} D\lambda(\bar{\omega}) \\ Dx(\bar{\omega}) \end{pmatrix} = - \begin{pmatrix} 0_{J \times J} & D_x g(x^*, \bar{\omega}) \\ D_x g(x^*, \bar{\omega})^\top & D^2_{x,x} \mathcal{L}(x^*, \lambda^*, \bar{\omega}) \end{pmatrix}^{-1} \begin{pmatrix} D_\omega g(x(\bar{\omega}), \bar{\omega}) \\ D^2_{\omega,x} \mathcal{L}(x(\bar{\omega}), \lambda(\omega), \bar{\omega}) \end{pmatrix}.$$

It is then immediate that v is differentiable at $\bar{\omega}$ and

$$Dv(\bar{\omega}) = D_x f(x(\bar{\omega}), \bar{\omega}) Dx(\bar{\omega}).$$

A simpler method, however, is given by the following theorem

Theorem 8 (The Envelope Theorem). *If, under the assumptions of this subsection, v is continuously differentiable at $\bar{\omega}$, then $Dv(\bar{\omega}) = D_\omega \mathcal{L}(x(\bar{\omega}), \lambda(\bar{\omega}), \bar{\omega})$.*

Proof: One just needs to use the Chain Rule: by assumption,

$$D_x f(x(\omega), \omega) + D_x g(x(\omega), \omega)^\top \lambda(\omega) = 0,$$

whereas $g(x(\omega), \omega) = 0$, so

$$D_x g(x(\omega), \omega) Dx(\omega) + D_\omega g(x(\omega), \omega) = 0;$$

meanwhile,

$$\begin{aligned} Dv(\omega) &= Dx(\omega)^\top D_x f(x(\omega),\omega) + D_\omega f(x(\omega),\omega) \\ &= -D_x(\omega)^\top D_x g(x(\omega),\omega)^\top \lambda(\omega) + D_\omega f(x(\omega),\omega) \end{aligned}$$

and

$$\begin{aligned} D_\omega \mathcal{L}(x(\omega),\lambda(\omega),\omega) &= D_\omega f(x(\omega),\omega) + D_\omega g(x(\omega),\omega)^\top \lambda(\omega) \\ &= D_\omega f(x(\omega),\omega) - D_x(\omega)^\top D_x g(x(\omega),\omega)^\top \lambda(\omega), \end{aligned}$$

which gives the result. *Q.E.D.*

Exercise 1. *Let* $f : \mathbb{R}^K \to \mathbb{R}$, $g : \mathbb{R}^K \to \mathbb{R}^J \in \mathbf{C}^2$, *with* $J \leq K \in \mathbb{N}$. *Suppose that for all* $\omega \in \mathbb{R}^m$, *the problem*

$$\max f(x) : g(x) = \omega$$

has a solution, which is characterized by the first order conditions of the Lagrangean defined by $\mathcal{L}(x,\lambda,\omega) = f(x) + \lambda \cdot (\omega - g(x))$. *Suppose furthermore that these conditions define differentiable functions* $x : \mathbb{R}^J \to \mathbb{R}^K$ *and* $\lambda : \mathbb{R}^J \to \mathbb{R}^J$. *Prove that* $Dv(\omega) = \lambda(\omega)$, *for all* ω, *where* $v : \mathbb{R}^J \to \mathbb{R}$ *is the value function of the problem.*