Module 5: Optimization

A central tenet – perhaps *the* central tenet – of economics is that agents make decisions in a manner that is consistent with optimizing an objective. Further, most interesting economics optimization problems involve constraints – scarcity is a central theme after all. The principal analytic tool for elementary optimization problems is the Kuhn-Tucker theorem. In chapter, besides stating the theorem, we use the total differential to develop the intuition underlying it – this "variational approach" will extend to many other types of optimization problems, including continuous-time optimal control, i.e. the *calculus of variations*.

1 Background notions

Let $f : \mathbb{R}^n \to \mathbb{R}$ by C^2 and $K \subset \mathbb{R}^n$. Interpret f as the *objective* function and K as the *constraint* set. When $K = \mathbb{R}^n$ the problem is said to be *unconstrained*. Recall that given $\varepsilon > 0$, the ball at x of radius ε, written $B(x, \varepsilon)$, is the set of points in $\mathbb{R}^n$ within a distance ε of x. A point $x^* \in K$ is a *local maximum* provided there is some $\varepsilon > 0$ such that for all $x \in K \cap B(x^*, \varepsilon)$ it follows that $f(x^*) \geq f(x)$; it is a *global maximum* provided that $x \in K \implies f(x^*) \geq f(x)$. To find the global maximum the standard approach is to first identify all local maxima and then pick the biggest one.

It is common for both f and K to depend on some vector of parameters $\eta \in \mathbb{R}^m$. The standard presentation of the optimization problem, which may include these parametric dependences, takes two forms:

$$V(\eta) = \max_{x \in K(\eta)} f(x, \eta)$$

$$x^*(\eta) = \operatorname*{argmax}_{x \in K(\eta)} f(x, \eta).$$

In the first case, $V(\eta)$ is the value obtained by f at the maximum, and in the second case $x^*(\eta)$ is the vector in K that attains the maximum.

Exercise 1 *Show that a minimization problem may be converted to a maximization problem.*

The *interior* of a given set $A \subset \mathbb{R}^n$ (or any metric space), often write A°, is the largest open set contained in A. The boundary of A is sometimes written ∂A, and comprises those points in A that are not in A°, i.e. $\partial A = A \setminus A^\circ$. As a simple example, let $A = [a, b] \subset \mathbb{R}$. Then $A^\circ = (a, b)$ and $\partial A = \{a, b\}$. Note that the interior may be empty: if $A = \{a\} \subset \mathbb{R}$ then there are no points A°. Note also that A is an open set if and only if $A = A^\circ$.

The constraint set K will often be taken as compact, i.e. closed and bounded; and, it will very often have a non-empty interior. In this case there are two types of solutions: *interior* solutions, which, naturally, are found in K°; and *corner* solutions, which are found on the boundary of K, i.e. in ∂K. Returning to our simple example, if $K = [a,b]$ then an interior solution to the maximization problem would be $x^* \in (a,b)$ and a corner solution would be $x^* \in \{a,b\}$. The conditions characterizing interior and corner solutions to a given problem are typically distinct, and *both types of conditions must be checked*.

Finally, there are two types of conditions used to characterize a given solution (be it an interior or a corner solution): *necessary* conditions and *sufficient* conditions. Necessary conditions are usually called *first-order conditions* because they involve first-order local approximations of the objective, and sufficient conditions are usually called *second-order conditions* for analogous reasons. Importantly, the FOC provide statements that are true about the solution x^* *assuming it exists*; SOC provide conditions that *guarantee* existence.

2 The univariate case

As usual, the univariate case provides a great deal of intuition for the more technically cumbersome, but much richer multivariate environment. Let $f : \mathbb{R} \to \mathbb{R}$ and let $K = [a,b]$. To derive the first-order conditions we always start the same way: *assume* x^* is a local maximum, and use total differentials to see what must be true about it. Said somewhat more intuitively, start at a local maximum and consider all feasible differential variations. These variations cannot increase the value of the objective – if they did, you would not have been at a local maximum.

More formally, if x^* is a local maximum then $df = f'(x^*)dx \leq 0$; after all, if $df = f'(x^*)dx > 0$ then $f(x^* + dx) > f(x^*)$ which would contradict that x^* is a local maximum. This "no-arbitrage argument" is at the heart of variational arguments. As mentioned, there are two case: interior and corner.

- First suppose x^* is an interior local maximum. Then we could have considered both positive variations, i.e. $dx > 0$, and negative variations, i.e. $dx < 0$. If $df = f'(x^*)dx \leq 0$ for both positive and negative dx then it must be the case that $f'(x^*) = 0$.
- Now turn to a corner, and suppose $x^* = a$. Then the only feasible differential variation is $dx > 0$. It follows that $f'(x^*) \leq 0$. Symmetrically, if $x^* = b$ then $f'(x^*) \geq 0$.

Now let's turn to second-order conditions. The idea is as follows: suppose we have found all possible local maxima by finding all points in $[a,b]$ that satisfy the FOC identified above. Which of these really are local maxima? This time let's start at the corners. Suppose $f'(a) \leq 0$. Since $dx > 0$ implies that $df = f'(a)dx \leq 0$, it follows that for $\varepsilon > 0$ small enough, if $x \in (a,\varepsilon)$ the $f(x) \leq f(x^*)$. We conclude no additional conditions are needed at either corner. Now consider a potential interior solution. That $f'(x^*) = 0$ is not enough to guarantee x^* is a local maximum – it could be a minimum or a horizontal point of inflection, e.g. x^3 at the origin. To handle this case, consider the second-order Taylor expansion around the candidate optimum:

$$\begin{aligned} f(x) &\approx f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2} f''(x^*)(x - x^*)^2 \\ &= f(x^*) + \frac{1}{2} f''(x^*)(x - x^*)^2, \end{aligned}$$

where the equality follows from the fact that $f'(x^*) = 0$. Near x^*, provided $f''(x^*) \neq 0$, the function f looks like a parabola; and, x^* is a local maximum if this parabola faces downward, i.e. if if the coefficient on x^2 is negative. Since this coefficient is $f''(x^*)$, we conclude that the candidate interior local maximum x^* really is a local maximum provided $f''(x^*) < 0$.

Exercise 2 *Solve the following problems.*

1. $f = x^2 - 3x + 2$, $a = 4/3$, $b = 3$
2. $f = -x^2 + 3x - 2$, $a = 2/3$, $b = 3$

3 The multivariate case

The multivariate case is more difficult principally because the constraint sets can be much more complicated. For this reason, let's begin with the unconstrained problem, as it will link very nicely to the univariate case.

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $K = \mathbb{R}^n$. We work just as before by assuming x^* is a local maximum. Now contemplate a differential variation dx. As before, it must be that $df = Df(x^*)dx \leq 0$, where we recall that $Df(x^*)$ is the row of partials $f_i(x^*)$. Since $K = \mathbb{R}^n$, any variation dx is feasible. Thus fix i and consider a variation only along the ith coordinate axis. It follows that $df = f_i(x^*)dx_i$. Since dx_i could be either positive or negative, we conclude that $f_i(x^*) = 0$ is a necessary condition for x^* to be a local maximum. This must be true for all i.

An important observation is warranted here: there are exactly n first-order conditions for this unconstrained problem, namely $f_i(x^*) = 0$ for $i = 1, \ldots n$. And there are, in effect, n unknowns: the x_i^*. This is no coincidence: the number of FOC correspond to the degrees of freedom you have in selecting potential variations, which, in the unconstrained case, corresponds to the number of choice variables. This insight will carry forward to the constrained case, and can be useful for determining when you have a complete set of FOC.

As in the univariate case, the sufficient conditions derive from the second-order Taylor expansion:

$$\begin{aligned} f(x) &\approx f(x^*) + Df(x^*)(x-x^*) + \frac{1}{2}(x-x^*)^T D^2 f(x^*)(x-x^*) \\ &= f(x^*) + \frac{1}{2}(x-x^*)^T D^2 f(x^*)(x-x^*). \end{aligned}$$

Now suppose that $D^2 f(x^*)$ is negative definite. Then, provided the approximation is accurate, i.e. provided x is near x^*,

$$f(x^*) - f(x) = -\frac{1}{2}(x-x^*)^T D^2 f(x^*)(x-x^*) > 0,$$

that is, x^* is a local maximum.

3.1 Equality constraints

Constraints in the multivariate case commonly come in two forms: equality constraints and inequality constraints; we treat them separately before wrapping the whole theory up into the Kuhn-Tucker theorem, and we begin with equality constraints.

Equality constraints present as the solution set to a system of equations, which can be written as follows: let $g : \mathbb{R}^n \to \mathbb{R}^m$ and $c \in \mathbb{R}^m$. Then

$$K = \{x \in \mathbb{R}^n : g(x) = c\}.$$

Of course we could always rewrite g to set $c = 0$, but it is often convenient in applications to have non-zero c. Note that g may be interpreted as a column vector of m functions $g^i : \mathbb{R}^n \to \mathbb{R}$.

Exercise 3 *Two quick examples: graph the follow constraint sets.*

1. $g(x,y) = x^2 + y^2$, $c = 1$
2. $g^1(x,y,z) = x^2 + y^2 + z^2$, $g^2(x,y,z) = z$, $c_1 = 1, c_2 = 0$.

To determine the FOC for this problem we adopt the same approach: assume x^* is a local maximum. Then given any feasible variation dx, we have that $df = Df(x^*)dx \leq 0$. But this time not all variations are feasible: we are only allows to choose differentials so that $x^* + dx$ remains in K, i.e. continues to satisfy the constraining equations. How do we do this? The implicit function theorem.

To illuminate the point, suppose that $m = 1$ and $n = 2$. To ensure the constraint continues to bind our variation dx must satisfy $Dg(x^*)dx = 0$. But this means

$$g_1(x^*)dx_1 + g_2(x^*)x_2 = 0$$

or

$$dx_2 = -(g_2(x^*))^{-1}g_1(x^*)dx_1,$$

provided $g_2(x^*) \neq 0$, which we will assume formally below. We conclude that

$$df = f_1(x^*)dx_1 + f_2(x^*)dx_2 = \left(f_1(x^*) - (g_2(x^*))^{-1}g_1(x^*)f_2(x^*)\right)dx_1 \leq 0$$

Since there is no sign restriction on dx_1 it follows that

$$f_1(x^*) - (g_2(x^*))^{-1}g_1(x^*)f_2(x^*) = 0,$$

or

$$\frac{f_1(x^*)}{g_1(x^*)} = \frac{f_2(x^*)}{g_2(x^*)}. \tag{1}$$

Thus in case $m = 1$ and $n = 2$ there is one FOC, as given by (**??**), and one constraint, $g(x) = 0$. That there is a single FOC even though there are two choice variables (x_1 and x_2) reflects that, because of the constraint, there is only one degree of freedom available when choosing the variation.

Exercise 4 *Rewrite (**??**) as*

$$\frac{f_1(x^*)}{f_2(x^*)} = \frac{g_1(x^*)}{g_2(x^*)}$$

and interpret graphically.

This same analysis can be completed for general m and n; however, it becomes quite tedious. Thankfully, there is an easier way to reach the same FOC. The trick is to add variables – multipliers – which convert the constrained problem into an unconstrained problem, and then to use the FOC already developed above. To this end, write

$$L(x,\lambda) = f(x) + \lambda^T(c - g(x)).$$

Here L is called the *Lagrangian* and $\lambda \in \mathbb{R}^m$ is a vector of real numbers, which you should treat as unknowns, or variables (really they are dual, or co-variables), called the Lagrange multipliers. Note that there is one multiplier for each constraint.

We have the following theorem:

Theorem 1 (Method of Lagrange) *If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are C^2, if $x^* \in \mathbb{R}^n$ is a local maximum of f subject to the constraint $g(x^*) = 0$ and if $\text{rank}(Dg(x^*)) = m$ then there exists $\lambda^* \in \mathbb{R}^m$ such that*

1. $L_{x_i}(x^*, \lambda^*) = 0$ *for* $i = 1, \dots, n$
2. $L_{\lambda_j}(x^*, \lambda^*) = 0$ *for* $j = 1, \dots, m$

This theorem provides for a mechanical procedure for solving a maximization problem with equality constraints for which you know a solution exists: check the constraint qualification $\operatorname{rank}(Dg(x^*)) = m$; write down the Lagrangian; take all the first partials,; set them equal to zero; and solve for the potential x^* and λ^*. Among them is the solution you are looking for.

Returning to the case $n = 2$ and $m = 1$, we get the following conditions:

$$\begin{aligned} L_{x_1} = 0 &\implies f x_1 - \lambda g_{x_1} = 0 \\ L_{x_2} = 0 &\implies f x_2 - \lambda g_{x_2} = 0 \\ L_\lambda = 0 &\implies g(x) = c \end{aligned}$$

Combining the first two equations, we obtain

$$\lambda = \frac{f_1(x^*)}{g_1(x^*)} = \frac{f_2(x^*)}{g_2(x^*)},$$

thus reproducing (**??**). The last equation $L_\lambda = 0$ simply reproduces the constraint.[1]

Exercise 5 *Some problems*

1. *Solve the following problem for* $c = 3$ *and* $c = 5$*:*

$$\begin{aligned} \max_{x,y} \quad & -(x-2)^2 - (y-2)^2 \\ s.t. \quad & x + y = c \end{aligned}$$

2. *Use the method of Lagrange to solve the following problem:*

$$\begin{aligned} \max_{x \geq 0, y \geq 0} \quad & x + 3y \\ & (x+1)^2 + (y+1)^2 = 4 \end{aligned}$$

3. *Consider the consumer choice problem given by*

$$\begin{aligned} \max \quad & u(x,y) \\ & p_x x + p_y y = I \end{aligned}$$

and assume an interior solution.

[1] L_λ will play a role that feels less trivial when inequality constraints are included.

(a) *Write down the FOC and interpret the multiplier.*

(b) *Assume utility satisfies $u_x > 0$, $u_y > 0$, $u_{xx} < 0$, and $u_{yy} < 0$ as usual. Assume $u_{xy} = 0$. Is x normal?*

As with the unconstrained problem, finding a critical point, i.e. one that satisfies the FOC obtained via Lagrange's method, does not imply the point is a local maximum: it could be a local minimum or a saddle point. Second-order conditions sufficient to guarantee that a critical point is a local maximum are available, but they require a modified version of the Hessian to account for the presence of the constraints.

Let $L(x,\lambda) = f(x) + \lambda^T(c - g(x))$ be the Lagrangian. Then

$$D^2L(x,\lambda) = \begin{pmatrix} L_{x_1x_1} & \cdots & L_{x_1x_n} & L_{x_1\lambda_1} & \cdots & L_{x_1\lambda_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ L_{x_nx_1} & \cdots & L_{x_nx_n} & L_{x_n\lambda_1} & \cdots & L_{x_n\lambda_m} \\ L_{\lambda_1x_1} & \cdots & L_{\lambda_1x_n} & L_{\lambda_1\lambda_1} & \cdots & L_{\lambda_1\lambda_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ L_{\lambda_mx_1} & \cdots & L_{\lambda_mx_n} & L_{\lambda_m\lambda_1} & \cdots & L_{\lambda_m\lambda_m} \end{pmatrix} = \begin{pmatrix} D_x^2L & -Dg^T \\ -Dg & 0 \end{pmatrix}.$$

Here D_x^2L is the matrix of cross partials of L with respect to x, Dg is the derivative of g with respect to x, and we observe that because λ enters the Lagrangian linearly, it follows that $L_{\lambda\lambda} = 0$. The matrix $D^2L(x,\lambda)$ is called the *bordered Hessian* because it is the Hessian of L with respect to x, i.e. D_x^2L, bordered by the derivatives of the constraint. Here is the result:

Theorem 2 *If (x^*,λ^*) satisfies Lagrange's FOC and if $v^TD_x^2L(x^*,\lambda^*)v < 0$ for all $v \notin \ker Dg(x^*,\lambda^*)$, then x^* is a local maximum.*

Thus the bordered Hessian need only be negative definite outside the kernel of Dg.

We now turn to a remarkably useful result called the *envelope theorem*, which allows for simple, and often very intuitive assessments of how the value of the maximization problem changes when a parameter changes. We begin with the case $n = 2$ and $m = 1$. Write

$$V(c) = \max_{x\in K} f(x),$$

where K is defined as $g(x) = c$. Thus the *value function* V depends on the parameter controlling the constraint, c. We want to compute $V'(c)$. Note that

$$g_1dx_1 + g_2dx_2 = dc \text{ or } dx_2 = \frac{1}{g_2}(dc - g_1dx_1).$$

Also, $dV = f_1 dx_1 + f_2 dx_2$. Combining, we get

$$dV = f_1 dx_1 + \frac{f_2}{g_2}(dc - g_1 dx_1) = \left(f_1 - \frac{g_1}{g_2} f_2 \right) dx + \frac{f_2}{g_2} dc = \lambda dc.$$

It follows that $V'(c) = \lambda = L_c$. This is an instance of the envelope theorem: to compute $V'(c)$, just differentiate the Lagrangian. Also, it provides a lovely intuition for the multiplier: λ measures the marginal value obtained by relaxing the constraint: it's the price of the constraint.

Theorem 3 (Envelope theorem) *Consider the maximization problem*

$$V(\eta) = \max_{x \in K(\eta)} f(x, \eta), \textit{ where } \eta \in \mathbb{R}^k \textit{ and } K(\eta) = \{x \in \mathbb{R}^n : g(x, \eta) = 0\}$$

If (x^*, λ^*) satisfy the Lagrange conditions and constitutes a local maximum then

$$V_{\eta_i}(\eta) = L_{\eta_i}(x^*, \lambda^*).$$

Exercise 6 *What does the Lagrange multiplier measure in the consumer problem?*

Exercise 7 *Consider the usual consumer choice problem:*

$$\begin{aligned} V(I, p) = \max_{x \geq 0, y \geq 0} \quad & u(x, y) \\ & x + py = I \end{aligned}$$

where $V(I, p)$ is the value function, and we have set the price of good x to 1. Assume quasi-linear utility: $u(x,y) = x + \nu(y)$, where $\nu' > 0$, $\nu'' < 0$. Finally, assume $\nu'(0) = 1$. Let x^*, y^* be the solutions to this problem.

1. *What is the smallest price p, so that $y^* = 0$?*
2. *Assume p is small enough so that $y^* > 0$. Compute $\partial y^* / \partial I$.*
3. *Compute $\partial V / \partial I$ using the envelope theorem, and discuss, referencing your answer to part (b).*

4 Inequality constraints

To handle inequality constraints we need to understand *complementary slackness*, which pertains to *pairs of weak inequalities*. We say that the inequalities $x \geq 0$ and $y \geq 0$ hold *with complementary slackness* (w/sc) provided that $xy = 0$. Exactly the same terminology is used for the other three possible permutations of inequality pairs, i.e. $x \geq 0$ and $y \leq 0$, etc. Here's how I remember it: if one inequality is strict then the other isn't.

Inequality constraints present similarly to equality constraints: let $g : \mathbb{R}^n \to \mathbb{R}^m$ and $c \in \mathbb{R}^m$. Then

$$K = \{x \in \mathbb{R}^n : g(x) \leq c\},$$

where the suggestive notation $g(x) \leq c$ means $g^i(x) \leq c_i$ for $i = 1, \ldots, m$. Again, we could always rewrite g to set $c = 0$, but it is often convenient in applications to have non-zero c.

Exercise 8 *Some examples: graph the follow constraint sets.*

1. $g(x,y) = x^2 + y^2$, $c = 1$
2. $g^1(x,y,z) = x^2 + y^2 + z^2$, $g^2(x,y,z) = -x$, $g^3(x,y,z) = -y$, $g^4(x,y,z) = -z$

Exercise 9 *Write a consumer choice problem with three goods using inequality constraints as denoted above.*

To gain intuition for the general result, consider again the case $n = 2$ and $m = 1$, and assume x^* is a local maximum. If the constraint is not binding, i.e. if $g(x^*) < c$, then it is as if the constraint doesn't even exist. Thus, the FOC in this case are the same as in the unconstrained case: $f_i(x^*) = 0$. If the constraint is binding then $g(x^*) = c$ and now we are back in the equality-constraint case, which, under the language of Lagrange, gives $f_i(x^*) + \lambda^* g_i(x^*) = 0$. Another point can be made in this binding case: $\lambda^* \geq 0$. Why? If it were negative then tightening the constraint would increase the value of the object; but this means we shouldn't have been on the constraint in the first place.

Notice that the unconstrained FOC is reproduced by constrained FOC when $\lambda^* = 0$. This observation, together with the language of comparative slackness, and the observation that λ^* is non-negative, allows us to succinctly combine the two cases into the following FOC:

$$f_i(x^*) + \lambda^* g_i(x^*) = 0, \text{ and } \lambda^* \geq 0 \text{ and } g(x^*) \leq c \text{ hold w/cs.}$$

Even more succinctly, we can write this in terms of the Lagrangian: the FOC are $L_i = 0$ and $\lambda^* \geq 0$, $L_\lambda \geq 0$ w/cs. Isn't that elegant?

5 Kuhn-Tucker theorem

Now we are ready to put this all together. Let $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ be C^2 (twice continuously differentiable). The general (finite-dimensional) constrained problem may be written

$$\max_{x \in \mathbb{R}^n} \quad f(x)$$
$$g^i(x) = c_i,\ i \in \{1, \ldots, m\}$$
$$h^i(x) \leq 0,\ i \in \{1, \ldots, k\}$$

Let $L : \mathbb{R}^n \oplus \mathbb{R}^m \oplus \mathbb{R}^k \to \mathbb{R}$ be given by

$$L(x, \lambda, \mu) = f(x) + \lambda^T (c - g(x)) - \mu^T h(x).$$

We say point $x \in \mathbb{R}^n$ satisfies the constraint qualification (CQ) if the derivatives of the equality constraints g^i and of the inequality constraints h^j *for which the constraint is binding, i.e.* $h^i(x) = 0$ are linearly independent. We have the following result.

Theorem 4 (Kuhn-Tucker) *If x^* solves (**??**) and satisfies (CQ) then there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^k$ so that*

1. $L_{x_i}(x^*, \lambda^*, \mu^*) = 0,\ i \in \{1, \ldots, n\}$
2. $L_{\lambda_i}(x^*, \lambda^*, \mu^*) = 0,\ i \in \{1, \ldots, m\}$
3. $L_{\mu_i}(x^*, \lambda^*, \mu^*) \geq 0$ *and* $\mu_i^* \geq 0,$ *w/cs* $i \in \{1, \ldots, k\}$

Exercise 10 *Solve the following problem:*

$$\max_{x_1 \geq 0, x_2 \geq 0} \quad -((x_1 - 1)^2 + (x_2 - 2)^2)$$
$$x_1 + 2x_2 \leq 10$$
$$2x_1 + x_2 \leq 10$$