Optimization

Written by Paul Schrimpf and modified by Hiro Kasahara
August 28, 2020
University of British Columbia

Economics 526

[1]

Today's lecture is about optimization. Useful references are chapters 1-5 of Dixit (1990), chapters 16-19 of Simon and Blume (1994).

We typically model economic agents as optimizing some objective function. Consumers maximize utility.

Example 0.1. A consumer with income y chooses a bundle of goods $x_1, \ldots, x_n$ with prices $p_1, \ldots, p_n$ to maximize utility $u(x)$.

$$\max_{x_1,\ldots,x_n} u(x_1, \ldots, x_n) \text{ s.t. } p_1 x_1 + \cdots + p_n x_n \leq y$$

Firms maximize profits.

Example 0.2. A firm produces a good y with price p, and uses inputs $x \in \mathbb{R}^k$ with prices $w \in \mathbb{R}^k$. The firm's production function is $f : \mathbb{R}^k \to \mathbb{R}$. The firm's problem is

$$\max_{y,x} py - w^T x \text{ s.t. } f(x) = y$$

[1]

1. Unconstrained optimization

Although most economic optimization problems involve some constraints it is useful to begin by studying unconstrained optimization problems. There are two reasons for this. One is that the results are somewhat simpler and easier to understand without constraints. The second is that we will sometimes encounter unconstrained problems. For example, we can substitute the constraint in the firm's problem from example 0.2 to obtain an unconstrained problem.

$$\max_x pf(x) - w^T x$$

1.1. **Notation and definitions.**

An optimization problem refers to finding the maximum or minimum of a function, perhaps subject to some constraints. In economics, the most common optimization problems are utility maximization and profit maximization. Because of this, we will state most of our definitions and results for maximization problems.

[1]See appendix A if any notation is unclear.

Of course, we could just as well state each definition and result for a minimization problem by reversing the direction of most inequalities.

We will start by examining generic unconstrained optimization problems of the form

$$\max_x f(x)$$

where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$. To allow for the possibility that the domain of f is not all of $\mathbb{R}^n$, we will let $U \subseteq \mathbb{R}^n$, be the domain, and write $\max_{x \in U} f(x)$ for the maximum of f on U.

If $F^* = \max_x f(x)$, we mean that $f(x) \leq F^*$ for all x and $f(x^*) = F^*$ for some x^*.

Definition 1.1. $F^* = \max_{x \in U} f(x)$ is the **maximum** of f on U if $f(x) \leq F^*$ for all x and $f(x^*) = F^*$ for some x^*.

There may be more than one such x^*. We denote the set of all x^* such that $f(x^*) = F^*$ by $\arg\max_{x \in U} f(x)$ and might write $x^* \in \arg\max_x f(x)$, or, if we know there is only one such, x^*, we sometimes write $x^* = \arg\max_x f(x)$.

Definition 1.2. $x^* \in U$ is a **maximizer** of f on U if $f(x*) = \max_{x \in U} f(x)$. The set of all maximizers is denoted $\arg\max_{x \in U} f(x)$.

Definition 1.3. $x^* \in U$ is a **strict maximizer** of f if $f(x*) > f(x)$ for all $x \neq x^*$

Definition 1.4. f has a **local maximum** at x if $\exists \delta > 0$ such that $f(y) \leq f(x)$ for all y in U and within δ distance of x (we will use the notation $y \in N_\delta(x) \cap U$, which could be read y in a δ neighborhood of x intersected with U). Each such x is called a **local maximizer** of f. If $f(y) < f(x)$ for all $y \neq x$, $y \in N_\delta(x) \cap U$, then we say f has a **strict local maximum** at x.

When we want to be explicit about the distinction between local maximum and the maximum in definition 1.1, we refer to the later as the global maximum.

Example 1.1. Here are some examples of functions from $\mathbb{R} \to \mathbb{R}$ and their maxima and minima.

(1) $f(x) = x^2$ is minimized at $x = 0$ with minimum 0.
(2) $f(x) = c$ has minimum and maximum c. Any x is a maximizer.
(3) $f(x) = \cos(x)$ has maximum 1 and minimum -1. $2\pi n$ for any integer n is a maximizer.
(4) $f(x) = \cos(x) + x/2$ has no global maximizer or minimizer, but has many local ones.

A related concept to the maximum is the supremum.

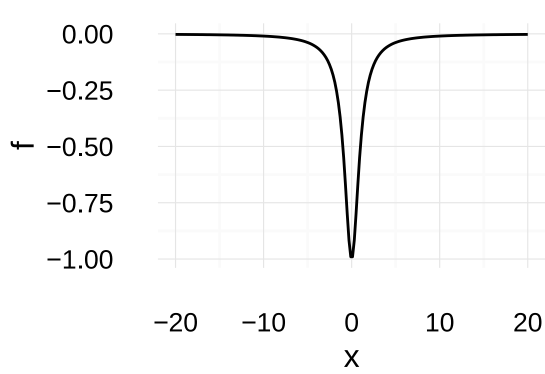
Definition 1.5. The **supremum** (or least upper bound) of f on U is S is written,

$$S = \sup_{x \in U} f(x),$$

and means that $S \geq f(x)$ for all $x \in U$ and for any $A < S$ there exists $x_A \in U$ such that $f(x_A) > A$.

The main difference between the supremum and maximum is that the supremum can exist when the maximum does not.

Example 1.2. Let $x \in \mathbb{R}$ and $f(x) = -\frac{1}{1+x^2}$. Then $\max_{x\in\mathbb{R}} f(x)$ does not exist, but $\sup_{x\in\mathbb{R}} f(x) = 0$.



An important fact about real numbers is that if $f : \mathbb{R}^n \to \mathbb{R}$ is bounded above, i.e. there exists $B \in \mathbb{R}$ such that $f(x) \leq B$ for all $x \in \mathbb{R}^n$, then $\sup_{x\in\mathbb{R}^n} f(x)$ exists and is unique.

The infimum is to the minimum as the supremum is to the maximum.

1.2. **First order conditions.** [2]Suppose we want to maximize $f(x)$. If we are at x and travel some small distance Δ in direction v, then we can approximate how much f will change by its directional derivative, i.e.

$$f(x + \Delta v) \approx f(x) + df(x; v)\Delta.$$

If there is some direction v with $df(x; v) \neq 0$, then we could take that Δv or $-\Delta v$ step and reach a higher function value. Therefore, if x is a local maximum, then it must be that $df(x; v) = 0$ for all v. From theorem B.1 this is the same as saying $\frac{\partial f}{\partial x_i} = 0$ for all i.

Theorem 1.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$, and suppose f has a local maximum x and f is differentiable at x. Then $\frac{\partial f}{\partial x_i} = 0$ for $i = 1, \ldots, n$.*

Proof. The paragraph preceding the theorem is an informal proof. The only detail missing is some care to ensure that the $\approx$ is sufficiently accurate. If you are interested, see the past course notes for details. http://faculty.arts.ubc.ca/pschrimpf/526/lec10optimization.pdf □

The first order condition is the fact that $\frac{\partial f}{\partial x_i} = 0$ is a necessary condition for x to be a local maximizer or minimizer of f.

Definition 1.6. Any point x such that f is differentiable at x and $Df_x = 0$ is call a **critical point** of f.

If f is differentiable, f cannot have local minima or maxima (=local extrema) at non-critical points. f might have a local extrema its critical points, but it does not have to. Consider $f(x) = x^3$ $f'(0) = 0$, but 0 is not a local maximizer or minimizer of F. Similarly, if $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x) = x_1^2 - x_2^2$, then the partial derivatives at 0 are 0, but 0 is not a local minimum or maximum of f.

[2]This section will use some results on partial and directional derivatives. See appendix B and the summer math review for reference.

2. Second order conditions

To determine whether a given critical point is a local minimum or maximum or neither we can look at the second derivative of the function. Let $f : \mathbb{R}^n \to \mathbb{R}$ and suppose x^* is a critical point. Then $\frac{\partial f}{\partial x_i}(x^*) = 0$. To see if x^* is a local maximum, we need to look at $f(x^* + h)$ for small h. We could try taking a first order expansion,

$$f(x^* + h) \approx f(x^*) + Df_{x^*}h,$$

but we know that $Df_{x^*} = 0$, so this expansion is not useful for comparing $f(x^*)$ with $f(x^* + h)$.

If f is twice continuously differentiable, we can instead take a second order expansion of f around x^*.

$$f(x^* + v) \approx f(x^*) + Df_{x^*}v + \frac{1}{2}v^T D^2 f_{x*}v$$

where $D^2 f_{x^*}$ is the matrix of f's second order partial derivatives. Since x^* is a critical, point $Df_{x^*} = 0$, so

$$f(x^* + v) - f(x^*) \approx \frac{1}{2}v^T D^2 f_{x^*}v.$$

We can see that x^* is a local maximum if

$$\frac{1}{2}v^T D^2 f_{x^*}v < 0$$

for all $v \neq 0$. The above inequality will be true if $v^T D^2 f_{x^*}v < 0$ for all $v \neq 0$. The Hessian, $D^2 f_{x^*}$ is just some symmetric n by n matrix, and $v^T D^2 F_{x^*}v$ is a quadratic form in v. This motivates the following definition.

Definition 2.1. Let A be a symmetric matrix, then A is

- **Negative definite** if $x^T Ax < 0$ for all $x \neq 0$
- **Negative semi-definite** if $x^T Ax \leq 0$ for all $x \neq 0$
- **Positive definite** if $x^T Ax > 0$ for all $x \neq 0$
- **Positive semi-definite** if $x^T Ax \geq 0$ for all $x \neq 0$
- **Indefinite** if $\exists\ x_1$ s.t. $x_1^T Ax_1 > 0$ and some other x_2 such that $x_2^T Ax_2 < 0$.

Later, we will derive some conditions on A that ensure it is negative (semi-)definite. For now, just observe that if $D^2 F_{x*}$ is negative semi-definite, then x^* must be a local maximum. If $D^2 F_{x^*}$ is negative definite, then x^* is a strict local maximum. The following theorem restates the results of this discussion.

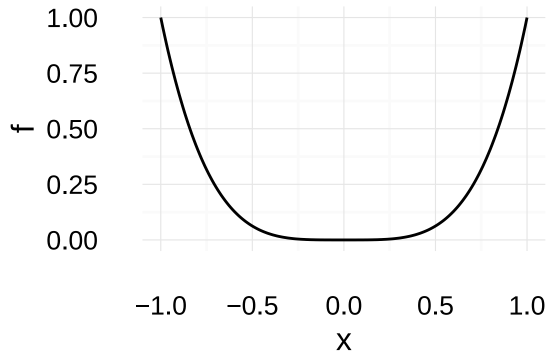
Theorem 2.1. *Let $F : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and let x^* be a critical point. If*

1. *The Hessian, $D^2 F_{x^*}$, is negative definite, then x^* is a strict local maximizer.*
2. *The Hessian, $D^2 F_{x^*}$, is positive definite, then x^* is a strict local minimizer.*
3. *The Hessian, $D^2 F_{x^*}$, is indefinite, x^* is neither a local min nor a local max.*
4. *The Hessian is positive or negative semi-definite, then x^* could be a local maximum, minimum, or neither.*

Proof. The main idea of the proof is contained in the discussion preceding the theorem. The only tricky part is carefully showing that our approximation is good enough. $\square$

When the Hessian is not positive definite, negative definite, or indefinite, the result of this theorem is ambiguous. Let's go over some examples of this case.

Example 2.1. $F : \mathbb{R} \to \mathbb{R}$, $F(x) = x^4$. The first order condition is $4x^3 = 0$, so $x^* = 0$ is the only critical point. The Hessian is $F''(x) = 12x^2 = 0$ at x^*. However, x^4 has a strict local minimum at 0.



Example 2.2. $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x_1, x_2) = -x_1^2$. The first order condition is $DF_x = (-2x_1, 0) = 0$, so the $x_1^* = 0$, $x_2^* \in \mathbb{R}$ are all critical points. The Hessian is

$$D^2F_x = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}$$

This is negative semi-definite because $h^T D^2 F_x h = -2h_1^2 \leq 0$. Also, graphing the function would make it clear that $x_1^* = 0$, $x_2^* \in \mathbb{R}$ are all (non-strict) local maxima.

Example 2.3. $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x_1, x_2) = -x_1^2 + x_2^4$. The first order condition is $DF_x = (-2x_1, 4x_2^3) = 0$, so the $x^* = (0, 0)$ is a critical point. The Hessian is

$$D^2F_x = \begin{pmatrix} -2 & 0 \\ 0 & 12x_2^2 \end{pmatrix}$$

This is negative semi-definite at 0 because $h^T D^2 F_0 h = -2h_1^2 \leq 0$. However, 0 is not a local maximum because $F(0, x_2) > F(0, 0)$ for any $x_2 \neq 0$. 0 is also not a local minimum because $F(x_1, 0) < F(0, 0)$ for all $x_1 \neq 0$.

In each of these examples, the second order condition is inconclusive because $h^T D^2 F_{x^*} h = 0$ for some h. In these cases we could determine whether x^* is a local maximum, local minimum, or neither by either looking at higher derivatives of F at x^*, or look at D^2F_x for all x in a neighborhood of x^*. We will not often encounter cases where the second order condition is inconclusive, so we will not study these possibilities in detail.

Example 2.4 (Competitive multi-product firm)**.** Suppose a firm has produces k goods using n inputs with production function $f : \mathbb{R}^n \to \mathbb{R}^k$. The prices of the goods are p, and the prices of the inputs are w, so that the firm's profits are

$$\Pi(x) = p^T f(x) - w^T x.$$

The firm chooses x to maximize profits.

$$\max_x p^T f(x) - w^T x$$

The first order condition is

$$p^T Df_{x^*} - w = 0.$$

or without using matrices,

$$\sum_{j=1}^{k} p_j \frac{\partial f_j}{\partial x_i}(x^*) = w_i$$

for $i = 1, ..., n$.

The second order condition is that

$$D^2[p^T f]_{x^*} = \begin{pmatrix} \sum_{j=1}^k p_j \frac{\partial^2 f_j}{\partial x_1^2}(x^*) & \cdots & \sum_{j=1}^k p_j \frac{\partial^2 f_j}{\partial x_1 \partial x_n}(x^*) \\ \vdots & & \vdots \\ \sum_{j=1}^k p_j \frac{\partial^2 f_j}{\partial x_1 \partial x_n}(x^*) & \cdots & \sum_{j=1}^k p_j \frac{\partial^2 f_j}{\partial x_n^2}(x^*) \end{pmatrix}$$

must be negative semidefinite.[a]

[a] There is some awkwardness in the notation for the second derivative of a function from $\mathbb{R}^n \to \mathbb{R}^k$ when $k > 1$. The first partial derivatives can be arranged in an $k \times n$ matrix. There are then $k \times n \times n$ second partial derivatives. When $k = 1$ it is convenient to write these in an $n \times n$ matrix. When $k > 1$, one could arrange the second partial derivates in 3-d array of dimension $k \times n \times n$, or one could argue that an $kn \times n$ matrix makes sense. These higher order derivatives are called tensors and physicists and mathematicians do have some notation for working with them. However, tensors do not come up that often in economics. Moreover, we can usually avoid needing any such notation by rearranging some operations. In the above example, $f : \mathbb{R}^n \to \mathbb{R}^k$, so $D^2 f$ should be avoided (especially if we combine it with other matrix and vector algebra), but $p^T f(x)$ is a function from $\mathbb{R}^n$ to $\mathbb{R}$, so $D^2[p^T f]$ is unambiguously an $n \times n$ matrix.

3. Constrained optimization

To begin our study of constrained optimization, let's consider a consumer in an economy with two goods.

Example 3.1. Let x_1 and x_2 be goods with prices p_1 and p_2. The consumer has income y. The consumer's problem is

$$\max_{x_1, x_2} u(x_1, x_2) \text{ s.t. } p_1 x_1 + p_2 x_2 \leq y$$

In economics 101, you probably analyzed this problem graphically by drawing indifference curves and the budget set, as in figure 1. In this figure, the curved lines are indifference curves. That is, they are regions where utility is constant. So going outward from the axis they show all (x_1, x_2) such that $u(x_1, x_2) = c_1$, $u(x_1, x_2) = c_2$, $u(x_1, x_2) = c_3$, and $u(x_1, x_2) = c_4$ with $c_1 < c_2 < c_3 < c_4$. The diagonal line is the budget constraint. All points below this line satisfy the constraint $p_1 x_1 + p_2 x_2 \leq y$. The optimal x_1, x_2 is the point x^* where an indifference curve is tangent to the budget line. The slope of the budget line is $-\frac{p_1}{p_2}$. The slope of the indifference curve can be found by implicitly differentiating,

$$\frac{\partial u}{\partial x_1}(x^*)dx_1 + \frac{\partial u}{\partial x_2}(x^*)dx_2 = 0$$
$$\frac{dx_2}{dx_1}(x^*) = \frac{-\frac{\partial u}{\partial x_1}}{\frac{\partial u}{\partial x_2}}$$

So the optimum satisfies

$$\frac{\frac{\partial u}{\partial x_1}}{\frac{\partial u}{\partial x_2}} = \frac{p_1}{p_2}.$$

The ratio of marginal utilities equals the ratio of marginal prices. Rearranging, we can also say that

$$\frac{\frac{\partial u}{\partial x_1}}{p_1} = \frac{\frac{\partial u}{\partial x_2}}{p_2}.$$

Suppose we have a small amount of additional income Δ. If we spend this on x_1, we would get Δ/p_1 more of x_1, which would increase our utility by $\frac{\frac{\partial u}{\partial x_1}}{p_1}\Delta$. At the optimum, the marginal utility from spending additional income on either good should be the same. Let's call this marginal utility of income μ, then we have

$$\mu = \frac{\frac{\partial u}{\partial x_1}}{p_1} = \frac{\frac{\partial u}{\partial x_2}}{p_2}$$

or

$$\frac{\partial u}{\partial x_1} - \mu p_1 = 0$$
$$\frac{\partial u}{\partial x_2} - \mu p_2 = 0.$$

This is the standard first order condition for a constrained optimization problem.

Now, let's look at a generic maximization problem with equality constraints.

$$\max f(x) \text{ s.t. } h(x) = c$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$. As in the unconstrained case, consider a small perturbation of x by some small v. The function value is then approximately,

$$f(x+v) \approx f(x) + Df_x v$$

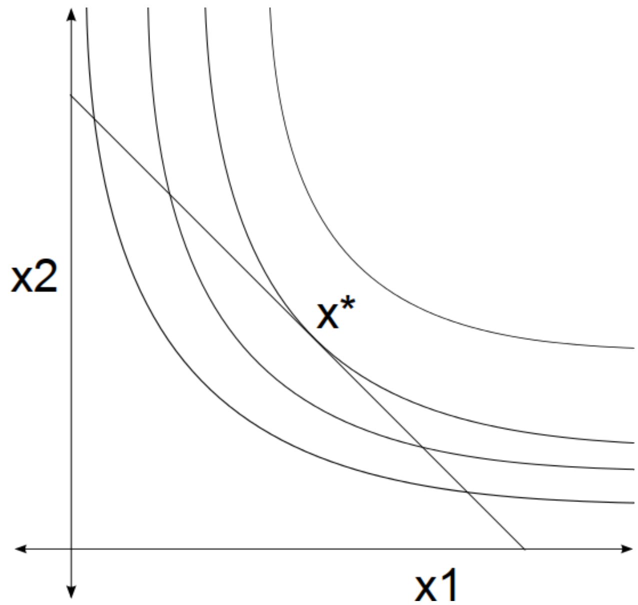
and the constraint value is

$$h(x+v) \approx h(x) + Dh_x v$$

$x + v$ will satisfy the constraint only if $Dh_x v = 0$. Thus, x is a constrained local maximum if for all v such that $Dh_x v = 0$ we also have $Df_x v = 0$.

Now, we want to express this condition in terms of Lagrange multipliers. Remember that Dh_x is an $m \times n$ matrix of partial derivatives, and Df_x is a $1 \times n$ row vector of partial

Figure 1. Indifference curves



derivatives. If $n = 2$ and $m = 1$ this condition can be written as

$$\frac{\partial h}{\partial x_1}(x)v_1 + \frac{\partial h}{\partial x_2}(x)v_2 = 0$$

implies

$$\frac{\partial f}{\partial x_1}(x)v_1 + \frac{\partial f}{\partial x_2}(x)v_2 = 0.$$

Solving the first of these equations for v_2, substituting into the second and rearranging gives

$$\frac{\frac{\partial f}{\partial x_1}(x)}{\frac{\partial h}{\partial x_1}(x)} = \frac{\frac{\partial f}{\partial x_2}(x)}{\frac{\partial h}{\partial x_2}(x)},$$

which as in the consumer example above, we can rewrite as

$$\begin{aligned} \frac{\partial f}{\partial x_1} - \mu \frac{\partial h}{\partial x_1} &= 0 \\ \frac{\partial f}{\partial x_2} - \mu \frac{\partial h}{\partial x_2} &= 0. \end{aligned}$$

for some constant μ. Similar reasoning would show that for any n and $m = 1$, if for all v such that $Dh_x v = 0$ we also have $Df_x v = 0$, then it is also true that there exists μ such that

$$\frac{\partial f}{\partial x_i} - \mu \frac{\partial h}{\partial x_i} = 0$$

for $i = 1, ..., n$. Equivalently, using matrix notation $Df_x = \mu Dh_x$. Note that conversely, if such a μ exists, then trivially for any v with $Dh_x v = 0$, we have $Df_x v = \mu Dh_x v = 0$.

When there are multiple constraints, i.e. $m > 1$, it would be reasonable to guess that $Dh_x v = 0$ implies $Df_x v = 0$ if and only if there exists $\mu \in \mathbb{R}^m$ such that

$$\frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i} = 0$$

or in matrix notation $Df_x = \mu^T Dh_x$. This is indeed true. It is tedious to show using algebra, but will be easy to show later using some results from linear algebra.

A heuristic geometric argument supporting the previous claims is as follows. Imagine the level sets or contour lines of $h(\cdot)$. The feasible points that satisfy the constraint lie on the contour line with $h(x) = 0$. Similarly imagine the contour lines of $f(\cdot)$. If x^* is a constrained maximizer, then as we move alone the contour line $h(\cdot) = 0$, we must stay on the same contour line (or set if f is flat) of f. In other words, the contour lines of h and f must be parallel at x^*. Gradients give the direction of steepest ascent and are perpendicular to contour lines. Therefore, if the gradients of h and f are parallel, then so are the contour lines. Gradients being parallel just means that they should be multiples of one another, i.e. $Df_x = \mu^T Dh_x$ for some μ^T.

The following theorem formally states the preceding result.

Theorem 3.1 (First order condition for maximization with equality constraints)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable. Suppose x^* is a constrained local maximizer of f subject to $h(x) = c$. Also assume that Dh_{x^*} has rank m. Then there exists $\mu^* \in \mathbb{R}^m$ such that (x^*, μ^*) is a critical point of the Lagrangian,*

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

i.e.

$$\begin{aligned} \frac{\partial L}{\partial x_i}(x^*, \mu^*) &= \frac{\partial f}{\partial x_i} - \mu^{*T} \frac{\partial h}{\partial x_i}(x^*) = 0 \\ \frac{\partial L}{\partial \mu_j}(x^*, \mu^*) &= h(x^*) - c = 0 \end{aligned}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

The assumption that Dh_{x*} has rank m is needed to make sure that we don't divide by zero when defining the Lagrange multipliers. When there is a single constraint, it simply says that at least one of the partial derivatives of the constraint is non-zero. This assumption is called the non-degenerate constraint qualification. It is rarely an issue for equality constraints, but can sometimes fail as in the following examples.

Exercise 3.1. (1) Solve the problem:

$$\max_x \alpha_1 \log x_1 + \alpha_2 \log x_2 \text{ s.t. } (p_1 x_1 + p_2 x_2 - y)^3 = 0$$

What goes wrong? How does it differ from the following problem:

$$\max_x \alpha_1 \log x_1 + \alpha_2 \log x_2 \text{ s.t. } (p_1 x_1 + p_2 x_2 - y) = 0$$

(2) Solve the problem:

$$\max_{x_1,x_2} x_1 + x_2 \text{ s.t. } x_1^2 + x_2^2 = 0$$

3.1. **Lagrange multipliers as shadow prices.** Recall that in our example of consumer choice, the Lagrange multiplier could be interpreted as the marginal utility of income. A similar interpretation will always apply.

Consider a constrained maximization problem,

$$\max_x f(x) \text{ s.t. } h(x) = c$$

From 3.1, the first order conditions are

$$\begin{aligned} Df_{x^*} - \mu^T Dh_{x*} =& 0 \\ h(x^*) - c =& 0. \end{aligned}$$

What happens to x^* and $f(x^*)$ if c changes? Let $x^*(c)$ denote the maximizer as a function of c. Differentiating the constraint with respect to c shows that

$$Dh_{x^*(c)} Dx_c^* = I$$

By the chain rule,

$$D_c\left(f(x^*(c))\right) = Df_{x^*(c)} Dx_c^*.$$

Using the first order condition to substitute for $Df_{x^*(c)}$, we have

$$\begin{aligned} D_c\left(f(x^*(c))\right) =& \mu^T Dh_{x^*(c)} Dx_c^* \\ =& \mu^T \end{aligned}$$

Thus, the multiplier, μ, is the derivative of the maximized function with respect to c. In economic terms, the multiplier is the marginal value of increasing the constraint. Because of this μ_j is called the shadow price of c_j.

Example 3.2 (Cobb-Douglas utility). Consider the consumer's problem in example 3.1 with Cobb-Douglas utility,

$$\max_{x_1,x_2} x_1^{\alpha_1} x_2^{\alpha_2} \text{ s.t. } p_1x_1 + p_2x_2 = y$$

The first order conditions are

$$\begin{aligned} \alpha_1 x_1^{\alpha_1 - 1} x_2^{\alpha_2} - p_1\mu = 0 \\ x_1^{\alpha_1} \alpha_2 x_2^{\alpha_2 - 1} - p_2\mu = 0 \\ p_1x_1 + p_2x_2 - y = 0 \end{aligned}$$

Solving for x_1 and x_2 yields

$$\begin{aligned} x_1 =& \frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{y}{p_1} \\ x_2 =& \frac{\alpha_2}{\alpha_1 + \alpha_2} \frac{y}{p_2}. \end{aligned}$$

The expenditure share of each good $\frac{p_j x_j}{y}$, is constant with respect to income. Many economists, going back to Engel (1857) have studied how expenditure shares vary with income. See Lewbel (2008) for a brief review and additional references.

If we solve for μ, we find that

$$\mu = \frac{\alpha_1^{\alpha_1}\alpha_2^{\alpha_2}}{(\alpha_1+\alpha_2)^{\alpha_1+\alpha_2}} \frac{y^{\alpha_1+\alpha_2-1}}{p_1^{\alpha_1} p_2^{\alpha^2}}$$

μ is the marginal utility of income in this model. As an exercise you may want to verify that if we were to take a monotonic transformation of the utility function (such as $\log(x_1^{\alpha_1}x_2^{\alpha_2})$ or $(x_1^{\alpha_1}x_2^{\alpha_2})^\beta$) and re-solve the model, then we would obtain the same demand functions, but a different marginal utility of income.

3.2. **Inequality constraints.** Now let's consider an inequality instead of equality constraint.

$$\max_x f(x) \text{ s.t. } g(x) \leq b.$$

When the constraints binds, i.e. $g(x^*) = b$, the situation is exactly the same as with an equality constraint. However, the constraints do not necessarily bind at a local maximum, so we must allow for that possibility.

Suppose x^* is a constrained local maximum. To begin with, let's consider a simplified case where there is only one constraint, i.e. $g : \mathbb{R}^n \to \mathbb{R}$. If the constraint does not bind, then for small changes to x^*, the constraint will still not bind. In other words, locally it as though we have an unconstrained problem. Then, as in our earlier analysis of unconstrained problems, it must be that $Df_{x^*} = 0$. Now, suppose the constraint does bind. Consider a small change in x, v. Since x^* is a local maximum, it must be that if $f(x^*+v) > f(x^*)$, then x^*+v must violate the constraint, $g(x^*+v) > b$. Taking first order expansions, we can say that if $Df_{x^*}v > 0$, then $Dg_{x^*}v > 0$. This will be true if $Df_{x^*} = \lambda Dg_{x^*}$ for some $\lambda > 0$. Note that the sign of the multiplier λ matters here.

To summarize the results of the previous paragraph. If x^* is a constrained local maximum, then there exists $\lambda \geq 0$ such that $Df_{x^*} - \lambda Dg_{x^*} = 0$. Furthermore if $\lambda > 0$, then $g(x^*) = b$. If $g(x^*) < b$, then $\lambda = 0$ (it is also possible, but rare for both $\lambda = 0$ and $g(x^*) = b$). This situation where if one inequality is strict and then another holds with equality and vice versa is called a complementary slackness condition.

Essentially the same result holds with multiple inequality constraints.

Theorem 3.2 (First order condition for maximization with inequality constraints)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable. Suppose x^* is a local maximizer of f subject to $g(x) \leq b$. Suppose that the first $k \leq m$ constraints, bind*

$$g_j(x^*) = b_j$$

for $j = 1...k$ and that the Jacobian for these constraints,

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_k}{\partial x_1} & \cdots & \frac{\partial g_k}{\partial x_n} \end{pmatrix}$$

has rank k. Then, there exists $\lambda^* \in \mathbb{R}^m$ such that for

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

we have

$$\begin{aligned} \frac{\partial L}{\partial x_i}(x^*, \lambda^*) =& \frac{\partial f}{\partial x_i} - \lambda^{*T} \frac{\partial g}{\partial x_i}(x^*) = 0 \\ \lambda_j^* \frac{\partial L}{\partial \lambda_j}(x^*, \lambda^*) =& \lambda_j^* \left(g_j(x^*) - b \right) = 0 \\ \lambda_j^* \geq& 0 \\ g_j(x^*) \leq& b_j \end{aligned}$$

for $i = 1, ..., n$ and $j = 1, ..., m$. Moreover for each j at least if $\lambda_j^* > 0$ then $g_j(x^*) = b_j$ and if $g_j(x^*) < b$, then $\lambda_j^* = 0$ (complementary slackness).

Some care needs to be taken regarding the sign of the multipliers and the setup of the problem. Given the setup of our problem, $\max_x f(x)$ s.t. $g(x) \leq b$, if we write the Lagrangian as $L(x, \lambda) = f(x) - \lambda^T (g(x) - b)$, then the $\lambda_j \geq 0$. If we instead wrote the Lagrangian as $L(x, \lambda) = f(x) + \lambda^T (g(x) - b)$, then we would get negative multipliers. If we have a minimization problem, $\min_x f(x)$ s.t. $g(x) \leq b$, the situation is reversed. $f(x) - \lambda^T (g(x) - b)$ leads to negative multipliers, and $f(x) + \lambda^T (g(x) - b)$ leads to positive multipliers. Reversing the direction of the inequality in the constraint, i.e. $g(x) \geq b$ instead of $g(x) \leq b$, also switches the sign of the multipliers. Generally it is not very important if you end up with multipliers with the "wrong" sign. You will still end up with the same solution for x. The sign of the multiplier does matter for whether it is the shadow price of b or $-b$.

To solve the first order conditions of an inequality constrained maximization problem, you would first determine or guess which constraints do and do not bind. Then you would impose $\lambda_j = 0$ for the non-binding constraints and solve for x and the remaining components λ. If the resulting x do lead to the constraints not binding that you guessed not binding, then that x is a possible local maximum. To find all possible local maxima, you would have to repeat this for all possible combinations of constraints binding or not. There 2^m such possibilities, so this process can be tedious.

Example 3.3. Let's solve

$$\begin{aligned} \max_x x_1 x_2 \text{ s.t. } & x_1^2 + 2x_2^2 \leq 3 \\ & 2x_1^2 + x_2^2 \leq 3 \end{aligned}$$

The Lagrangian is

$$L(x, \lambda) = x_1 x_2 - \lambda_1 (x_1^2 + 2x_2^2 - 3) - \lambda_2 (2x_1^2 + x_2^2 - 3)$$

The first order conditions are

$$
\begin{aligned}
0 =& x_2 - 2\lambda_1 x_1 - 4\lambda_2 x_1 \\
0 =& x_1 - 4\lambda_1 x_2 - 2\lambda_2 x_2 \\
0 =& \lambda_1(x_1^2 + 2x_2^2 - 3) \\
0 =& \lambda_2(2x_1^2 + x_2^2 - 3)
\end{aligned}
$$

Now, we must guess which constraints bind. Since the problem is symmetric in x_1 and x_2, we only need to check three possibilities instead of all four. The three possibilities are neither constraint binds, both bind, or one binds and one does not.

If neither constraint binds, then $\lambda_1 = \lambda_2 = 0$, and the first order conditions imply $x_1 = x_2 = 0$. This results in a function value of 0. This is feasible, but it is not the maximum since for example, small positive x_1 and x_2 would also be feasible and give a higher function value.

Instead, suppose both constraints bind. Then the solutions to $x_1^2 + 2x_2^2 = 3$ and $2x_1^2 + x_2^2 = 3$ are $x_1 = \pm 1$ and $x_2 = \pm 1$. Substituting into the first order condition and solving for λ_1 and λ_2 gives

$$
\begin{aligned}
0 =& 1 - 2\lambda_1 - 4\lambda_2 \\
0 =& 1 - 4\lambda_1 - 2\lambda_2 \\
\frac{1}{6} =& \lambda_1 = \lambda_2
\end{aligned}
$$

Both λ_1 and λ_2 are positive, so complementary slackness is satisfied.

Finally, let's consider the case where the first constraint binds but not the second. In this case $\lambda_2 = 0$, and taking the ratio of the first order conditions gives

$$
\frac{x_2}{x_1} = \frac{1}{2}\frac{x_1}{x_2}
$$

or $x_2^2 = \frac{1}{2}x_1^2$. Substituting this into the first constraint yields $x_1^2 + x_1^2 = 3$, so $x_1 = \sqrt{3/2}$, and $x_2 = \sqrt{3/4}$. However, then $2x_1^2 + x_1^2 = 3 + 3/4 > 3$. The second constraint is violated, so this cannot be the solution.

Fortunately, the economics of a problem often suggest which constraints will bind, and you rarely actually need to investigate all 2^m possibilities.

Example 3.4 (Quasi-linear preferences)**.** A consumer's choice of some good(s) of interest x and spending on all other goods z is often modeled using quasi-linear preferences. $U(x, z) = u(x) + z$. The consumer's problem is

$$
\begin{aligned}
\max_{x,z} u(x) + z \text{ s.t. } & px + z \leq y \\
& x \geq 0 \\
& z \geq 0
\end{aligned}
$$

Let λ, μ_x, and μ_z be the multipliers on the constraints. The first order conditions are

$$\begin{aligned} u'(x) - \lambda p + \mu_x &= 0 \\ 1 - \lambda + \mu_z &= 0 \end{aligned}$$

along with the complementary slackness conditions

$$\begin{aligned} \lambda &\geq 0 & px + z &\leq y \\ \mu_x &\geq 0 & x &\geq 0 \\ \mu_z &\geq 0 & z &\geq 0 \end{aligned}$$

There are $2^3 = 8$ possible combinations of constraints binding or not. However, we can eliminate half of these possibilities by observing that if the budget constraint is slack, $px + z < y$, then increasing z is feasible and increases the objective function. Therefore, the budget constraint must bind, $px + z = y$ and $\lambda > 0$. There are four remaining possibilities. Having $0 = x = z$ is not possible because then the budget constraint would not bind. This leaves three possibilities

(1) $0 < x$ and $0 < z$. That implies $\mu_x = \mu_z = 0$. Rearranging the first order conditions gives that $u'(x) = p$ and from the budget constraint $z = y - px$. Depending on u, p, and y, this could be the solution.
(2) $0 = x$ and $0 < z$. That implies that $\mu_z = 0$ and $\mu_x > 0$. From the budget constraint, $z = y$. From the first order conditions we have $u'(0) = p + \mu_x$. Since μ_x must be positive, this equation requires that $u'(0) < p$. This also could be the solution depending on u, p, and y
(3) $0 < x$ and $0 = z$. That implies that $\mu_x = 0$. From the budget constraint, $x = y/p$. Combining the first order conditions to eliminate λ gives $1 + \mu_z = \frac{u'(y/p)}{p}$. Since $\mu_z > 0$, this equation requires that $u'(y/p) > p$. This also could be the solution depending on u, p, and y.

Exercise 3.2. Show that replacing the problem

$$\max_x f(x) \text{ s..t } g(x) \leq b.$$

with an equality constrained problem with slack variables s,

$$\max_{x,s} f(x) \text{ s..t } g(x) - s = b, s \geq 0$$

leads to the same conclusions as theorem 3.2

3.3. **Second order conditions.** As with unconstrained optimization, the first order conditions from the previous section only give a necessary condition for x^* to be a local maximum of $f(x)$ subject to some constraints. To verify that a given x^* that solves the first order condition is a local maximum, we must look at the second order condition. As in the unconstrained case, we can take a second order expansion of $f(x)$ around x^*.

$$
\begin{aligned}
f(x^*+v) - f(x^*) \approx & Df_{x^*}v + \frac{1}{2}v^T D^2 f_{x^*} v \\
\approx & \frac{1}{2}v^T D^2 f_{x^*} v
\end{aligned}
$$

This is a constrained problem, so any x^*+v must satisfy the constraints as well. As before, what will really matter are the equality constraints and binding inequality constraints. To simplify notation, let's just work with equality constraints, say $h(x)=c$. We can take a first order expansion of h around x^* to get

$$h(x^*+v) \approx h(x^*) + Dh_{x*}v = c.$$

Then v satisfies the constraints if

$$
\begin{aligned}
h(x^*) + Dh_{x^*}v = & c \\
Dh_{x^*}v = & 0
\end{aligned}
$$

Thus, x^* is a local maximizer of f subject to $h(x)=c$ if

$$v^T D^2 f_{x^*} v \leq 0$$

for all v such that that $Dh_{x^*}v=0$. The following theorem precisely states the result of this discussion.

Theorem 3.3 (Second order condition for constrained maximization)**.** *Let $f: U \to \mathbb{R}$ be twice continuously differentiable on U, and $h: U \to \mathbb{R}^l$ and $g: U \to \mathbb{R}^m$ be continuously differentiable on $U \subseteq \mathbb{R}^n$. Suppose $x^* \in \operatorname{interior}(U)$ and there exists $\mu^* \in \mathbb{R}^l$ and $\lambda^* \in \mathbb{R}^m$ such that for*

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

we have

$$
\begin{aligned}
\frac{\partial L}{\partial x_i}(x^*,\lambda^*) = & \frac{\partial f}{\partial x_i} - \lambda^{*T}\frac{\partial g}{\partial x_i}(x^*) - \mu^{*T}\frac{\partial h}{\partial x_i}(x^*) = 0 \\
\frac{\partial L}{\partial \mu_\ell}(x^*,\lambda^*) = & h_\ell(x^*) - c = 0 \\
\lambda_j^* \frac{\partial L}{\partial \lambda_j}(x^*,\lambda^*) = & \lambda_j^*\left(g(x^*) - c\right) = 0 \\
\lambda_j^* \geq & 0 \\
g(x^*) \leq & b
\end{aligned}
$$

Let B be the matrix of the derivatives of the binding constraints evaluated at x^*,

$$B = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_l}{\partial x_1} & \cdots & \frac{\partial h_l}{\partial x_n} \\ \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_k}{\partial x_1} & \cdots & \frac{\partial g_k}{\partial x_n} \end{pmatrix}$$

If

$$v^T D^2 f_{x*} v < 0$$

for all $v \neq 0$ such that $Bv = 0$, then x^* is a strict local constrained maximizer for f subject to $h(x) = c$ and $g(x) \leq b$.

Recall from above that an n by n matrix, A, is negative definite if $x^T Ax < 0$ for all $x \neq 0$. Similarly, we say that A is negative definite on the null space[3] of B if $x^T Ax < 0$ for all $x \in \mathcal{N}(B) \setminus \{0\}$. Thus, the second order condition for constrained optimization could be stated as saying that the Hessian of the objective function must be negative definite on the null space of the Jacobian of the binding constraints. The proof is similar to the proof of the second order condition for unconstrained optimization, so we will omit it.

4. Comparative statics

Often, a maximization problem will involve some parameters. For example, a consumer with Cobb-Douglas utility solves:

$$\max_{x_1,x_2} x_1^{\alpha_1} x_2^{\alpha_2} \text{ s.t. } p_1 x_1 + p_2 x_2 = y.$$

The solution to this problem depends α_1, α_2, p_1, p_2, and y. We are often interested in how the solution depends on these parameters. In particular, we might be interested in the maximized value function (which in this example is the indirect utility function)

$$v(p_1, p_2, y, \alpha_1, \alpha_2) = \max_{x_1,x_2} x_1^{\alpha_1} x_2^{\alpha_2} \text{ s.t. } p_1 x_1 + p_2 x_2 = y.$$

We might also be interested in the maximizers $x_1^*(p_1, p_2, y, \alpha_1, \alpha_2)$ and $x_2^*(p_1, p_2, y, \alpha_1, \alpha_2)$ (which in this example are the demand functions).

4.1. Envelope theorem.

The envelope theorem tells us about the derivatives of the maximum value function. Suppose we have an objective function that depends on some parameters θ. Define the value function as

$$v(\theta) = \max_x f(x, \theta) \text{ s.t. } h(x) = c.$$

Let $x^*(\theta)$ denote the maximizer as function of θ, i.e.

$$x^*(\theta) = \arg\max_x f(x, \theta) \text{ s.t. } h(x) = c.$$

[3] The null space of an $m \times n$ matrix B is the set of all $x \in \mathbb{R}^n$ such that $Bx = 0$. We will discuss null spaces in more detail later in the course.

By definition $v(\theta) = f(x^*(\theta), \theta)$. Applying the chain rule we can calculate the derivative of v. For simplicity, we will treat x and θ as scalars, but the same calculations work when they are vectors.

$$\frac{dv}{d\theta} = \frac{\partial f}{\partial \theta}(x^*(\theta), \theta) + \frac{\partial f}{\partial x}(x^*(\theta), \theta)\frac{dx^*}{d\theta}(\theta)$$

From the first order condition, we know that

$$\frac{\partial f}{\partial x}(x^*(\theta), \theta)\frac{dx^*}{d\theta} = \mu \frac{\partial h}{\partial x}(x^*(\theta))\frac{dx^*}{d\theta}.$$

However, since $x^*(\theta)$ is a constrained maximum for each θ, it must be that $h(x^*(\theta)) = c$ for each θ. Therefore,

$$0 = \frac{d}{d\theta}h(x^*(\theta)) = \frac{\partial h}{\partial x}(x^*(\theta))\frac{dx^*}{d\theta} = 0.$$

Thus, we can conclude that

$$\frac{dv}{d\theta} = \frac{\partial f}{\partial \theta}(x^*(\theta), \theta).$$

More generally, you might also have parameters in the constraint,

$$v(\theta) = \max_x f(x, \theta) \text{ s.t. } h(x, \theta) = c.$$

Following the same steps as above, you can show that

$$\frac{dv}{d\theta} = \frac{\partial f}{\partial \theta} - \mu \frac{\partial h}{\partial \theta} = \frac{\partial L}{\partial \theta}.$$

Note that this result also covers the previous case where the constraint did not depend on θ. In that case, $\frac{\partial h}{\partial \theta} = 0$ and we are left with just $\frac{dv}{d\theta} = \frac{\partial f}{\partial \theta}$.

The following theorem summarizes the above discussion.

Theorem 4.1 (Envelope)**.** *Let $f : \mathbb{R}^{n+k} \to \mathbb{R}$ and $h : \mathbb{R}^{n+k} \to \mathbb{R}^m$ be continuously differentiable. Define*

$$v(\theta) = \max_x f(x, \theta) \text{ s.t. } h(x, \theta) = c$$

where $x \in \mathbb{R}^n$ and $\theta \in \mathbb{R}^k$. Then

$$D_\theta v_\theta = D_\theta f_{x^*,\theta} - \mu^T D_\theta h_{x^*,\theta}$$

where μ are the Lagrange multipliers from theorem 3.1, $D_\theta f_{x^*,\theta}$ denotes the $1 \times k$ matrix of partial derivatives of f with respect to θ evaluated at (x^*, θ) , and $D_\theta h_{x^*,\theta}$ denotes the $m \times k$ matrix of partial derivatives of f with respect to θ evaluated at (x^*, θ).

Example 4.1 (Consumer demand)**.** Some of the core results of consumer theory can be shown as simple consequences of the envelope theorem. Consider a consumer choosing goods $x \in \mathbb{R}^n$ with prices p and income y. Define

$$v(p, y) = \max_x u(x) \text{ s.t. } p^T x - y \leq 0.$$

In this context, v is called the indirect utility function. From the envelope theorem,

$$\frac{\partial v}{\partial p_i} = -\mu x_i^*(p,y)$$
$$\frac{\partial v}{\partial y} = \mu$$

Taking the ratio we have a relationship between the (Marshallian) demand function, x^*, and the indirect utility function,

$$x_i^*(p,y) = \frac{-\frac{\partial v}{\partial p_i}}{\frac{\partial v}{\partial y}}.$$

This is known as Roy's identity.

Now consider the mirror problem of minimizing expenditure given a target utility level,

$$e(p,\bar{u}) = \min_x p^T x \text{ s.t. } u(x) \geq \bar{u}$$

In general, e is a maximum value function, but in this particular context, it is the consumer's expenditure function. Using the envelope theorem again,

$$\frac{\partial e}{\partial p_i}(p,\bar{u}) = x_i^h(p,\bar{u})$$

where $x_i^h(p,\bar{u})$ is the constrained minimizer. It is the Hicksian (or compensated) demand function.

Finally, using the fact that $x_i^h(p,\bar{u}) = x_i^*(p, e(p,\bar{u}))$ and differentiating with respect to p_k gives Slutsky's equation.

$$\begin{aligned}\frac{\partial x_i^h}{\partial p_k} &= \frac{\partial x_i^*}{\partial p_k} + \frac{\partial x_i^*}{\partial y}\frac{\partial e}{\partial p_k} \\ &= \frac{\partial x_i^*}{\partial p_k} + \frac{\partial x_i^*}{\partial y} x_k^*\end{aligned}$$

Slutsky's equation is useful because we can determine the sign of some of these derivatives. From above we know that

$$\frac{\partial e}{\partial p_i}(p,\bar{u}) = x_i^h(p,\bar{u})$$

so

$$\frac{\partial x_i^h(p,\bar{u})}{\partial p_j} = \frac{\partial^2 e}{\partial p_j \partial p_i}(p,\bar{u})$$

If we fix p and $x_0 = x^h(p,\bar{u})$, we know that $u(x_0) \geq \bar{u}$, so x_0 satisfies the contraint in the minimum expenditure problem. Therefore, for any $\tilde{p}$,

$$\tilde{p}^T x_0 \geq e(\tilde{p},\bar{u}) = \min_x \tilde{p}^T x \text{ s.t. } u(x) \geq \bar{u}$$

Taking a second order expansion, this gives

$$\begin{aligned}\tilde{p}^T x_0 - p^T x_0 \geq & e(\tilde{p}, \bar{u}) - e(p, \bar{u}) \approx D_p e_{(p,\bar{u})}(\tilde{p} - p) + \frac{1}{2}(\tilde{p} - p)^T D_p^2 e_{(p,\bar{u})}(\tilde{p} - p) \\ 0 \geq & (\tilde{p} - p)^T D_p^2 e_{(p,\bar{u})}(\tilde{p} - p)\end{aligned}$$

where the second line uses the fact that $D_p e_{(p,\bar{u})} = x_0^T$. Hence, we know that $D_p^2 e_{(p,\bar{u})} = D_p x^h_{(p,\bar{u})}$ is negative semi-definite. In particular, the diagonal, $\frac{\partial x_i^h}{\partial p_i}$, must be less than or equal to zero. Hicksian demand curves must slope down. Marshallian demand curves usually do too, but might not when the income effect, $\frac{\partial x_i^*}{\partial y} x_i^*$, is large.

We can also analyze how the maximizer varies with the parameters. We do this by totally differentiating the first order condition. Consider an equality constrained problem,

$$\max_x f(x, \theta) \text{ s.t. } h(x, \theta) = c.$$

where $x \in \mathbb{R}^n$, $\theta \in \mathbb{R}^s$, and $c \in R^m$. The optimal x much satisfy the first order condition and the constraint.

$$\begin{aligned}\frac{\partial f}{\partial x_j} - \sum_{k=1}^m \mu_k \frac{\partial h_k}{\partial x_j} =& 0 \\ h_k(x, \theta) - c =& 0\end{aligned}$$

for $j = 1, ..., n$, and $k = 1, ..., m$. Suppose θ changes by $d\theta$, let dx and $d\mu$ be the amounts that x and μ have to change by to make the first order conditions still hold. These must satisfy,

$$\begin{aligned}\sum_{\ell=1}^n \frac{\partial^2 f}{\partial x_j \partial x_\ell} dx_\ell + \sum_{r=1}^s \frac{\partial f}{\partial x_j \partial \theta_r} d\theta_r - \sum_{k=1}^m \mu_k \left(\sum_{\ell=1}^n \frac{\partial h_k}{\partial x_j \partial x_\ell} dx_\ell + \sum_{r=1}^s \frac{\partial h_k}{\partial x_j \partial \theta_r} d\theta_r \right) - \sum_{k=1}^m d\mu_k \frac{\partial h_k}{\partial x_j} =& 0 \\ \sum_{\ell=1}^n \frac{\partial h_k}{\partial x_\ell} dx_\ell + \sum_{r=1}^s \frac{\partial h_k}{\partial \theta_r} d\theta_r =& 0\end{aligned}$$

where the first equation is for $j = 1, ..., n$ and the second is for $k = 1, ..., m$. We have $n + m$ equations to solve for the $n + m$ unknown dx and $d\mu$. We can express this system of equations more compactly using matrices.

$$\begin{pmatrix} D_{xx}^2 f - D_{xx}^2 \mu^T h & -(D_x h)^T \\ -D_x h & 0 \end{pmatrix} \begin{pmatrix} dx \\ d\mu \end{pmatrix} = - \begin{pmatrix} D_{x\theta}^2 f - D_{x\theta}^2 \mu^T h \\ -D_\theta h \end{pmatrix} d\theta$$

where $D_{xx}^2 f$ is the $n \times n$ matrix of second partial derivatives of f with respect to x, $D_{x\theta}^2 \mu^T h$ is the $n \times s$ matrix of second partial derivatives of $\mu^T h$ with respect to combinations of x and θ, etc.

Whether expressed using matrices or not, this system of equations is a bit unwieldy. Fortunately in most applications, there will be some simplification. For example, if the constraints do not depend on θ, we simply get

$$\underbrace{\frac{dx}{d\theta}}_{\equiv D_\theta x} = (D^2_{xx}f)^{-1}D^2_{x\theta}f.$$

In other situations, the constraints might depend on θ, but s, m, and/or n might just be one or two.

A similar result holds with inequality constraints, except at θ where changing θ changes which constraints bind or do not. Such situations are rare, so we will not worry about them.

Example 4.2 (Production theory)**.** Consider a competitive multiple product firm facing output prices $p \in \mathbb{R}^k$ and input prices $w \in \mathbb{R}^n$. The firm's profits as function of prices is

$$\pi(p, w) = \max_{y,x} p^T y - w^T x \text{ s.t. } y - f(x) \leq 0.$$

The first order conditions are

$$\begin{aligned} p^T - \lambda^T =& 0 \\ -w^T + \lambda^T D_x f =& 0 \\ y - f(x) =& 0 \end{aligned}$$

Total differentiating with respect to p, holding w constant gives

$$\begin{aligned} dp^T - d\lambda^T =& 0 \\ d\lambda^T D_x f + dx^T D^2_{xx}\lambda^T f =& 0 \\ dy - D_x f dx =& 0 \end{aligned}$$

Combining we can get

$$dp^T dy = -dx^T D^2_{xx}\lambda^T f dx.$$

Notice that if we assume the constraint binds and substitute it into the objective function, then the second order condition for this problem is that $v^T D^2_{xx} p^T f v < 0$ for all v. If the second order condition holds, then we must have

$$-dx^T D^2_{xx}\lambda^T f dx > 0.$$

Therefor $dp^T dy > 0$. Increasing output prices increases output.

As an exercise, you could use similar reasoning to show that $dw^T dx < 0$. Increasing input prices decreases input demand.

Appendix A. Notation

- $\mathbb{R}$ is the set of real numbers $\mathbb{R}^n$ is the set of vectors of n real numbers.
- $x \in \mathbb{R}^k$ is read "x in $\mathbb{R}^k$" and means that x is vector of k real numbers.
- $f : \mathbb{R}^n \to \mathbb{R}$ means f is a function from $\mathbb{R}^n$ to $\mathbb{R}$. That is, f's argument is an n-tuple of real numbers and its output is a single real number.
- $U \subseteq \mathbb{R}^n$ means U is a subset of $\mathbb{R}^n$.
- When convenient we will treat $x \in \mathbb{R}^k$ as a $k \times 1$ matrix, so $w^T x = \sum_{i=1}^k w_i x_i$

- $N_\delta(x)$ is a δ neighborhood of x, meaning the set of points within δ distance of x. For $x \in \mathbb{R}^n$, we will use Euclidean distance, so that $N_\delta(x)$ is the set of $y \in \mathbb{R}^n$ such that $\sqrt{\sum_{i=1}^n (x_i - y_i)^2} < \delta$.

Appendix B. Review of derivatives

Partial and directional derivatives were discussed on the summer math review, so we will just briefly restate their definitions and some key facts here.

Definition B.1. Let $f : \mathbb{R}^n \to R$. The ith **partial derivative** of f is

$$\frac{\partial f}{\partial x_i}(x_0) = \lim_{h \to 0} \frac{f(x_{01}, ..., x_{0i} + h, ...x_{0n}) - f(x_0)}{h}.$$

The ith partial derivative tells you how much the function changes as its ith argument changes.

Definition B.2. Let $f : \mathbb{R}^n \to \mathbb{R}^k$, and let $v \in \mathbb{R}^n$ the **directional derivative** in direction v at x is

$$df(x; v) = \lim_{\alpha \to 0} \frac{f(x + \alpha v) - f(x)}{\alpha}.$$

where $\alpha \in \mathbb{R}$ is a scalar.

An important result relating partial to directional derivatives is the following.

Theorem B.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and suppose its partial derivatives exist and are continuous in a neighborhood of x_0. Then*

$$df(x; v) = \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i}(x_0)$$

in this case we will say that f is differentiable at x_0.

It is convenient to gather partial derivatives of a function into a matrix. For a function $f : R^n \to \mathbb{R}$, we will gather its partial derivatives into a $1 \times n$ matrix,

$$Df_x = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

We will simply call this matrix the derivative of f at x. This helps reduce notation because for example we can write

$$df(x; v) = \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i}(x_0) = Df_x v.$$

Similarly, we can define second and higher order partial and directional derivatives.

Definition B.3. Let $f : \mathbb{R}^n \to R$. The ijth **partial second derivative** of f is

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_i}(x_{01}, ..., x_{0j} + h, ...x_{0n}) - \frac{\partial f}{\partial x_i}(x_0)}{h}.$$

Definition B.4. Let $f : \mathbb{R}^n \to \mathbb{R}^k$, and let $v, w \in \mathbb{R}^n$ the **directional derivative** in directions v and w at x is

$$d^2 f(x; v, w) = \lim_{\alpha \to 0} \frac{df(x + \alpha w; v) - df(x; v)}{\alpha}.$$

where $\alpha \in \mathbb{R}$ is a scalar.

Theorem B.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and suppose its first and second partial derivatives exist and are continuous in a neighborhood of x_0. Then*

$$d^2 f(x; v, w) = \sum_{j=1}^{n} \sum_{i=1}^{n} v_i \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) w_j$$

in this case we will say that f is twice differentiable at x_0. Additionally, if f is twice differentiable, then $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_j}$.

We can gather the second partials of f into an $n \times n$ matrix,

$$D^2 f_x = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n}^2 \end{pmatrix},$$

and then write $d^2 f(x; v, w) = w^T D^2 f_x w$. $D^2 f_x$ is also called the Hessian of f.

References

Carter, Michael. 2001. *Foundations of mathematical economics*. MIT Press.

De la Fuente, Angel. 2000. *Mathematical methods and models for economists*. Cambridge University Press.

Dixit, Avinash K. 1990. *Optimization in economic theory*. Oxford University Press Oxford.

Engel, Ernst. 1857. "Die Productions- und Consumptionsverhaeltnisse des Koenigsreichs Sachsen." *Zeitschrift des Statistischen Bureaus des Koniglich Sachsischen Ministeriums des Inneren* (8-9).

Lewbel, Arthur. 2008. "Engel curve." In *The New Palgrave Dictionary of Economics*, edited by Steven N. Durlauf and Lawrence E. Blume. Basingstoke: Palgrave Macmillan. URL http://www.dictionaryofeconomics.com/article?id=pde2008_E000085.

Simon, Carl P and Lawrence Blume. 1994. *Mathematics for economists*. Norton New York.