Mathematics for Economists: Constrained Optimization

Juuso Välimäki

Spring 2020

Contents

1 Analysis

We start by considering the notions of distance, convergence and continuity in a bit more detail. This will help us understand when optimization problems are well posed in the sense that they have optimal solutions.

1.1 Length and distance in $\mathbb{R}^n$

The only spaces that we will be interested in these notes are the various Cartesian products of the real line $\mathbb{R}$ denoted by $\mathbb{R}^n$. The exponent n is also called the dimension of the Euclidean space. Hence an element $x \in \mathbb{R}^n$ is an ordered n-tuple $(x_1, ..., x_n)$ where each $x_i \in \mathbb{R}$.

Distance $d(x, y)$ between two vectors $x, y \in \mathbb{R}^n$ is usually based on the Euclidean norm or the length of a vector in $x \in \mathbb{R}^n$ defined by

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}. \tag{1}$$

This is just the generalization of the Pythagorean theorem to an arbitrary dimension. A distance for $\mathbb{R}^n$ can be derived from this norm as

$$d(x, y) = \|x - y\|.$$

Proposition 1. Let x and y denote points in $\mathbb{R}^n$. Then we have:
(a) $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = \mathbf{0}$,
(b) $\|ax\| = a\|x\|$ for every real a,
(c) $\|x-y\| = \|y-x\|$,
(d) $x \cdot y \leq \|x\| \|y\|$ (Cauchy-Schwarz inequality),
(e) $\|x+y\| \leq \|x\| + \|y\|$ (Triangle inequality).

Remark 1. To see why the Cauchy-Schwarz inequality is true, consider the sum of squares

$$\sum_{i=1}^{n} (x_i + ty_i)^2.$$

This is a quadratic polynomial in t, and as a sum of squares, it is also non-negative. Hence its discriminant is non-positive, i.e.

$$(2\sum_{i=1}^{n} x_i y_i)^2 \leq 4(\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2).$$

Dividing both sides by 4 and taking square roots on both sides gives Cauchy-Schwarz inequality.

This simple result is one of the most important results in all of mathematics. Equality holds in the result if and only if $x = \lambda y$, i.e. x is proportional to y. We have used this observation to argue that the gradient $\nabla f(\hat{x})$ gives the direction of steepest ascent for a function f at point $\hat{x}$.

From Cauchy-Schwarz, we get easily the triangle inequality:

$$\|x+y\|^2 = (x+y)\cdot(x+y) = \|x\|^2 + \|y\|^2 + 2x\cdot y$$
$$\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2.$$

The triangle inequality follows by taking square roots on both sides of the inequality. The inequality in the above expression results from the Cauchy-Schwarz inequality.

Exercise In general, any function $\hat{d}(x,y)$ satisfying (a), (c) and (e) in the above list is a distance. It is a good exercise to show that $\hat{d}(x,y) := \max_i |x_i - y_i|$ is a distance in this sense. Are all the other properties above also satisfied by this distance?

By the segment (a,b) we mean the set of all real number x such that $a < x < b$. By the interval $[a,b]$, we mean the set of all real numbers such that $a \leq x \leq b$. If $a_i < b_i$ for $i = 1, ..., n$, the set of all points $\mathbf{x} = (x_1, ..., x_n)$ in $\mathbb{R}^n$ whose coordinates satisfy $a_i \leq x_i \leq b_i$ for $(1 \leq i \leq n)$, is called an n-cell. If $\mathbf{x} \in \mathbb{R}^n$ and $\varepsilon > 0$, the open (or closed) neighborhood $B^\varepsilon(x)$ with center at x and radius ε is defined to be the set of all $y \in \mathbb{R}^n$, such that $\|y - x\| < (\leq)\, r$.

1.2 Open and closed sets

In this subsection, we give some basic definitions on sets in $\mathbb{R}^n$.

Definition 1. A point x is a limit point of the set $E \subset \mathbb{R}^n$ if every neighborhood of x contains a point $y \in E$ with $y \neq x$.

We say that E is *closed* if every limit point of E is an element of E. A point x is an interior point of E if there is a neighborhood $B^\varepsilon(x)$ of x such that $B^{\varepsilon}(x) \subset E$. We say that E is *open* if every point of E is an interior point.

The *complement* of E, denoted by E^c is the set of all points $x \in \mathbb{R}^n$ such that $x \notin E$.

The set E is *bounded* if there is a real number M such that $\|x\| < M$ for all $x \in E$.

Exercise Is the empty set open or closed? Show that $A = \{x : a < x < b\}$ is an open set and that $A = \{x : a \leq x \leq b\}$ is a closed set. Show that the set $\{1, \frac{1}{2}, \frac{1}{3}, ...\}$ is neither open nor closed (hint: is 0 a limit point? Is it in the set?)

Proposition 2. A set $E \subset \mathbb{R}^n$ is open if and only if its complement is closed. A set $F \subset \mathbb{R}^n$ is closed if and only if its complement is open.

A very important property for sets in mathematical analysis is called *compactness*. we give here a definition of compactness for sets in $\mathbb{R}^n$ that should really be derived as a theorem starting from a more fundamental notion, but for practical matters, this is all we need.

Definition 2 (Compact sets)**.** A set $E \subset \mathbb{R}^n$ is called *compact* if it is closed and bounded.

1.3 Sequences

Definition 3. If S is any set, a sequence in S is a function on the set $\mathbb{N} = \{1, 2, 3, ...\}$ of natural numbers and whose range is in S.

Definition 4. A sequence $\{x_n\}$ in $\mathbb{R}^n$ is said to converge if there is a point $x \in \mathbb{R}^n$ with the following property: For every $\epsilon > 0$, there is an integer N such that $n \geq N$ implies that $d(x_n, x) < \epsilon$.

We say that x_n converges to x, x is the limit of $\{x_n\}$ and we write $x_n \to x$,

$$\lim_{n \to \infty} x_n = x.$$

Theorem 1. Let $\{x_n\}$ be a sequence in $\mathbb{R}^n$.

(i) $\{x_n\}$ converges to $x \in \mathbb{R}^n$ if and only if every neighborhood of x contains all but finitely many of the terms of $\{x_n\}$.

(ii) If $x \in \mathbb{R}^n, x' \in \mathbb{R}^n$, and if $\{x_n\}$ converges to x and to x', then $x = x'$.

(iii) If $\{x_n\}$ converges, then $\{x_n\}$ is bounded.

(iv) If $E \subset \mathbb{R}^n$ and x is a limit point of E, then there is a sequence $\{x_n\}$ in E such that $x = \lim_{n\to\infty} x_n$.

(v) $x_n = (x_{1,n}, ..., x_{k,n}) \to x = (x_1, ..., x_k) \Leftrightarrow x_{i,n} \to x_i$ for all $i \in \{1, ..., k\}$.

The last part of the proposition claims that a sequence of vectors converges if and only if all of its coordinates converge.

Definition 5. Given a sequence $\{x_n\}$, consider an infinite sequence $\{n_k\}$ of positive integers, such that $n_1 < n_2 < \cdots$. Then the sequence $\{x_{n_i}\}$ is called a subsequence of $\{x_n\}$. If $\{x_{n_i}\}$ converges, its limit is called a subsequential limit of $\{x_n\}$.

Exercise Show that if $\{x_n\}$ converges to x, then all of its subsequences also converge to x.

Definition 6. A sequence $\{x_n\}$ is said to be a Cauchy sequence if for every $\epsilon > 0$ there is an integer N such that $d(x_n, x_m) < \epsilon$, if $n \geq N$ and $m \geq N$.

Real numbers are constructed in such a way that Cauchy sequences in $\mathbb{R}$ converge, i.e. have limits in $\mathbb{R}$. By part (v) of the previous theorem, the same is true for real vectors.

Theorem 2 (Weierstrass)**.** Every bounded subset $E \subset \mathbb{R}^n$ with infinitely many elements has a limit point in $\mathbb{R}^n$.

Idea of proof for $\mathbb{R}$: Since E is bounded, it is contained in an interval $[-M, M]$ of length $2M$ for some $M < \infty$. Since E has infinitely many elements, either $[-M, 0]$ or $[0, M]$ or both have infinitely many elements. Hence some interval of length M also contains infinitely many elements of E. Continue this process of halving the interval to show that you can come up with a sequence of intervals of length $2^{-k}M$ containing infinitely many elements of E. The midpoints of the sequences form a Cauchy sequence and hence they converge to a point $x \in \mathbb{R}$. This x is a limit point of E. The same construction generalizes easily to $\mathbb{R}^n$

An immediate consequence of this is the following theorem.

Theorem 3. (Bolzano-Weierstrass Theorem)

Every bounded sequence in $\mathbb{R}^n$ contains a convergent subsequence and every sequence in a compact set $E \in \mathbb{R}^n$ has a convergent subsequence whose limit is in E.

1.4 Continuous functions

Definition 7. Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$. We write $f(x) \to \hat{y}$ as $x \to \hat{x}$, or

$$\lim_{x \to \hat{x}} f(x) = \hat{y}, \tag{2}$$

if there is a point $y \in \mathbb{R}^m$ with the following property: For every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x \in B^{\delta}(\hat{x}) \Rightarrow f(x) \in B^{\varepsilon}(\hat{y}).$$

We say that f is *continuous at* $\hat{x}$ if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x \in B^{\delta}(\hat{x}) \Rightarrow f(x) \in B^{\varepsilon}(f(\hat{x})).$$

Another way of writing this is given in the following simple proposition.

Proposition 3. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\hat{x}$ if for every sequence $\{x_n\}$ that converges to $\hat{x}$, the sequence $\{f(x_n)\}$ converges to $f(\hat{x})$; in symbols,

$$\lim_{n \to \infty} f(x_n) = f\left(\lim_{n \to \infty} x_n\right).$$

A function is said to be continuous if it is continuous at all points in its domain. Continuity of a function f at a point $\hat{x}$ is called a local property of f because it depends on the behavior of f only in the immediate vicinity of $\hat{x}$. A property of f which concerns the whole domain of f is called a global property. Thus, continuity of f on its domain is a global property.

The following proposition gives yet another way of looking at continuity.

Proposition 4. A function f is continuous if and only if the inverse image $f^{-1}(V)$ is open (closed) for every open (closed) set V in Y.

Proposition 5. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$ be continuous functions, and let h be the composite function defined by

$$h(x) = g(f(x)) \qquad \text{for } x \in \mathbb{R}^n.$$

If f is continuous at $\hat{x}$ and if g is continuous at $f(\hat{x})$, then h is continuous at $\hat{x}$.

1.5 Global properties of continuous functions

Definition 8. A function $f : E \to \mathbb{R}$ is said to be bounded if there is a real number M such that $|f(x)| \leq M$ for all $x \in E$.

Recall the definition of the least upper bound and greatest lower bound for a set A of real numbers. We say that $\overline{a}$ is the least upper bound of A if for all $x \in A$, $x \leq \overline{a}$ and for all $a' < \overline{a}$, there is some $x \in A$ such that $x > a'$. Similarly, we say that $\underline{a}$ is the greatest lower bound of A if for all $x \in A$, $x \geq \underline{a}$ and for all $a' > \underline{a}$, there is some $x \in A$ such that $x < a'$.

We write:

$$\overline{a} := \sup A, \underline{a} := \inf A.$$

Theorem 4 (Weierstrass' Theorem)**.** Suppose f is a continuous function on a compact set E, and

$$M = \sup_{x \in E} f(x), \qquad m = \inf_{x \in E} f(x).$$

Then there exists a point $\overline{x}, \underline{x} \in E$ such that $f(\overline{x}) = M$ and $f(\underline{x}) = m$.

Proof. We show this for the supremum. The case for the infimum is analogous. Let $M = \sup_{x \in E} f(x)$. Let $\{m_n\} \to M$ with $m_n < M$ for all n. Then By the definition of supremum, there must be a sequence $\{x_n\} \in E$ with $x_n \geq m_n$. Since E is compact, $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\} \to x \in E$. Since $\{m_n\} \to M$, we also know that $\{m_{n_k}\} \to M$. By continuity of f,

$$M \geq f(x) = \lim f(x_{n_k}) \geq \lim m_{n_k} = M.$$

□

This theorem ensures that our maximization and minimization problems have solutions as long as the objective function is continuous and the feasible set is compact.

Remark 2. To see that E must be closed and bounded and that f has to be continuous, consider the following examples:

1. $f(x) = x$ and $E = \mathbb{R}$.
2. $f(x) = x$ and $E = \{x : 0 < x < 1\}$.
3. $f(x) = x$ for $0 \leq x < 1, f(1) = 0$ and $E = \{x : 0 \leq x \leq 1\}$.

2 Constrained Optimization

We start with some simple examples of constrained optimization to set up expectations for the more general case to follow.

Example 1. Consider finding the maximum for $f(x) = 3 + 2x - x^2$ on the feasible set $F = \{x : -\infty < a \leq x \leq b < \infty$. Since f is continuous and the feasible set F is compact. Therefore Weierstrass' theorem guarantees the existence of a maximizer, i.e. an $x \in F$ such that for all $y \in F$, we have $f(x) \geq f(y)$.

Notice that f is strictly increasing for $x < 1$ and strictly decreasing for $x > 1$. If $a \leq 1 \leq b$, then the function is maximized at its critical point $x = 1$. We say that a direction $(x - x_0)$ is feasible from $x_0 \in F$ if for a small Δ, we have $x_0 + \Delta(x - x_0) \in F$. Linear approximation by the derivative gives:

$$f(x_0 + \Delta(x - x_0)) - f(x_0) = f'(x_0)\Delta(x - x_0).$$

If we have a maximum at x_0, then for all feasible direction

$$f'(x_0)\Delta(x - x_0) \leq 0.$$

If $a < x_0 < b$, then we must have $f'(x_0) = 0$ since both directions $x > x_0, x < x_0$ are feasible. If $f'(x_0) > 0$, then $x > x_0$ cannot be feasible if x_0 is a maximum. Therefore $x_0 = b$ if x_0 is the optimal choice and $f'(x_0) > 0$. Similarly, if $f'(x_0) < 0$ and x_0 is the optimum, then $x_0 = a$.

If all directions are feasible from x_0 and x_0 is a maximum, then just as in the case of unconstrained optimization, we must have $f'(x_0)$. For the other cases, the derivative of the objective function at optimum is closely related to the constraint that binds (i.e. restricts the feasible directions).

In the next subsections, we will generalize our findings to multidimensional optimization problems.

2.1 Optimization with a single equality constraint

We start with local considerations. Let $f : \mathbb{R}^n \to \mathbb{R}$ be the objective function to be maximized and suppose the constraints take the form $g(x) = g(x_1, ..., x_n) = 0$. In other words, $F = \{x : g(x) = 0\}$. We write the maximization problem often as:

$$\max_x f(x)$$

$$\text{subject to } g(x) = 0.$$

A solution to this problem finds a point $\hat{x}$ such that $f(\hat{x}) \geq f(x)$ for all $x \in F$. What can we say about such an $\hat{x}$? At this point, we do not know if it exists. If it exists, and f is differentiable, then for small Δ,

$$f(\hat{x} + \Delta(x - \hat{x})) - f(\hat{x}) = Df(\hat{x})(x - \hat{x})\Delta \leq 0$$

for all feasible directions $(x - \hat{x})$. But how do we know which directions are feasible? Assume that the function g defining the constraint is also differentiable. To find the feasible directions, we go back to implicit function theorem. If $\hat{x} \in F$ and $\frac{\partial g}{\partial x_i}(\hat{x}) \neq 0$ for some $i \in \{1, ..., n\}$, then we can find a write $x_i = h(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) =: h(x_{-i})$ in a neighborhood of $\hat{x}_{-i}$ so that

$$g(h(x_{-i}), x_{-i}) = 0.$$

Notice that it is not possible to use the implicit function theorem if at a critical point of the constraint function. Therefore, we must assume that $Dg(\hat{x}) \neq 0$. We call this the constraint qualification and we will see different versions of this in more complex situations with multiple constraints.

Since the function g is at constant value in the feasible set, we have for all feasible directions $(x - \hat{x})$:

$$\nabla g(\hat{x})(x - \hat{x}) = 0.$$

Notice also that if $(x - \hat{x})$ is feasible, then also $-(x - \hat{x})$ is feasible. From the linear approximation above, this means immediately that for all feasible directions,

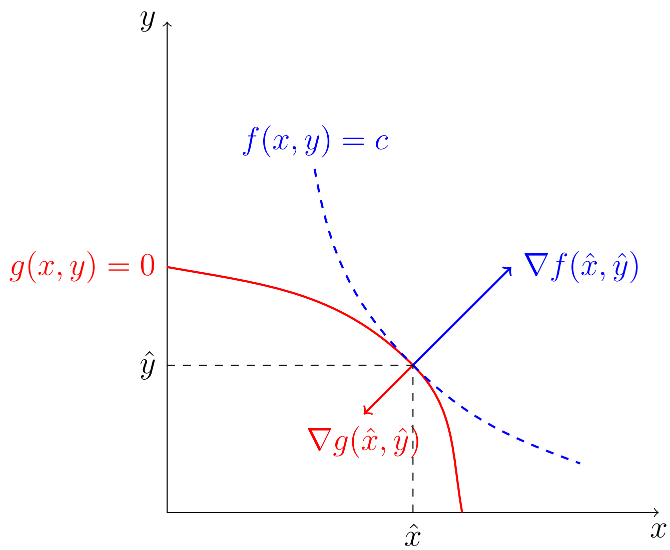
$$Df(\hat{x})(x - \hat{x}) = 0.$$

But therefore we have shown that at optimum $\hat{x}$,

$$\nabla f(\hat{x}) = \mu \nabla g(\hat{x}).$$

We have derived the following necessary condition for a constrained optimum at $\hat{x}$: the gradient of the objective function must be a scalar multiple of the gradient of the constraint function at the optimum. The second requirement is that the choice must be feasible, i.e. $g(\hat{x}) = 0$.

Figure 1: Single equality constraint



2.1.1 The Lagrangean function

The previous discussion motivates the following function that incorporates the constraints into an augmented objective function called the Lagrangean function.

For a constrained optimization problem, we define the following function of $n + 1$ variables:

$$\mathcal{L}(x, \mu) = f(x) - \mu g(x).$$

We call the new variable μ the Lagrange multiplier. We will give it a good economic interpretation later in the course. We are interested in the critical points of this augmented function. Therefore we look for $(\hat{x}, \hat{\mu})$ such that

$$\frac{\partial \mathcal{L}}{\partial x_i}(\hat{x}, \hat{\mu}) = \frac{\partial f}{\partial x_i}(\hat{x}) - \hat{\mu}\frac{\partial g}{\partial x_i}(\hat{x}) = 0 \text{ for all } i,$$
$$\frac{\partial \mathcal{L}}{\partial \mu}(\hat{x}, \hat{\mu}) = g(\hat{x}) = 0.$$

As argued above, these are the first-order conditions for the constrained optimization problem. In order to know if we have found a local maximum or a minimum, we need to look at the second-order Taylor -approximations and the definiteness of the Hessian matrix at the critical point.

As before, write the second-order Taylor approximation to $f : \mathbb{R}^n \to \mathbb{R}$ at $\hat{x}$ as:

$$f(x) = f(\hat{x}) + Df(\hat{x})(x - \hat{x}) + (x - \hat{x}) \cdot D^2(\hat{x})(x - \hat{x}).$$

If $\hat{x}$ is a maximum, then for all feasible directions $(x - \hat{x})$, we have

$$i)Df(\hat{x})(x - \hat{x}) = 0,$$

$$ii)(x - \hat{x}) \cdot D^2(\hat{x})(x - \hat{x}) \leq 0.$$

Since the feasible directions are give by vectors $(x - \hat{x})$ such that

$$\nabla g(\hat{x}) \cdot (x - \hat{x}) = 0,$$

the condition for having a local maximum at $\hat{x}$ is equivalent to checking the negative definiteness of the bordered Hessian where we need

1. The Lagrangean $\mathcal{L}$
2. The Equality constraint h

To get the bordered Hessian, start with the derivative of the Lagrangean with respect to the choice variables x at the critical point $\hat{x}$: $D_x^2\mathcal{L}(\hat{x})$ and 'border' it with the derivative of the constraint function (to capture the restriction to feasible directions).

$$D^2\mathcal{L} = \begin{bmatrix} 0 & Dg(\hat{x}) \\ [Dg(\hat{x})]^T & D_x^2\mathcal{L}(\hat{x}) \end{bmatrix}$$

In the special case where we have only two choice variables, I let the variables be x, y for notational ease, we need to examine

$$D^2\mathcal{L} = \begin{bmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{bmatrix}$$

How do we determine the negative definiteness of the bordered Hessian?

1. Leading principal minors must alternate in sign [1]
2. $\det D^2\mathcal{L}(\hat{x})$ must have the same sign as $(-1)^n$.

How many principal minors to examine?

- You need to check the sign of the last $(n-1)$ leading principal minors
- For completeness, I state here that with more constraints, you need to border the Hessian with the derivatives of all binding constraints. If you have k such constraints, then you need to examine the sign of $(n-k)$ leading principal minors.

Bordered Hessians are a bit of a nightmare for me. They are tedious to compute and they tell nothing of significance in the end. We will see later how we can bypass them to a large extent by concentrating on a subset of problems where first-order conditions also turn out to be sufficient.

In any case, here is eventually a concrete example:

Example 2. Find the minima and maxima of $f(x,y,z) = x + y + z^2$ subject to constraints

$$\begin{aligned} x^2 + y^2 + z^2 &= 1 \\ y &= 0 \end{aligned}$$

Start by substituting the second constraint to the objective function and the first constraint to get $f(x,z) = x + z^2$, and

$$x^2 + z^2 = 1$$

Form the Lagrangean

$$\mathcal{L}(x,z,\mu) = x + z^2 - \mu(x^2 + z^2 - 1)$$

Differentiate to get the first-order conditions (FOC):

[1] Recall that a leading principal minor of k^{th} order is obtained from a matrix A by deleting its last k rows and columns.

$$\frac{\partial \mathcal{L}}{\partial x} = 1 - 2\mu x = 0 \tag{3}$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2z - 2\mu z = 0 \tag{4}$$
$$\frac{\partial \mathcal{L}}{\partial \mu} = 1 - x^2 - z^2 = 0 \tag{5}$$

The second FOC gives:

$$z(2 - 2\mu) = 0$$

Therefore either $z = 0$, or $\mu = 1$.

Consider first the possibility that $z = 0$. In that case, (5) implies that $x = \pm 1$. We get two critical points from (3):

$$\left(x = 1, y = 0, z = 0, \mu = \tfrac{1}{2}\right) \text{ and } \left(x = -1, y = 0, z = 0, \mu = -\tfrac{1}{2}\right)$$

If $\mu = 1$, (3) implies that $x = \frac{1}{2}$. By substituting into (5) we get the critical points:

$$\left(x = \tfrac{1}{2}, y = 0, z = \tfrac{\sqrt{3}}{2}, \mu = 1\right) \text{ and } \left(x = \tfrac{1}{2}, y = 0, z = -\tfrac{\sqrt{3}}{2}, \mu = 1\right)$$

As a result, we have four critical points for the Lagrangean. Draw the constraint set and level curves for the objective function to get a guess of the classification of the critical points.

By examining the bordered Hessian, we see that $\left(x = -1, y = 0, z = 0, \mu = -\frac{1}{2}\right)$ and $\left(x = 1, y = 0, z = 0, \mu = \frac{1}{2}\right)$ are local minima, and $\left(x = \frac{1}{2}, y = 0, z = \pm\frac{\sqrt{3}}{2}, \mu = 1\right)$ are local maxima.

Can you show the existence of a maximum? Which of the local maxima is the global maximum?

2.2 Multiple equality constraints

Consider next the case, where we have k equality constraints $g(x) = (g_1(x), ..., g_k(x) : \mathbb{R}^n \to \mathbb{R}^k$. In this case, we have the problem:

$$\max_x f(x)$$
$$\text{subject to } g_1(x) = 0,$$
$$g_2(x) = 0,$$
$$\vdots$$
$$g_k(x) = 0.$$

Form the Lagrangean now with k constraints as a function of $n + k$ variables:

$$\mathcal{L}(x, \mu_1, ..., \mu_k) = f(x) - \sum_{j=1}^{k} \mu_j g_j(x).$$

We can proceed exactly as before to use the linear approximations to characterize the feasible directions from $\hat{x}$ as $\{(x - \hat{x}) : Dg(\hat{x})(x - \hat{x})\} = 0$. Since the objective function cannot increase at the optimum in any feasible direction, we have that

$$Df(\hat{x})(x - \hat{x}) = 0 \text{ whenever } Dg(\hat{x})(x - \hat{x}) = 0.$$

If $Dg(\hat{x})$ has full rank, then this is equivalent to requiring that $Df(\hat{x})$ and $Dg_j(\hat{x})$ must be linearly dependent. Since we assume that $Dg(\hat{x})$ has full rank, this means that there must exist $(\mu_1, ..., \mu_k)$ such that

$$\nabla f(\hat{x}) = \sum_{j=1}^{k} \mu_j \nabla g_j(\hat{x}).$$

Hence we can summarize the three necessary conditions for local maximum:

i) Gradient alignment: $\nabla f(\hat{x}) = \sum_{j=1}^{k} \mu_j \nabla g_j(\hat{x})$,

ii) Constraint holds: $g(\hat{x}) = 0$,

iii) Constraint qualification: $Dg_1(\hat{x}), ..., Dg_k(\hat{x})$ are linearly independent.

The first two can be achieved by requiring that $(\hat{x}, \hat{\mu}_1, ..., \hat{\mu}_k)$ be a critical point of the Lagrangean. The second-order conditions are based on bordered Hessian matrices as explained at the end of the previous subsection.

Let's end this section with another example

Example 3.

Consider the objective function

$$f(x, y, z) = xz + yz$$

and a maximization problem subject to:

$$g_1(x, y, z) = y^2 + z^2 - 1$$
$$g_2(x, y, z) = xz - 3$$

1. Find the critical points of f subject to constraints $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$.

2. How would you determine which of the critical points are local minima and which are local maxima? What is the bordered Hessian that you would use?

1. Find first the critical points of the Lagrangean

$$\mathcal{L}(x, y, z, \mu_1, \mu_2) = xz + yz - \mu_1(y^2 + z^2 - 1) - \mu_2(xz - 3)$$

First-order conditions:

$$\frac{\partial \mathcal{L}}{\partial x} = z - \mu_2 z = 0 \tag{6}$$
$$\frac{\partial \mathcal{L}}{\partial y} = z - 2\mu_1 y = 0 \tag{7}$$
$$\frac{\partial \mathcal{L}}{\partial z} = x + y - 2\mu_1 z - \mu_2 x = 0 \tag{8}$$
$$\frac{\partial \mathcal{L}}{\partial \mu_1} = y^2 + z^2 - 1 = 0 \tag{9}$$
$$\frac{\partial \mathcal{L}}{\partial \mu_2} = xz - 3 = 0 \tag{10}$$

We need to solve this system of equations to find the critical points. Start with (6), giving

$$z(1-\mu_2) = 0, \Leftrightarrow \; z = 0 \text{ or } \mu_2 = 1$$

If $z = 0$, then (10) is not true for any x and as a result, we must have$z \neq 0$. Therefore, we can only have $\mu_2 = 1$ as a candidate solution. The second FOC (7) gives

$$y - 2\mu_1 z = 0, \; \Leftrightarrow y = \frac{z}{2\mu_1}.$$

Plug in the solutions for y and μ_2 into (8) :

$$\frac{z}{2\mu_1} - 2\mu_1 z = 0 \;\Leftrightarrow\; z\left(\frac{1}{2\mu_1} - 2\mu_1\right) = 0.$$

We already know that $z \neq 0$, and therefore

$$\frac{1}{2\mu_1} - 2\mu_1 = 0 \;\Leftrightarrow\; 4\mu_1^2 = 1 \;\Leftrightarrow\; \mu_1 = \pm\frac{1}{2}$$

We have now solved for possible Lagrange multipliers μ_1 ja μ_2, i.e. we have:

$$\mu_1 = \pm\tfrac{1}{2} \text{ and } \mu_2 = 1$$

To get the values of the choice variables, plug in the values of the multipliers into (8) to get:

$$y = \pm z.$$

Substituting into (9), we get (by squaring):

$$2z^2 - 1 = 0 \;\Leftrightarrow z = \pm\frac{1}{\sqrt{2}}$$

The fifth FOC (10) gives:

$$x = \frac{3}{z},$$

or $x = 3\sqrt{2}$ if $z = \frac{1}{\sqrt{2}}$ and $x = -3\sqrt{2}$ if $z = -\frac{1}{\sqrt{2}}$. We have now found all that we need for the critical points of f subject to the constraints. If $z = \frac{1}{\sqrt{2}}$, then $x = 3\sqrt{2}$, $y = \pm z$. This yields two critical points (x, y, z):

$$1 : \left(3\sqrt{2}, \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right)$$

$$2 : \left(3\sqrt{2}, -\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right)$$

If $z = -\frac{1}{\sqrt{2}}$, then $x = -3\sqrt{2}$, $y = \pm z$. This gives also two critical points (x, y, z):

$$3 : \left(-3\sqrt{2}, -\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}\right)$$

$$4 : \left(-3\sqrt{2}, \tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}\right)$$

We know that for all critical points, $\mu_2 = 1$, and we can check the sign of μ_1 from FOC (7). After this, we have all the critical points of the problem as:

Critical points for the problem (x, y, z, μ_1, μ_2):

$$
\begin{aligned}
1 &: \left(3\sqrt{2}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{2}, 1\right)\\
2 &: \left(3\sqrt{2}, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{2}, 1\right)\\
3 &: \left(-3\sqrt{2}, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \frac{1}{2}, 1\right)\\
4 &: \left(-3\sqrt{2}, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, -\frac{1}{2}, 1\right)
\end{aligned}
$$

We can plug these into the objective function to see which of the critical points could be the true maximum. Do we know now that the problem has a maximum?

2. In order to determine which of the critical points are maxima or minima, we need to consider the bordered Hessian by bordering the Hessian of $\mathcal{L}$ with respect to (x, y, z), i.e. $D_x^2 L$ with the derivatives of the constraints $D_x g_1$ and $D_x g_2$.

$$
D^2\mathcal{L} = \begin{bmatrix} 0 & Dg(\hat{x}) \\ [Dg(\hat{x})]^T & D_x^2\mathcal{L} \end{bmatrix}
$$

With three dimensional choice variable, we know that $D_x^2 L$ is a $[3 \times 3]$ -matrix. With two constraints $Dg(\hat{x})$ is a $[2 \times 3]$ -matrix. Hence we have a (5×5) square matrix for our bordered Hessian:

$$
D^2 L = \begin{bmatrix} 0 & 0 & g_{1x} & g_{1y} & h_{1z} \\ 0 & 0 & g_{2x} & g_{2y} & g_{2z} \\ g_{1x} & g_{2x} & \mathcal{L}_{xx} & \mathcal{L}_{xy} & \mathcal{L}_{xz} \\ g_{1y} & g_{2y} & \mathcal{L}_{yx} & \mathcal{L}_{yy} & \mathcal{L}_{yz} \\ g_{1z} & g_{2z} & \mathcal{L}_{zx} & \mathcal{L}_{zy} & \mathcal{L}_{zz} \end{bmatrix},
$$

where we have denoted the partial derivatives by g_{ix}, and second-order partial derivatives by of the Lagrangean by L_{xx} etc. As you can imagine, it is not a pure pleasure to check the definiteness of this bordered Hessian for the four critical points (even though you have to compute a single determinant).

3 Optimization with inequality constraints

The most important class of optimization problems in economics considers maximizing (or minimizing) an objective function subject to k inequality constraints. In these problems, the feasible set takes the form

$$F = \{x \in \mathbb{R}^n : h(x) \leq 0\},$$

where $h(x) \leq 0$ can written more fully as:

$$\begin{array}{rcl} h_1(x_1, ..., x_n) & \leq & 0 \\ \vdots & & \vdots \\ h_k(x_1, ..., x_n) & \leq & 0 \end{array}$$

Notice that we can incorporate equality constrains into these problems since $\{x \in \mathbb{R}^n : g(x) = 0\}$ is the same set as $\{x \in \mathbb{R}^n : g(x) \leq 0, \quad -g(x) \leq 0\}$.

subsectionKuhn-Tucker or Karush-Kuhn-Tucker first-order conditions

We shall concentrate on the first-order conditions for an optimum. From this point on, we restrict attention to i) maximization problems where the objective function is quasiconcave and ii) minimization problems where the objective function is quasiconvex. For each of these cases, we assume that the constraint functions h_j are quasiconvex so that the feasible set that is given as the intersection of lower level sets of these functions is convex. This restriction on the constraint functions means that the only types of equality constraints that are allowed are affine linear constraints (since the only functions of $n > 1$ variables that are both quasiconvex and quasiconcave are affine linear functions (i.e. linear plus a constant vector).

Under these conditions, any point satisfying the first-order conditions is a global optimum if the derivative of the objective function at the point in question is non-zero.

We say that an inequality constraint $h_j(x_1, ..., x_n) \leq 0$ is binding at $\hat{x}$ if $h_j(\hat{x}) = 0$. If $h_j(\hat{x}) < 0$, then we say that the constraint is not binding. A non-binding constraint does not restrict the feasible directions for small changes in $\hat{x}$. For binding constraints $h_j(\hat{x})$, the feasible directions Δx are given again by:

$$Dh_j(\hat{x})\Delta x \leq 0.$$

Hence the binding constraints are just like the equality constraints that we discussed in the previous section. Non-binding constraints can be ignored. The problem in general is that we do not know a priori which constraints are binding and which are not.

Let's write the Lagrangean function for the optimization problem as before:

$$\mathcal{L}(x, \lambda_1, ..., \lambda_k) = f(x) - \sum_{j=1}^{k} \lambda_j h_j(x).$$

I have adopted the notation for the textbook to denote the Lagrange multipliers in inequality constrained problems by λ_j. If a constraint is not binding, it can be ignored in the problem. If it binds, then it cannot be ignored. But both of these cases are incorporated in the following complementary slackness condition. For all j, we have:

$$\lambda_j h_j(\hat{x}_1, ..., \hat{x}_n) = 0.$$

This simply says that if $h_j(/hatx) < 0$, then $\lambda_j = 0$ and the constraint vanishes from the Lagrangean. If the constraint binds, then $h_j(\hat{x}) = 0$ and the complementary slackness is also satisfied.

Based on these considerations, we formulate the first order conditions for $(\hat{x}, \hat{\lambda})$ as follows. We consider a point where the constraint qualification holds (i.e. the derivatives of the binding constraints are linearly independent so that we can use implicit function theorem). [2]

The first-order conditions for the problem also known as the Kuhn-Tucker or Karush-Kuhn-Tucker conditions for the problem are given by:

$$\frac{\partial \mathcal{L}}{\partial x_i}(\hat{x}, \hat{\lambda}) = \frac{\partial f}{\partial x_i}(\hat{x}) - \sum_{j=1}^{k} \hat{\lambda}_j \frac{\partial h_j}{\partial x_i}(\hat{x}) = 0 \text{ for all } i,$$

$$\hat{\lambda}_j h_j(\hat{x}) = 0 \text{ for all } j \in \{1, ..., k\},$$

[2] I note here that for the case of quasiconvex constraint functions, a sufficient condition for constraint qualification is that the feasible set has an interior point.

$$\hat{\lambda}_j \geq 0 \text{ for all } j \in \{1, ..., k\},$$
$$h_j(\hat{x}) \leq 0 \text{ for all } j \in \{1, ..., k\}.$$

Let me sum up: at the optimal point $\hat{x}$, we need i) the usual first-order condition for the Lagrangean with respect to the choice variables. ii) we need that $\hat{x}$ be feasible, i.e. $h_j(\hat{x}) >\leq 0$ for all j, iii) the complementary slackness conditions, and the non-negativity of the multipliers.

We have not discussed the non-negativity of the multipliers yet, but it is easy to see why this must be true in the case of a single inequality constraint. Assume constraint qualification, i.e. $Dh(\hat{x}) \neq 0$By the first order conditions with respect to the x_i, we see that as before,

$$\nabla h(\hat{x}) = \lambda \nabla f(\hat{x}).$$

If the multiplier was strictly negative at an optimal point $\hat{x}$, where the constraint binds, then

$$Dh(\hat{x}) \nabla f(\hat{x}) = \lambda \nabla h(\hat{x}) \cdot \nabla h(\hat{x}) \leq 0.$$

Hence movement in the direction of the fastest increase of f is feasible and $\hat{x}$ cannot be an optimum unless $\nabla f(\hat{x}) = 0$. But in this case, $\lambda = 0$ since $\nabla h(\hat{x}) \neq 0$ by constraint qualification.

The general case for the positive sign of the multipliers is proved using either separating hyperplane theorem or Farkas' Lemma and it is left for future studies. The following picture gives you an idea why the gradient of the objective function must be a positive combination of the gradients of the constraint functions.

3.1 Concave programming

We are now ready to see why the first-order conditions are sufficient for maxima of quasiconcave functions with a non-vanishing derivative on a convex set. Recall from Lecture 6 that a differentiable function f on a convex set X is quasiconcave if and only if for all $x, y \in X$:

$$f(y) \geq f(x) \Rightarrow Df(x)(y - x) \geq 0.$$

This implies the following (almost converse) result:

Figure 2: Single inequality constraint

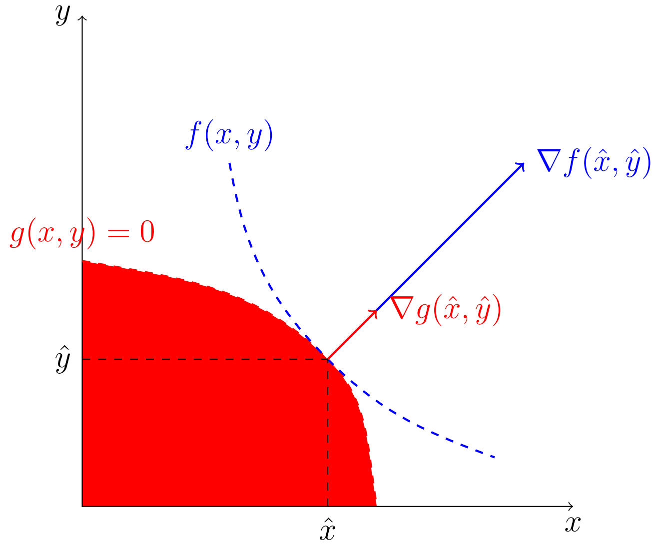
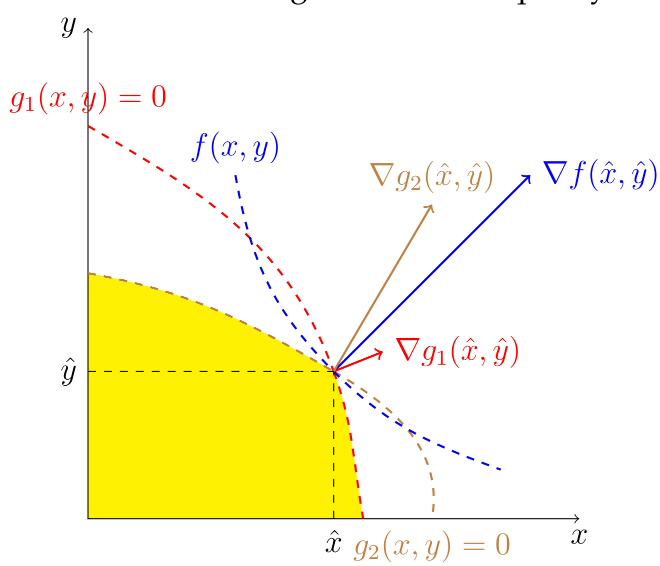


Figure 3: Two inequality constraints



Proposition 6. Suppose $Df(x)$ is non-zero for all $x \in X$ and f is quasiconcave on X. Then $\hat{x}$ is a global maximum for f on X if $Df(\hat{x})(y - \hat{x}) \leq 0$ for all $y \in X$

Theorem 5. Suppose that f is quasiconcave and $Df(x) \neq 0$ on a the convex set $X = \{x \in \mathbb{R}^n : h_j(x) \leq 0$ for $j \in \{1, ..., k\}$, where each $h_j(x)$ is a quasicinvex function. Then any point satisfying the first-order conditions is a global maximum for f on X.

Proof. Write the first-order condition with respect to x as:

$$Df(\hat{x}) - \sum_{j=1}^{k} \hat{\lambda}_j Dh_j(\hat{x}) = 0. \tag{11}$$

Multiply on the right by $(y - \hat{x})$ to get

$$Df(\hat{x})(y - \hat{x}) - \sum_{j=1}^{k} \hat{\lambda}_j Dh_j(\hat{x})(y - \hat{x}) = 0. \tag{12}$$

For feasible directions for binding constraints, we have $Dh_j(\hat{x})(y-\hat{x}) \leq 0$ since each h_j is assumed to be quasiconvex. For nonbinding constraints, $\hat{\lambda}_j = 0$. Therefore since $\hat{\lambda}_j \geq 0$ for all j, we have

$$\hat{\lambda}_j Dh_j(\hat{x})(y - \hat{x}) \leq 0 \text{ for all } j.$$

Thus by equation (12), we see that

$$Df(\hat{x})(y - \hat{x}) \leq 0$$

for all feasible y. Therefore by the proposition above, $f(\hat{x}) \geq f(y)$ for all feasible y. □

Before getting into economic applications proper, let's conclude this section with a couple of numerical examples demonstrating how to find constrained maxima.

Example 4. Maximize the objective function $f(x, y, z) = xyz + z$, subject to

$$
\begin{aligned}
x^2 + y^2 + z &\leq 6 \\
x &\geq 0 \\
y &\geq 0 \\
z &\geq 0.
\end{aligned}
$$

1. Find the points that satisfy the first-order condition
2. Investigate whether the constraint $x^2 + y^2 + z \leq 6$ binds at optimum.
3. Find a point satisfying the first order conditions with $x = 0$.

Let's do this mechanically and think about what we do only afterwards.

1. From the Lagrangean:

$$\mathcal{L}(x, y, z, \lambda_i) = xyz + z - \lambda_1 \left[x^2 + y^2 + z - 6\right] + \lambda_2 x + \lambda_3 y + \lambda_4 z$$

   The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial x} = yz - 2\lambda_1 x + \lambda_2 = 0 \tag{13}$$
$$\frac{\partial \mathcal{L}}{\partial y} = xz - \lambda_1 + \lambda_3 = 0 \tag{14}$$
$$\frac{\partial \mathcal{L}}{\partial z} = xy + 1 - \lambda_1 + \lambda_4 = 0 \tag{15}$$
$$\lambda_1 \left[x^2 + y^2 + z - 6\right] = 0 \tag{16}$$
$$\lambda_2 x = 0 \tag{17}$$
$$\lambda_3 y = 0 \tag{18}$$
$$\lambda_4 z = 0 \tag{19}$$
$$x^2 + y^2 + z \leq 6 \tag{20}$$
$$-x \leq 0 \tag{21}$$
$$-y \leq 0 \tag{22}$$
$$-z \leq 0 \tag{23}$$
$$\lambda_i \geq 0 \;\; i \in \{1, 2, 3, 4\} \tag{24}$$

2. If at optimum, $\lambda_1 = 0$, i.e. the first constraint is not binding, we get from (15):

$$xy + 1 + \lambda_4 = 0$$

This is not possible for any feasible (x, y), since $\lambda_4 \geq 0$, implying that either x or y must be negative. By the non-negativity constraints, we conclude that

$$\lambda_1 > 0,$$

and $x^2 + y^2 + z \leq 6$ binds at the optimum.

3. Find a critical point with $x = 0$:

$$\frac{\partial \mathcal{L}}{\partial x} = yz + \lambda_2 = 0 \tag{25}$$
$$\frac{\partial \mathcal{L}}{\partial y} = -\lambda_1 + \lambda_3 = 0 \tag{26}$$
$$\frac{\partial \mathcal{L}}{\partial z} = 1 - \lambda_1 + \lambda_4 = 0 \tag{27}$$
$$\lambda_1 \left[y^2 + z - 6\right] = 0 \tag{28}$$
$$\lambda_2 x = 0 \tag{29}$$
$$\lambda_3 y = 0 \tag{30}$$
$$\lambda_4 z = 0 \tag{31}$$
$$y^2 + z \leq 6 \tag{32}$$
$$-x \leq 0 \tag{33}$$
$$-y \leq 0 \tag{34}$$
$$-z \leq 0 \tag{35}$$
$$\lambda_i \geq 0 \;\; i \in \{1, 2, 3, 4\} \tag{36}$$

By part b), we know that $\lambda_1 > 0$. The second FOC gives $\lambda_3 = \lambda_1$. By the non-negativity of y, we get $y = 0$. The first FOC then requires that $\lambda_2 = 0$, and therefore constraint (32) must bind at optimum. This yields $z = 6$. We have therefore found the point $(x = 0, y = 0, z = 6)$ and $(\lambda_1 = 1, \lambda_2 = 0, \lambda_3 = 0, \lambda_4 = 0)$ satisfying the first-order conditions and $x = 0$.

Can you show that this is a local maximum (do not compute the bordered Hessian but argue directly using the directional derivative in all feasible directions)? Can you find another local maximum that gives a higher value to the objective function? (Use symmetry to argue that at any local maximum, $\hat{x} = \hat{y}$ reducing the problem essentially to a bivariate problem. look for interior solutions).

The next problem is to find a maximum in a problem that is really a consumer optimization problem without the economics terminology.

Example 5. Maximize

$$f(x, y) = \alpha x + \sqrt{y}$$

subject to

$$\begin{aligned} px + y &\leq 1, \\ x &\geq 0, \\ y &\geq 0. \end{aligned}$$

Let's assume that $p > 0$ and let's find interior solutions, i.e. $(\hat{x}, \hat{y}) > 0$. Are there other kinds of solutions?

Form the Lagrangean:

$$\mathcal{L}(x, y, \lambda_i) = \alpha x + \sqrt{y} - \lambda_1 [px + y - 1] + \lambda_2 x + \lambda_3 y$$

First-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial x} = \alpha - \lambda_1 p + \lambda_2 = 0 \tag{37}$$
$$\frac{\partial \mathcal{L}}{\partial y} = \frac{1}{2} y^{-\frac{1}{2}} - \lambda_1 + \lambda_3 = 0 \tag{38}$$
$$\lambda_1 \left[px + y - 1 \right] = 0 \tag{39}$$
$$\lambda_2 x = 0 \tag{40}$$
$$\lambda_3 y = 0 \tag{41}$$
$$px + y \leq 1 \tag{42}$$
$$-x \leq 0 \tag{43}$$
$$-y \leq 0 \tag{44}$$
$$\lambda_i \geq 0 \;\; i \in \{1, 2, 3\} \tag{45}$$

If $x, y > 0$, we have $\lambda_2, \lambda_3 = 0$:
First-order conditions give:

$$\lambda_1 = \frac{1}{2} y^{-\frac{1}{2}} \Leftrightarrow \lambda_1 > 0$$

Substitute the obtained multiplier into (37) to get:

$$\alpha - \frac{1}{2} y^{-\frac{1}{2}} p = 0 \Leftrightarrow y^* = \frac{p^2}{4\alpha^2}$$

Since $\lambda_1 > 0$, we have $px + y = 1$ at optimum. This gives:

$$x^* = \frac{4\alpha^2 - p^2}{4p\alpha^2}$$

Note that this solution is valid only if $2\alpha \geq p$. Constraint qualification holds since the derivative of the binding constraint is nonzero at optimum: $D(h_1(x, y)) = \begin{bmatrix} p & 1 \end{bmatrix}$. If $p > 2\alpha$, then the optimum is a corner solution. From (38), we see that at any optimum, $y > 0$. Therefore the only other possibility is that $(\hat{x}, \hat{y})$ = (0,1). What is the value of λ_2 in this case?

4 Economic applications

4.1 Utility maximization problem (UMP)

A consumer allocates her budget of $w > 0$ to n goods. Her consumption vector is an element of the positive orthant of the n Euclidean space $X = \{x \in \mathbb{R}^n_+\}$. We assume that the consumer has a continuous utility function $u(x)$ defined on X. Economic scarcity is present through the budget constraint:

$$p \cdot x \leq w \text{ or } \sum_{i=1}^{n} p_i x_i \leq w,$$

where $p = (p_1, ..., p_n) > 0$ is the vector of strictly positive prices for the goods.

We can write this problem then as

Maximize

$$u(x_1, ..., x_n)$$

subject to

$$\sum_{i=1}^{n} p_i x_i \leq w,$$

$$x_i \geq 0 \text{ for all } i.$$

By writing the constraints in the equivalent form:

subject to

$$\sum_{i=1}^{n} p_i x_i - w \leq 0,$$

$$-x_i \leq 0 \text{ for all } i,$$

the problem is in the standard form that we always write for inequality constrained optimization problems.

Let's pause to see what we know about this problem already. To see that the feasible set is bounded, let $p^{min} = \min_j p_j$ (i.e. one of the smallest prices p_j). Then we know that for all feasible x, we have $p_i x_i \leq w$ for all i since $x_i \geq 0$ and $p_i > 0$ for all i. Therefore for all feasible x, $x_i \leq \frac{w}{p^{min}}$ for all i. In other words, the feasible set is bounded since $0 \leq x_i \leq \frac{w}{p^{min}}$ for all i.

To see that the feasible set is closed, we need to show that all limit points of the feasible belong to the feasible set. We show this by arguing that when y is not in the feasible set, it is not a limit point. If y is not feasible, then either $y_i < 0$ for some i or $\sum_i p_i y_i > w$. In the first case, if $y_i = -a$ for some $a > 0$, then no $B^\varepsilon(y)$ contains any feasible point for $\varepsilon < a$. Hence y is not a limit point. For the second case $\sum_i p_i y_i - w = b$ for some $b > 0$, let $p^{max} = \max_j p_j$. Then no point in $B^\varepsilon(y)$ for $\varepsilon < \frac{b}{p^{max}}$ is feasible. Hence y is not a limit point.

Remark 3. You do not have to prove this closedness property every time. It is sufficient to note that $\{x \in \mathbb{R}^n : g(x) \leq 0\}$ is a closed set whenever g is continuous. In our case here, all g_k are linear and therefore continuous. For multiple constraints, just observe that intersections of closed sets are closed.

Hence we know by Weierstrass' theorem that a maximum exists. Since the constraint functions are linear, the feasible set is convex. If u is strictly increasing (as we usually assume) and quasiconvex, then the first order Kuhn-Tucker conditions are necessary and sufficient for optimum. In words, whenever we find a point satisfying the K-T conditions, we have solved the problem.

Let's turn our attention next to the Lagrangean and the K-T conditions:

$$\mathcal{L}(x, \lambda) = u(x) - \lambda_0 \left[\sum_{i=1}^{n} p_i x_i - w \right] + \sum_{i=1}^{n} \lambda_i x_i$$

The first-order K-T conditions are:

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial u}{\partial x_i} - \lambda_0 p_i + \lambda_i = 0 \text{ for all } i, \tag{46}$$

$$\lambda_0 \left[\sum_{i=1}^{n} p_i x_i - w \right] = 0, \tag{47}$$

$$\lambda_i x_i = 0 \text{ for all } i, \tag{48}$$

$$\sum_{i=1}^{n} p_i x_i - w \leq 0, \tag{49}$$

$$-x_i \leq 0 \text{ for all } i, \tag{50}$$

$$\lambda_i \geq 0 \;\; i \in \{0, 1, ..., n\}. \tag{51}$$

If the utility function has a strictly positive partial derivative for some x_i at the optimum, then the budget constraint must bind and $\lambda_0 > 0$. This follows immediately from the first line of the K-T conditions. For the other inequality constraints, consider the partial derivatives at $x \in X$ with $x_i \to 0$ for some i. If

$$\lim_{x_i \to 0} \frac{\partial u}{\partial x_i} = \infty,$$

then we know again from the first line of the K-T conditions that at optimum $x_i > 0$. (Intuitively, if you have an infinitely large marginal utility for some good, you would want to consume more of it). If this is true for all i, then we can ignore the non-negativity constraints and we are effectively back to a problem with a single equality constraint.

If $\frac{\partial u}{\partial x_i} < \infty$ for $x = (x_i, x_{-i}) = (0, x_{-i})$, then we must also consider corner solutions where $x_i = 0$ at optimum. For interior solutions, we get from the first equation by eliminating λ the familiar condition:

$$\frac{\frac{\partial u}{\partial x_i}}{\frac{\partial u}{\partial x_k}} = \frac{p_i}{p_k}. \tag{52}$$

This is of course the familiar requirement that $MRS_{x_i,x_k} = \frac{p_i}{p_k}$ that we saw in Principles of Economics 1. Now we see that the same condition extends for many goods and the economic intuition is exactly the same. The price ration gives the marginal rate of transformation between the different goods and at an interior optimum, that rate must coincide with the marginal rate of substitution.

In many cases, the functional form is such that the problem can be solved explicitly. For example, with constant elasticity of substitution, where

$$u(x) = (a_1 x_1^\rho + \cdots + a_n x_n^\rho)^{\frac{1}{\rho}},$$

for $\rho > 1$. You have already shown in problem sets that functions of this type are quasiconcave. Compute the marginal utility for each x_i:

$$\frac{\partial u}{\partial x_i} = \rho a_1 x_i^{\rho-1} \frac{1}{\rho} (a_1 x_1^\rho + \cdots + a_n x_n^\rho)^{\frac{1}{\rho}1}.$$

Note that since $\rho < 1$, we have $\frac{\partial u}{\partial x_i} > 0$, and

$$\lim_{x_i \to 0} \frac{\partial u}{\partial x_i} = \infty.$$

Since the feasible set is convex and the objective function is quasiconvcave with a non-vanishing derivative, the first order conditions are necessary and sufficient for optimum. Since the marginal utility is unbounded at the boundary, we know that we have an interior solution and that the budget constraint is binding. Hence the K-T conditions require simply that for all i, k:

$$\frac{\frac{\partial u}{\partial x_i}}{\frac{\partial u}{\partial x_k}} = \frac{p_i}{p_k},$$

and the budget constraint holds as an equality:

$$\sum_{i=1}^n p_i x_i = 0.$$

Hence we have that

$$\frac{a_1 x_1^{\rho-1}}{a_k x_k^{\rho-1}} = \frac{p_1}{p_k},$$

or

$$\frac{x_1}{x_k} = (\frac{a_k p_1}{a_1 p_k})^{\frac{1}{\rho-1}},$$

or

$$x_k = x_1(\frac{a_k p_1}{a_1 p_k})^{\frac{1}{1-\rho}}. \tag{53}$$

Substituting into the budget constraint, we get:

$$p_1 x_1 + \sum_{k=2}^{n} p_k x_1 (\frac{a_k p_1}{a_1 p_k})^{\frac{1}{1-\rho}} = w.$$

We can solve for x_1 to get

$$x_1 = \frac{w}{p_1 + \sum_{k=2}^{n} p_k (\frac{a_k p_1}{a_1 p_k})^{\frac{1}{1-\rho}}}.$$

Substituting into (53), we can solve the other x_j.

To get a bit nicer expression, let $r = \frac{\rho}{\rho - 1}$ and assume that $a_i = 1$ for all i. Then we have for each j:

$$x_j = \frac{w p_j^{r-1}}{\sum_{k=1}^{n} p_k^r}.$$

In this case, we are able to solve the optimal demands as explicit functions of the exogenous variables. We call the optimal solutions to the utility maximization problem the Marshallian demands. You will see these demand functions for CES utility functions in your future studies in models of monopolistic competition as needed in growth theory, international trade and industrial organization.

If you want to understand where the name constant elasticity of substitution comes from, you should note that

$$\frac{x_i}{x_k} = (\frac{p_i}{p_k})^{\frac{1}{\rho-1}} (\frac{a_k}{a_i})^{\frac{1}{\rho-1}}.$$

Hence a small percentage change in the price ratio between any two goods induces the same percentage change in the optimal consumptions. The size of this change is given by $\frac{1}{\rho-1}$ and hence ρ measures the elasticity of substitution between any two goods. The higher, ρ, the higher the substitution away from a good when its price increases.

You should consider the comparative statics of the optimal demands in prices and income. In other words, compute the partial derivatives $\frac{\partial x_i(p,w)}{\partial p_i}$, $\frac{\partial x_i(p,w)}{\partial p_j}$ and $\frac{\partial x_i(p,w)}{\partial w}$. For example, when does the demand for good i increase in the price of another good p_j?

Let's look at some special cases. In Problem set 1, you showed that as $\rho \to 0$, the CES -function converges to the Cobb-Douglas utility function $u(x) = x_1^{\alpha_1} \dots x_n^{\alpha_n}$.

If we just substitute $\rho = 0$ into the optimal demand, we get

$$x_i = \frac{\alpha_i w}{p_i(\sum_{i=1}^n \alpha_i)}.$$

For the Cobb-Douglas utility function, you get the result that the expenditure share $\frac{p_i x_i}{w}$ on each good is equal to $\frac{\alpha_i}{(\sum_{i=1}^n \alpha_i)}$. In this case, the consumer's expenditure share does not depend on her wealth. In other words, rich and poor consumers use the same fraction of their income on food, clothing, yachts etc. This is clearly not a very good description of reality.

By equation (53), you can see that CES -functions do not offer that much help either. The expenditure shares are still constant in wealth (even though they depend now on the entire price vector). One way to get more realistic consumption patters is to define the utility function for consumptions above a level needed for subsistence. Let $\underline{x} = (\underline{x_1}, ..., \underline{x_n})$ be the levels of each good needed for survival and assume that $w \geq p \cdot \underline{x}$. The utility function for $x \in \mathbb{R}^n$ such that $x_i \geq \underline{x_i}$ is of Cobb-Douglas -like form:

$$u(x) = (x_1 - \underline{x_1})^{\alpha_1} \dots (x_n - \underline{x_n})^{\alpha_n},$$

where $0 < \alpha_i < 1$ for all i and $\sum_{i=1}^n \alpha_i = 1$. Notice that the marginal utility for good i is infinite if $x_i = \underline{x_i}$ and that the utility function is strictly increasing in all of its components. Hence we still have an interior solution and the budget constraint binds.

We get as above:

$$\frac{\frac{\partial u}{\partial x_i}}{\frac{\partial u}{\partial x_k}} = \frac{\alpha_i(x_i - \underline{x_i})}{\alpha_k(x_k - \underline{x_k})} = \frac{p_i}{p_k} \text{ for all } i, k,$$

and

$$\sum_{i=1}^n p_i x_i = w.$$

We get that

$$x_i - \underline{x_i} = \frac{\alpha_i p_1}{\alpha_1 p_i}(x_1 - \underline{x_1}) \text{ for all } i. \tag{54}$$

Multiplying both sides by p_i and summing over i gives:

$$\sum_{i=1}^{n} p_i(x_i - \underline{x_i}) = \frac{p_1 \sum_{i=1}^{n} \alpha_i}{\alpha_1}(x_1 - \underline{x_1}).$$

So we can solve:

$$x_1 - \underline{x_1} = \frac{\alpha_1(w - \sum_{i=1}^{n} p_i \underline{x_i})}{p_1},$$

where we used the budget constraint $\sum_{i=1}^{n} p_i x_i = w$ and $\sum_{i=1}^{n} \alpha_i = 1$

By (54), we see that

$$x_k - \underline{x_k} = \frac{\alpha_k(w - \sum_{i=1}^{n} p_i \underline{x_i})}{p_k}.$$

Now you can see that the consumer uses a constant fraction of her excess income (above what is needed for the necessities $\underline{x}$) in constant shares given by the α_i. Since the poor have less excess wealth, their consumption fractions are closer to the ones given by the subsistence levels $\beta_i := \frac{\underline{x_i}}{\sum_i \underline{x_i}}$. Hence the richest spend fractions α_i 0n good i and the poorest spend β_i.

4.2 Expenditure minimization problem

We cover briefly the related problem of minimizing expenditure subject to the constraint of reaching a specified level of utility. All the notation is exactly as in the previous subsection. We assume that the utility function that we have is quasiconcave.

$$\min_{x \in X} p \cdot x = \sum_{i=1}^{n} p_i x_i,$$

subject to

$$u(x) \geq \overline{u}.$$

This means that we have a linear and thus quasiconvex objective function for our minimization problem and since the utility function is quasiconcave, the feasible set is convex. Hence we know that K-T necessary conditions are also sufficient. Notice that the feasible set is now not bounded (why?), but the solution exists because we can take any $x*$ such that $u(x*) \geq \overline{u}$ and restrict attention to x such that $p \cdot x \leq p \cdot x*$ since $x*$ is a feasible solution. But this set is convex and bounded since it is a budget set.

The Lagrangean to the problem is:

$$\mathcal{L}(x,\lambda) = \sum_{i=1}^{n} p_i x_i - \lambda_0(\overline{u} - u(x)) + \sum_{i=1}^{n} \lambda_i x_i.$$

The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial x} = p_i + \lambda_0 \frac{\partial u}{\partial x_i} + \lambda_i = 0 \text{ for all } i, \tag{55}$$

$$\lambda_0[u(x) - \overline{u}] = 0, \tag{56}$$

$$\lambda_i x_i = 0 \text{ for all } i, \tag{57}$$

$$\overline{u} - u(x) \leq 0, \tag{58}$$

$$-x_i \leq 0 \text{ for all } i, \tag{59}$$

$$\lambda_i \geq 0 \;\; i \in \{0, 1, ..., n\}. \tag{60}$$

Notice that for interior solutions (where $\lambda_1 = \lambda_2 = ... = \lambda_n = 0$, we get again (after eliminating the multiplier) from the first line of the K-T conditions that

$$\frac{\frac{\partial u}{\partial x_i}}{\frac{\partial u}{\partial x_k}} = \frac{p_i}{p_k}.$$

We have exactly the same situation as before. Now the ratio of marginal utilities is really the MRT for the problem since it describes the feasible set. The price ratio is now the MRS of this new problem. We will relate these two problems in the next lecture.

4.3 Cost minimization problem for a firm

A firm chooses its inputs k, l to minimize the cost of reaching a production target of $\overline{q}$ at given input prices r, w. The production function is assumed to be a strictly increasing and quasiconcave function $f(k, l)$.

$$\min_{(k,l)\in\mathbb{R}^2_+} rk + wl$$

subject to

$$f(k,l) \geq \overline{q}.$$

Notice that this is the same mathematical problem as in expenditure minimization. Only the names of variables have changed. The solution to the problem is therefore also identical and we do not repeat it here.

5 The value function

5.1 The value function for utility maximization

Consider an unconstrained maximization problem of a function of a single real variable x, where the objective function depends on a parameter $\alpha \in \mathbb{R}$.

$$\max_{x \in \mathbb{R}} f(x,\alpha).$$

Let $x(\alpha)$ be the solution to this problem. Consider the maximum value of the objective function that is achievable at the exogenous variable (or parameter) $\hat{\alpha}$, i.e. $f(x(\hat{\alpha}), \hat{\alpha})$.

We call this new function the *value function* of the problem and denote

$$V(\alpha) := f(x(\alpha), \alpha).$$

At the (unconstrained) optimum $x(\hat{\alpha})$, by the first-order condition:

$$\frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial x} = 0.$$

Suppose that f is twice continuously differentiable and that the second order condition is satisfied so that

$$\frac{\partial^2 f(x(\hat{\alpha}), \hat{\alpha})}{\partial x^2} < 0.$$

Then we can use implicit function theorem to see that $x(\alpha)$ satisfying the first-order condition $\frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial x} = 0$ exists in some neighborhood of $\hat{\alpha}$. We can compute via the chain rule:

$$V'(\hat{\alpha}) = \frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial x} x'(\hat{\alpha}) + \frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial \alpha}.$$

Since $\frac{\partial f(x(\hat{\alpha}),\hat{\alpha})}{\partial x} = 0$, we get

$$V'(\alpha) = \frac{\partial f(x(\alpha), \alpha)}{\partial \alpha}.$$

This observation is called the envelope theorem. In words, it states that when a parameter changes, the maximum value of the problem changes only through the direct effects on the objective function. The indirect effects on the value vanish because of the first-order condition on x.

In the more general case, where $x \in \mathbb{R}^n$, the message is exactly the same. The first order-condition is now:

$$\frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial x_i} = 0 \text{ for all } i \in \{1, ..., n\}.$$

Assuming the conditions for implicit function theorem, we have by chain rule:

$$V'(\hat{\alpha}) = \sum_{i=1}^{n} \frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial x_i} x_i'(\hat{\alpha}) + \frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial \alpha}.$$

Again, the first term vanishes by first-order condition and we are left with

$$V'(\hat{\alpha}) = \frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial \alpha}.$$

The situation is slightly different with constrained optimization problems. Suppose that we have an equality constrained parametric maximization problem for $x \in \mathbb{R}^n$:

$$\max_{x} f(x, \alpha)$$

$$\text{subject to } g(x, \alpha) = 0.$$

Please note that the problem may have many parameters so that α is a vector, but the analysis here is with respect of a single component of the parameter vector.

Let $x(\alpha)$ denote the optimal solution to the problem and assume again that he conditions for the implicit function theorem around the solution are satisfied as before.

The value function is still defined as before:

$$V(\alpha) = f(x(\alpha), \alpha).$$

Begin the analysis by forming the Lagrangean:

$$\mathcal{L}(x, \mu; \alpha) = f(x, \alpha) - \mu g(x, \alpha).$$

The envelope theorem relates the derivative of the value function with respect to the parameter to the partial derivatives of the Lagrangean.

Theorem 6 (Envelope theorem)**.** In an optimization problem subject to an equality constraint, we have:

$$V'(\alpha) = \frac{\partial \mathcal{L}(x, \mu; \alpha)}{\partial \alpha}.$$

Proof.

$$V'(\hat{\alpha}) = \sum_{i=1}^{n} \frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial x_i} x_i'(\hat{\alpha}) + \frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial \alpha}.$$

Now the first-order condition implies that

$$\frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial x_i} = \mu \frac{\partial g(x(\hat{\alpha}), \hat{\alpha})}{\partial x_i}.$$

Since the constraint holds for all α, we have

$$\sum_{i=1}^{n} \frac{\partial g(x(\hat{\alpha}), \hat{\alpha})}{\partial x_i} x_i'(\hat{\alpha}) = -\frac{\partial g(x(\hat{\alpha}), \hat{\alpha})}{\partial \alpha}.$$

Combining these gives:

$$V'(\hat{\alpha}) = \frac{\partial f(x(\hat{\alpha}), \hat{\alpha})}{\partial \alpha} - \mu \frac{\partial g(x(\hat{\alpha}), \hat{\alpha})}{\partial \alpha}.$$

□

The envelope theorem gives us a nice way of understanding the Lagrange multipliers in utility maximization problems. The Lagrangeam for the UMP with a single binding equality constraint is:

$$\mathcal{L}(x, \lambda) = u(x) - \mu \left[\sum_{i=1}^{n} p_i x_i - w \right].$$

The maximum value function

$$v(p, w) = \max u(x) \text{ subject to } p \cdot x = w,$$

is called the indirect utility function. It computes the optimal utility level for all combinations of prices $p \in \mathbb{R}^n_{++}$ and income $w > 0$.

Envelope theorem tells us that:

$$\frac{\partial v(p, w)}{\partial w} = \mu.$$

This means that if your income is increased by one unit, your maximal utility increases the amount given by the multiplier. By reducing income dw you lose μdw of utility and this is why the multiplier is sometimes called the shadow price of income. It evaluates the utility consequences from relaxing or strengthening the constraint.

Envelope theorem also tells us that:

$$\frac{\partial v(p, w)}{\partial p_i} = -\mu x_i.$$

Combining these two, we have Roy's identity:

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

In other words, if you have an indirect utility function, you can compute the demand function by simple partial differentiation. In later courses, you will learn what properties on $v(p, w)$ guarantee that it is the indirect utility function of some UMP for some $u(x)$.

5.2 Expenditure minimization

Consider next the expenditure minimization problem (EMP) from Lecture 9.

$$\min_{h \in X} p \cdot x = \sum_{i=1}^{n} p_i h_i,$$

subject to

$$u(h) = \overline{u}.$$

Denote the solution to this problem by $h(p, \overline{u})$. We call $h_i(p, \overline{u})$ the Hicksian or compensated demand for good i. The (minimum) value function of this problem,

$$e(p, \overline{u}) = \sum_{i=1}^{n} p_i h_i(p, \overline{u}),$$

is called the expenditure function.

The objective function is linear in p and hence by the results in Lecture 6, we know that $e(p, \overline{u})$ is concave in p. Therefore the Hessian matrix of $e(p, \overline{u})$ is negative semidefinite.

We turn next to the The Lagrangean function for the case where we can ignore the inequality constraints:

$$\mathcal{L}(x, \mu) = \sum_{i=1}^{n} p_i h_i - \mu(\overline{u} - u(h)).$$

The envelope theorem tells us that:

$$\frac{\partial e(p, \overline{u})}{\partial p_i} = h_i.$$

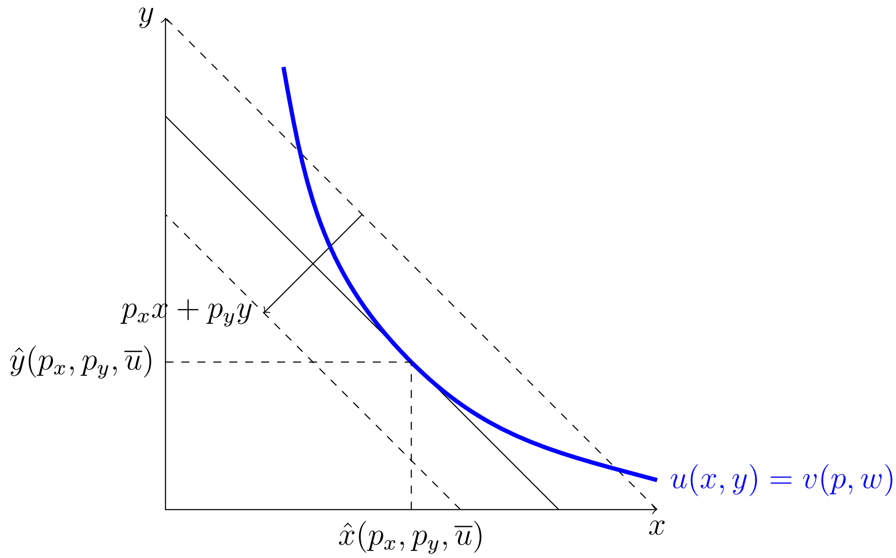
The partial derivatives of $h_i(p, \overline{u})$ with respect to p_j are the elements of the Hessian matrix of $e(p, \overline{u})$.

5.3 Connecting UMP and EMP

The main reason for considering the expenditure minimization problem is that it gives us a nice tool for understanding the solution to the utility maximization problem. Hold prices $\hat{p}$ fixed for a moment and ask how high utility you can achieve with income w. The answer is given by the indirect utility function $v(\hat{p}, w)$.

Ask next what is the minimum expenditure that you must use to achieve utility $v(\hat{p}, w)$. The following figures should convince you that for all $\hat{p}$,

Figure 4: Expenditure minimization for $\overline{u} = v(p, w)$



$$e(\hat{p}, v(\hat{p}, w)) = w.$$

Similarly, if it costs you $e(\hat{p}, \overline{u})$ to reach utility $\overline{u}$. If your budget is $e(\hat{p}, \overline{u})$, then the maximal utility that you can reach is for all $\hat{p}$,

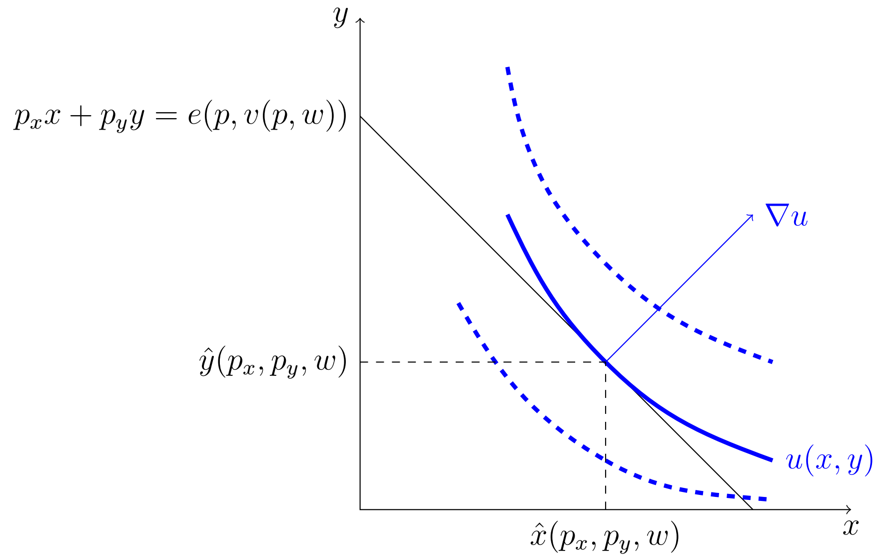
$$\overline{u} = v(\hat{p}, e(\hat{p}, \overline{u})).$$

From these pictures (or alternatively from the K-T constraints of the two problems), you can also see that for $\overline{u} = v(p, e(p, \overline{u}))$ and $e(p, v(p, w)) = w$ the solutions to expenditure minimization and UMP coincide for all p:

$$h_i(p, \overline{u}) = x_i(p, e(p, \overline{u})) \text{ for all } i,$$

$$h_i(p, v(p, w)) = x_i(p, w) \text{ for all } i.$$

Figure 5: UMP for $w = e(p, v(p, w))$



Differentiate the first of these identities with respect to p_j to get:

$$
\begin{aligned}
\frac{\partial h_i(p, \overline{u})}{\partial p_j} &= \frac{\partial x_i(p, w)}{\partial p_j} + \frac{\partial x_i(p, w)}{\partial w}\frac{\partial e(p, \overline{u})}{\partial p_j} \\
&= \frac{\partial x_i(p, w)}{\partial p_j} + \frac{\partial x_i(p, w)}{\partial w} h_j(p, \overline{u}) \\
&= \frac{\partial x_i(p, w)}{\partial p_j} + \frac{\partial x_i(p, w)}{\partial w} x_j(p, e(p, \overline{u})) \\
&= \frac{\partial x_i(p, w)}{\partial p_j} + \frac{\partial x_i(p, w)}{\partial w} x_j(p, w).
\end{aligned}
$$

This is the famous Slutsky equation for income and substitution effects. It relates the changes in the Marshallian (UMP) demands to the Hicksian (EMP) demands. Since the Marshallian demands depend on prices and income, they are in principle observable from demand data. The Hicksian demands depend on the utility level and hence they cannot be directly observed. Nevertheless, we know from the Hessian of the expenditure function that e.g. the Hicksian demand is downward sloping in own demand. With Stltsky equation, we can translate this knowledge to the Marshallian demands where the results are very hard to obtain directly.

The observable change in Marshallian $\frac{\partial x_i(p,w)}{\partial p_j}$ demands can be decomposed into a substitution effect, i.e. the change in compensated demand $\frac{\partial h_i(p,\overline{u})}{\partial p_j}$ and the observable income effect $\frac{\partial x_i(p,w)}{\partial w} x_j(p,w)$.

$$\frac{\partial x_i(p,w)}{\partial p_j} = \frac{\partial h_i(p,\overline{u})}{\partial p_j} - \frac{\partial x_i(p,w)}{\partial w} x_j(p,w).$$

Since we know that the Hessian of $e(p,\overline{u})$ is negative definite, we know that its diagonal elements are non-positive. Hence the effect of increasing p_i on x_i is negative whenever the demand for i is increasing in income (we say then that i is a non-inferior good).

5.4 Cost minimization

A firm chooses its inputs k, l to minimize the cost of reaching a production target of q at given input prices r, w. The production function is assumed to be a strictly increasing and quasiconcave function $f(k,l)$.

$$\min_{(k,l)\in\mathbb{R}^2_+} rk + wl$$

subject to

$$f(k,l) = q.$$

The value function of this problem is called the *cost function* of the firm and denoted by $c(r,w,q)$. We write:

$$c(r,w,q) = rk(r,w,q) + wl(r,w,q),$$

where $k(r,w,q), l(r,w,q)$ solve the cost minimization problem. These are called the conditional factor demands.

As in the case with expenditure minimization, we see that the cost function is concave in r, w since it is the minimum of linear functions of r, w. Therefore the Hessian of the cost function is negative semidefinite. By envelope theorem, we have the result known as Shephard's lemma:

$$\frac{\partial c(r,w,q)}{\partial r} = k(r,w,q), \qquad \frac{\partial c(r,w,q)}{\partial w} = l(r,w,q).$$

Negative semi-definiteness of the Hessian of c implies that (since the diagonal elements must be non-positive)

$$\frac{\partial k(r,w,q)}{\partial r} \leq 0, \frac{\partial l(r,w,q)}{\partial w} \leq 0.$$

In words, conditional factor demands are decreasing in own price (not surprisingly).

5.5 Profit function of a competitive firm

We end this part of the course with the analysis of profit maximization for a price taking firm. There are two ways to think about this. Either minimize cost for each production level q to get $c(r, w, q)$ and then choose q optimally to maximize

$$pq - c(r, w, q),$$

where p is the price of the output.

Alternatively, you can write directly:

$$\max_{k,l,q} pq - rk - wl$$

subject to

$$q = f(k, l).$$

An advantage of the second approach is that the problem is immediately seen to be linear in the input and output prices p, r, w. Let $q(p, r, w), k(p, r, w), l(p, r, w)$ be the optimal output and input choices in the problem. The value function $\pi(p, r, w)$ is called the profit function of the firm.

Since π is the maximum of linear functions in p, r, w, we get by Lecture 6 that π is convex and hence its Hessian is positive semi-definite.

The envelope theorem gives us Hotelling's lemma:

$$\frac{\partial \pi(p,r,w)}{\partial p} = q(p,r,w), \quad \frac{\partial \pi(p,r,w)}{\partial r} = -k(p,r,w), \quad \frac{\partial \pi(p,r,w)}{\partial w} = -l(p,r,w).$$

Since π is positive semi-definite, its diagonal elements are non-negative. This gives the 'Law of Supply' (supply increases in output price):

$$\frac{\partial q(p,r,w)}{\partial p} \geq 0,$$

and the 'Law of Factor Demands' (factor demand decrease in factor price):

$$\frac{\partial k(p,r,w)}{\partial r} \leq 0, \quad \frac{\partial l(p,r,w)}{\partial w} \leq 0.$$

As you can see, the theory of the competitive firm is easier than consumer theory since changes in prices do not change the constraint set (as with the budget set). You will see the firm's problem in some form in almost all branches of economics and in particular in Intermediate Microeconomics. Of course, in many industries firms are not competitive $\rightarrow$ Industrial organization.

On the other hand, firms do not make decisions but people do and people may have different objectives $\rightarrow$ Organizational economics, Contract theory.

6 Difference equations

Difference equations are perhaps the simplest example of dynamical systems. In a dynamical system the state of the system at point n in time, x_t, and a system equation can be used to extrapolate the entire future for the system, i.e. the values of the state x_{t+k} for all $k \in \mathbb{N}$. We will discuss the cases where $x_t \in \mathbb{R}$ as well as the case where $x_t \in \mathbb{R}^n$. Except for some motivating examples, we shall deal with linear systems where the system equation takes the form

$$x_{t+1} = Ax_t$$

for some $n \times n$ matrix A.

6.1 Motivating examples

6.1.1 Fibonacci sequence

Let's start with some examples. The most famous difference equation is the one giving rise to the Fibonacci sequence. We set $x_0 = 0$, $x_1 = 1$ and for all $t > 2$,

$$x_t = x_{t-1} + x_{t-2}.$$

You can compute by repeatedly substituting previous values the sequence $0, 1, 1, 2, 3, 5, 8, 13, ...$ In some sense the system equation is already a solution to the problem. Nevertheless it is not easy to see what x_{200} is or even to know its approximate size. By a solution to a difference equation, we mean here a formula that depends on the initial values x_0 and t so that $x_t = g(x_0, t)$ for some function g. In fact, we can look at the problem as one where the function g is the unknown to be solved. We will see how to compute the values for x_t in the Fibonacci sequence.

6.1.2 Solow growth model

The Solow growth model is perhaps the most important motivating example from economics. In this model, labor is kept fixed at L over time and capital K_t changes over time as a result of savings. The aggregate production function is $y_t = F(K_t, L_t)$ and it is usually assumed to be an increasing concave, linearly homogenous (constant returns to scale) so that

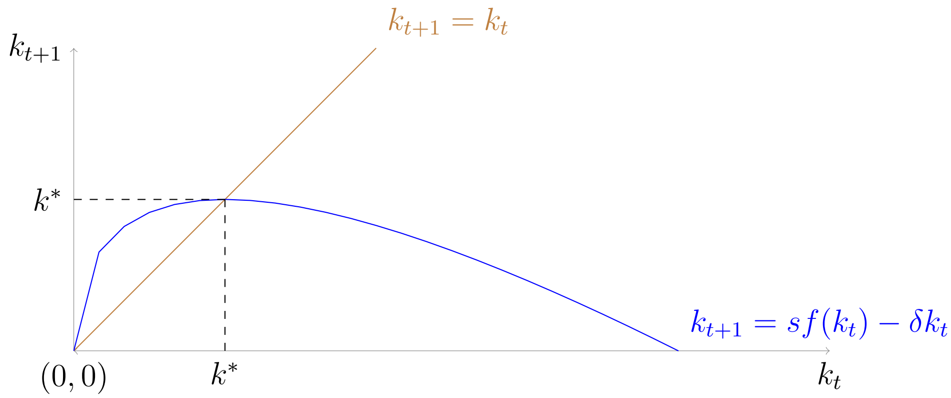
$$F(K_t, L_t) = L_t F(\frac{K_t}{L_t}, 1) := L_t f(k_t) = Lf(k_t),$$

where $k_t := \frac{K_t}{L_t}$. It is often assumed that $\lim_{k\to 0} = \infty$ and $\lim_{k\to\infty} = 0$. Let's assume here that $y_t = k_t^\alpha$ for some $0 < \alpha < 1$.

The output y_t is divided between savings and consumption. The assumption is that a constant fraction sy_t is saved. As you can recall from Principles of Economics II, savings equals investment and investment goes into next period's capital. Capital depreciates at rate δ per period. Taking all this together, we get

$$k_{t+1} = sk_t^\alpha - \delta k_t.$$

Let's draw the graph of $sk_t^\alpha - \delta k_t$ in the (k_t, k_{t+1}) plane with the 45-degree line. Of special interest is the point k^* such that $s(k^*)^{alpha} - \delta k^* = k^*$. If you start the system with $k_0 = k^*$, the system stays there forever since $k_1 = sk_0^\alpha - \delta k_0 = k_0$ and therefore also $k_t + n = k_0$ for all n. We call k^* the steady state or a rest point of the system.



We can picture the movement of the system by positing an initial point k_0 on the horizontal axis. If $k_0 < k^*$, then $k_1 = f(k_0) > k_0$. You can locate the k_1 on the horizontal axis by reflecting on the 45-degree line. Repeating this process, you can show that from any initial point, k_t converges to k^* as $t \to \infty$. We say that k^* is a globally stable steady state.

You may want to note that

$$k^* = (\frac{s}{1+\delta})^{\frac{1}{1-\alpha}}.$$

The steady state capital level is determined by the savings rate s for a fixed technology (fixed α and δ). Hence countries that save more grow more. You will see more sophisticated versions of the model in your intermediate macroeconomics course.

6.1.3 Markov process

A population consists of three income classes $i \in \{1,2,3\}$. If you are in class i, your children are in income class j with probability p_{ji}. Let P be the matrix with a typical element p_{ij}.

Let $x_0 = \mathbf{e_i}$ if you are in class. Then the probability that you child is in class j is given by the column vector

$$x_1 = Px_0 = P\mathbf{e_i}.$$

But then the probability that your grandchild is in class j is given by the column vector

$$x_2 = Px_1 = P^2 x_0,$$

and in general,

$$x_{t+1} = Px_t.$$

This is a linear difference equation system with constant coefficients and we will see how to solve such systems. The interesting questions here include: How does the probability distribution of your sixth successor generation depend on your income class? In other words, how much social mobility is there in the society.

6.2 Linear difference equations with constant coefficients

The simplest form of difference equations are linear difference equations with constant coefficients. These can be written as:

$$x_{t+1} = Ax_t + b_t,$$

where b_t is a given sequence. If $b_t = 0$ for all t, we have a homogenous equation. Wt start with the simplest homogenous equations where $x_n \in \mathbb{R}$ and $A = a \in \mathbb{R}$.

Solving the homogenous equation is very easy. If $x_{t+1} = ax_t$ for all t, then $x_{t+k} = a^k x_t$. Hence any sequence of the form $x_t = ca^t$ solves the difference equation. If we are given the initial value x_0, the solution is $x_t = x_0 a^t$. In other words, the initial value pins down the coefficient c of the general solution.

Consider next an inhomogenous equation,

$$x_{t+1} = ax_t + b,$$

where $b_t = b$ for all t. Clearly the constant solution $x_t = \frac{b}{1-a}$ for all t solves the equation. I claim that also $x_t = ca^t + \frac{b}{1-a}$ solves the equation. But this follows immediately from the fact that $ca^{t+1} = aca^t$.

This principle holds more generally. If you have a particular solution x_t^P to the inhomogenous equation and the general solution of the homogenous equation x_t^H, then the general solution to the problem is $x_t^P + x_t^H$. This is called the principle of superposition and it arises from the linearity of the equations in x_{t+1}, x_t. It is valid also for the case with $x_t \in \mathbb{R}^t$.

Consider next linear systems with constant coefficients. Let $x_t \in \mathbb{R}^n$ for all t and let A be an $n \times n$ matrix of real numbers. A linear homogenous system is then given by:

$$x_{t+1} = Ax_t.$$

As before, we can 'solve' this by repeated substitution to get

$$x_{t+k} = A^k x_t.$$

Hence I could write the general solution as $x_t = A^t c$ for some vector $c = (c_1, ..., c_k)$. I do not consider this a real solution since it is almost impossible to see what A^t is except in some very special cases. If A is a diagonal matrix with diagonal elements $a_1, ..., a_n$, then the solution becomes

$$x_{i,} = ca_i^t \text{ for } i \in \{1, ..., n\}.$$

Here we have essentially independent variables and the difference equation for each can be solved separately.

6.3 Eigenvalues and eigenvectors

To deal with the general case, we want to change the basis in $\mathbb{R}^k$ so that A is diagonal in that basis. This involves the eigenvectors and eigenvalues of A. You can visualize the effect of matrix multiplication on vectors as consisting of two operations: a rotation and a stretching or shrinking. Eigenvectors of A are those vectors that are not rotated, i.e. if $x \neq 0$ is an eigenvector of A, then for some $\lambda \in \mathbb{R}$,

$$Ax = \lambda x.$$

We may write this more compactly as

$$(A - \lambda I)x = 0,$$

where I is the $n \times n$ identity matrix. But from basic linear algebra, we know that a homogenous linear equation can have a non-zero solution only if the matrix does not have full rank, i.e. if $\det(A - \lambda I) = 0$. The values of λ for which this determinant is zero are called the eigenvalues of A.

The determinant of $(A - \lambda I)$ is called the characteristic polynomial of A so the eigenvalues are the roots of the characteristic polynomial. If A has n distinct eigenvalues $\lambda_1, ..., \lambda_n$, then it has also n linearly independent eigenvectors $v_1, ..., v_n$ so that

$$Av_i = \lambda_i x_i.$$

Let' see an example on how to compute the eigenvalues and vectors. Let

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then

$$A - \lambda I = \begin{pmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{pmatrix},$$

and

$$\det(A - \lambda I) = \lambda^2 - lambda - 1.$$

We have $\det(A - \lambda I) = 0$ if

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

The corresponding eigenvectors are:

$$v_1 = (\frac{1+\sqrt{5}}{2}, 1), v_2 = (\frac{1-\sqrt{5}}{2}, 1).$$

A useful thing to keep in mind about eigenvalues is that the sum of the eigenvalues equals the trace of the matrix and the product of the eigenvalues equals the determinant of the matrix. This is particularly useful for inference about the signs of eigenvalues.

Since the characteristic polynomial may fail to have real roots, eigenvalues correspond to the case where the matrix does not have any directions that are not rotated. To see an easy example of such a matrix, consider the 90-degree rotation anticlockwise:

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The characteristic polynomial for this matrix is $\lambda^2 + 1$ which obviously does not have real roots. If the eigenvalues are complex numbers, the eigenvectors are also have complex coordinates. We do not have time in this course to pursue this, but it should be pointed out that the method outlined below for solving the difference equations extends also to the case with complex eigenvalues.

It should not come as a surprise that rotations rotate all vectors. You may want to pursue the geometric implications for a 3x3 matrix with a single real eigenvalue and hence a single real eigenvector.

6.4 Eigenvectors, eigenvalues and difference equations

I can express any $x \in \mathbb{R}^n$ given in the usual coordinate system in the coordinate system of spanned by the eigenvectors by simple matrix multiplication. Let $P = [v_1 \quad v_2 \quad \ldots v_n$ be the matrix formed by the eigenvectors. Then for any vector y expressed in the coordinate system of the eigenvectors, we can translate it to the standard system by $x = Py$. Similarly any x in the standard system is $y = P^{-1}x$ in the system of the eigenvectors.

$$y_{t+1} = P^{-1}x_{t+1} = P^{-1}Ax_t = P^{-1}APy_t.$$

Now we want to show that $P^{-1}AP = \Lambda$, where Λ is the diagonal matrix of eigenvalues. But this is the same claim as (premultiply by P):

$$AP = P\Lambda.$$

But this follows immediately from the fact that P consists of the eigenvectors of A.

Hence we have: $y_t = (y_{1,t}, ..., y_{k,t}) = (c_1\lambda_1^t, ..., c_n\lambda_n^t)$. Since $x_t = Py_t$, we have the general solution:

$$x_t = c_1\lambda_1^t v_1 + ... + c_n\lambda_n^t v_n.$$

Note that $A^k = P\Lambda P^{-1}$. Therefore we could have also concluded that

$$x_t = P\lambda^k P^{-1}x_0.$$

The two methods give the same results since $Pc = x_0$ or $c = P^{-1}x_0$.

Sometimes a matrix has a repeated eigenvalue. Consider for example

$$A = (\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}).$$

Then the characteristic equation is $(1-\lambda)^2 = 0$ and the matrix has a single eigenvalue $\lambda = 1$ and therefore a single eigenvector $(1, 0)$. This matrix cannot be diagonalized in the procedure that we had above. Luckily enough all matrices can be expressed as

$$A = Q^{-1}BQ,$$

where Q is a matrix of generalized eigenvalues and B is upper triangular. Since the powers of upper triangular matrices are easy to compute, the same procedure as before can be applied for solving the model. See the book for the details on this.

6.5 Properties of the solutions

For all (homogenous) systems of linear difference equations, 0 is a steady state. If A has full rank, it is the only steady state. Does the system eventually converge to its steady state?

Look at the general solution

$$x_t = c_1\lambda_1^t v_1 + ... + c_n\lambda_n^t v_n.$$

If $|\lambda_i| < 1$ for all i, then $x_t \to 0$ for all c. We say that in this case, the origin is a globally stable steady state or a sink. If $|\lambda_i| > 1$ for all i, then the length of x_t grows without bound for all c. We say that the origin is unstable or a source. Finally if If $|\lambda_i| < 1$ for some i and If $|\lambda_i| > 1$ for some i, then the length of x_t grows without bound if $c_i \neq 0$ for some i with $|\lambda_i| > 1$. If $c \neq 0$ only for i with $|\lambda_i| < 1$, then x_t converges to the origin. In this last case, we say that origin is a saddle point for the system. If $\lambda_i = 1$ for some i, then origin is neither stable, unstable nor a saddle.

6.6 Linearizing non-linear systems

For your future information, I note here that if x^* is a steady state of a nonlinear system, we can use Taylor's first order approximation to analyze the local behavior of the system around the steady state (you'll do this in macroeconomics a lot). Here is the idea.

Suppose that $x_{t+1} = f(x_t)$ and $x^* = f(x^*)$. Then we have

$$x_{t+1} = f(x_{t+1}) = f(x^*) + D_x f(x^*)(x_t - x^*) \text{ or}$$

$$x_{t+1} - x^* = D_x f(x^*)(x_t - x^*).$$

But this is a linear system in the deviations from the steady state and we can apply the analysis from the linear case in the for small deviations. You can classify the steady states of nonlinear models locally as we just did for the linear system (but globally). Just look at the absolute values of the eigenvalues and compare to 1.

6.7 Markov model

Consider the system

$$x_{t+1} = Px_t$$

for a stochastic matrix P, i.e. non-negative matrix whose elements in each column sum up to 1. You have already shown in Problem set 0 that $\lambda = 1$ is an eigenvalue for all Markov matrices.

It can be shown that in the case with strictly positive entries, all other eigenvalues are less that one in absolute value. Therefore x_t converges in the long run to the eigenvector (whose coordinates are normalized to sum to 1) corresponding to eigenvalue 1.

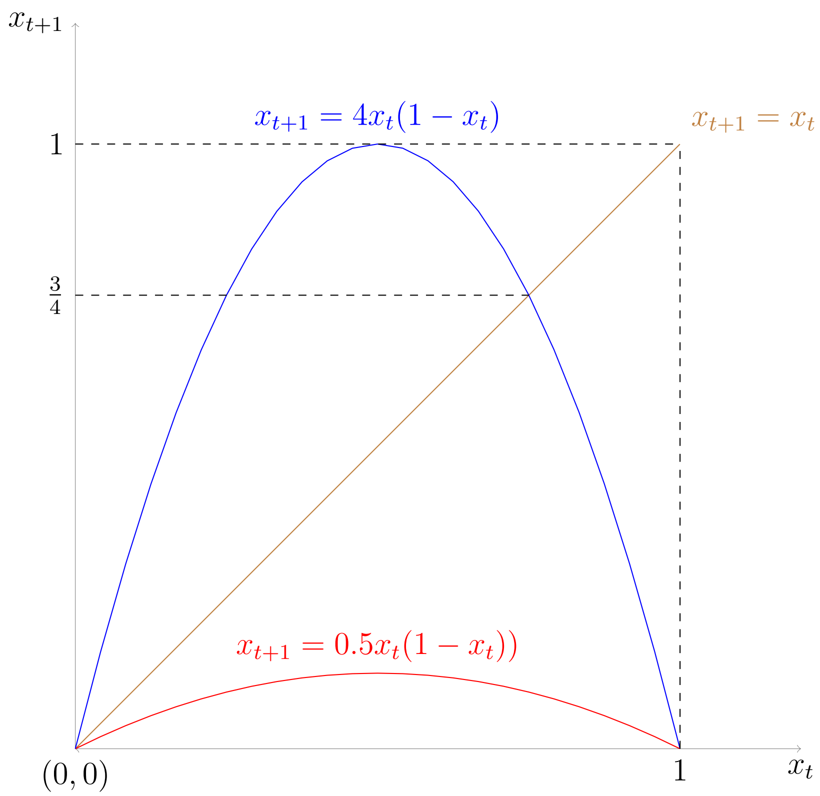
The second largest (in length) eigenvalue measures the speed of convergence to this eigenvector.

6.8 Extra: Non-linear systems

General difference equations for $x_t \in \mathbb{R}$ look deceptively simple, but can give rise to really surprising behavior. One famous example is the logistic equation on the unit interval: choose $0 < x_0 < 1$ and compute for $t > 1$:

$$x_{t+1} = rx_t(1 - x_t).$$

This is a nice differentiable function whose values remain in $(0, 1)$ for all t as long as $r < 4$. To analyze a difference equation on the real line, the first step is to look at the graph of the system equation.

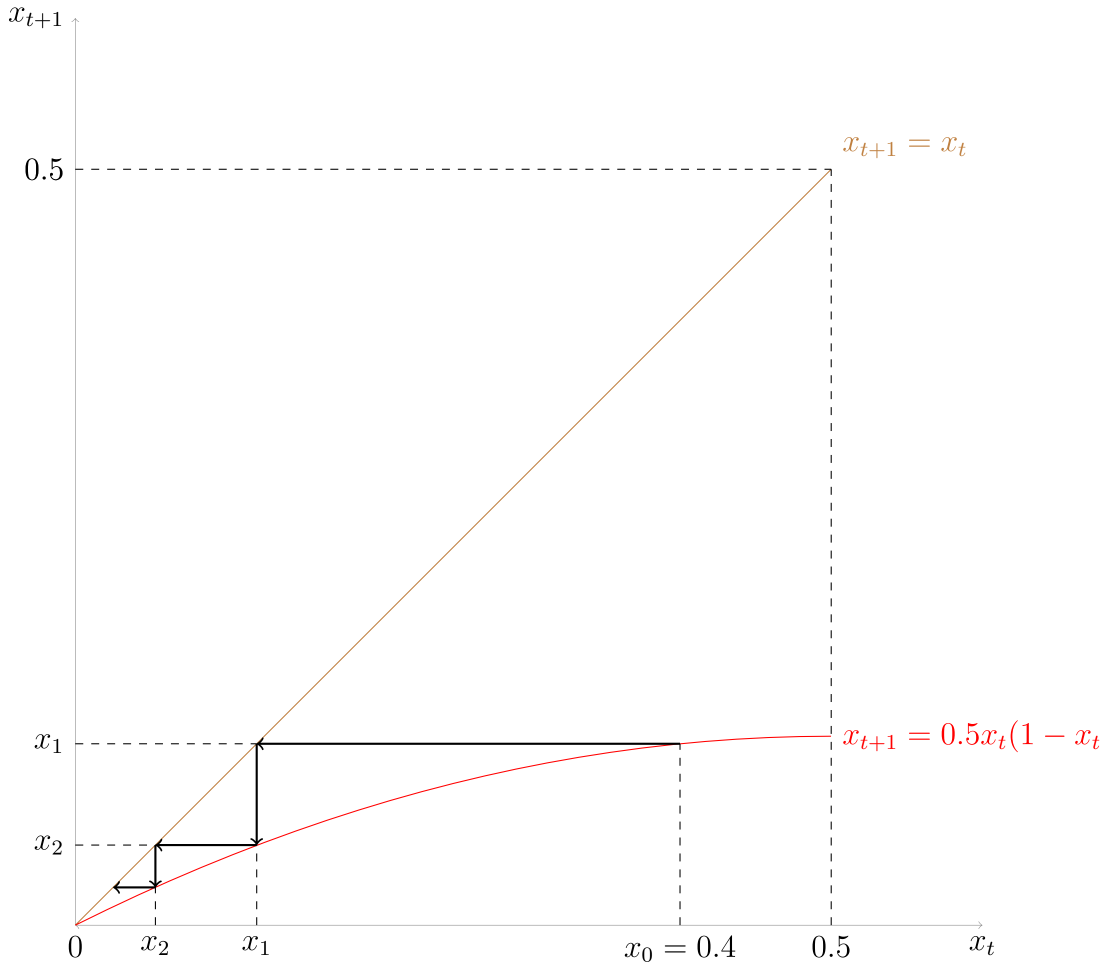


What is the significance of the intersections of $x_{t+1} = rx_t(1 - x_t)$ and the 45 -degree line $x_{t+1} = x_t$. The system stops at any such point, because if $x_{t+1} = x_t$, then also $x_{t+k} = x_t$ by repeated substitution into the system equation. These are called the steady states of the dynamical system.

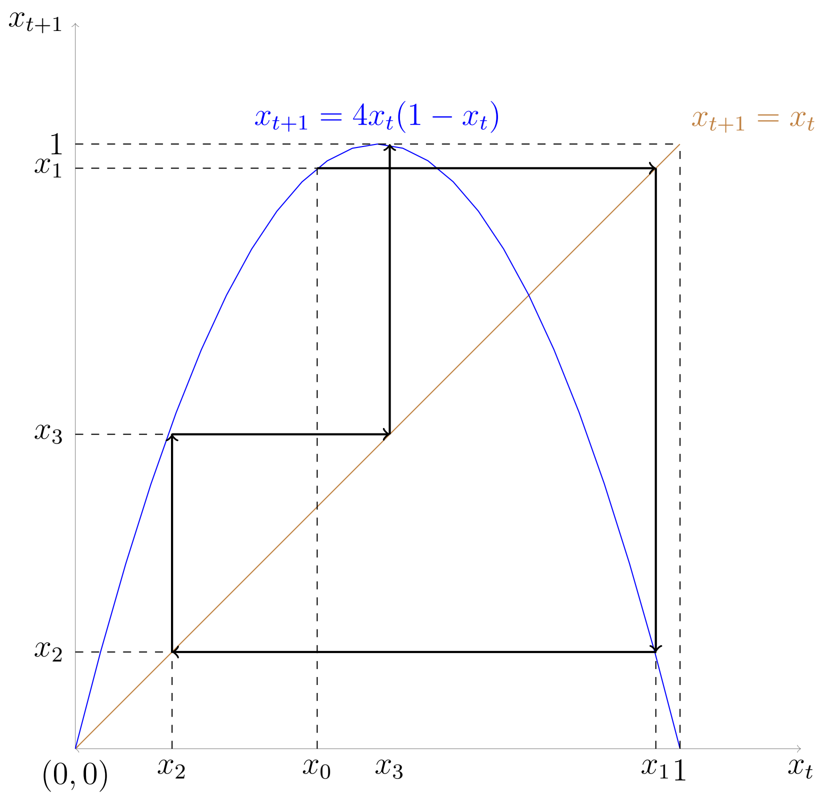
Notice that the system has a single steady state at $x = 0$ if $r < 1$ (can you show this?). For $4 > r > 1$, the system has another steady state at $x = \frac{r-1}{r}$. What happens to the values of x_t as t grows?

Here is a nice graphical way of seeing what happens to the sequence. Lets graph the function in a coordinate system where x_{t+1} is on the vertical and x_t is on the horizontal axis. Draw the graph of $x_{t+1} = f(x_t)$ and pick a starting point $x_0 = 0.4$ for example on the horizontal axis. You can read $x_1 = f(x_0)$ on the graph. You need to picture x_1 on the horizontal axis to see where x_2 is located. But you can do this by reflection through the 45 -degree line. The you just continue the procedure.

Lets look first at the case $x_t + 1 = \frac{1}{2}x : x(1 - x_t)$, i.e. lets take the red curve in the previous picture.



As you can see, for any starting x_0, the system x_t converges quite quickly to 0. If we take the blue graph from the first picture, things look quite different. Let's follow the system again for a few rounds starting at $x_0 = 0.4$.



It is much harder to see where the system might converge and actually the long run behavior of this simple dynamical system is very complicated. In fact, one can show that the system has cycles of all lengths, i.e. for all k, you can find $x_0, x_1, ..., x_{k-1}, x_k = x_0$ such that $x_j + 1 = 4x_j(1 - x_j)$ for all $j \in$.

This model is so simple that you can simulate it very easily with Excel. You can see for example how quickly the paths from nearby starting values diverge.