Topic I: One-variable optimization

EC123 Mathematical Techniques B

Professor Dennis Novy

University of Warwick

Topic I: One-variable optimization

Topic I's big picture: We will learn general techniques to find the maximum/minimum values of a univariate function on a given set $\Leftrightarrow$ We will find the "optimum" values of one-variable functions.

- I.1 Convexity and concavity
- I.2 Extreme points
- I.3 Existence of extreme points
- I.4 Local extreme points
- I.5 Inflection points
- I.6 Sufficient conditions for local extreme points

I.1 Convex and concave functions (Ch. 6.9)

Review: First derivative and increasing/decreasing functions

First, we introduce some notation:

- Let $f : D \rightarrow R$ be a function of domain D and range R be denoted by $f(x)$.
- We will denote its first derivative at $x \in D$ as $f'(x)$, which (just to refresh your memory) is defined as:

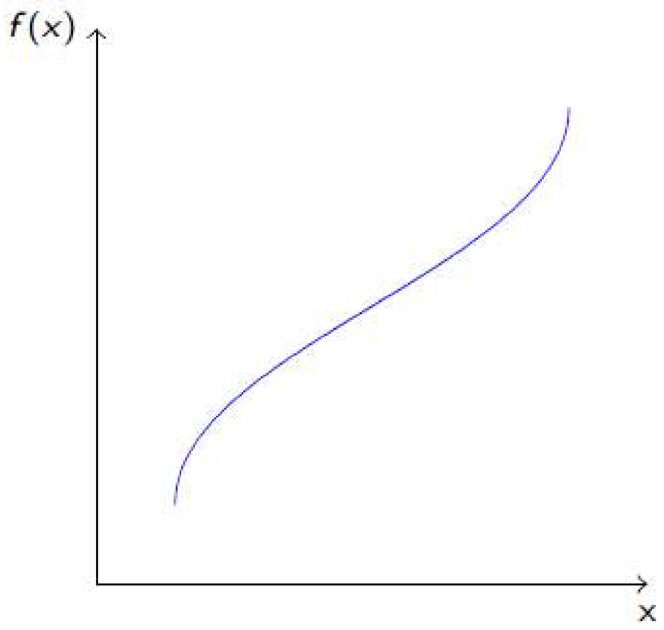
Definition

The first derivative of $f(\cdot)$ at $c \in D$ corresponds to

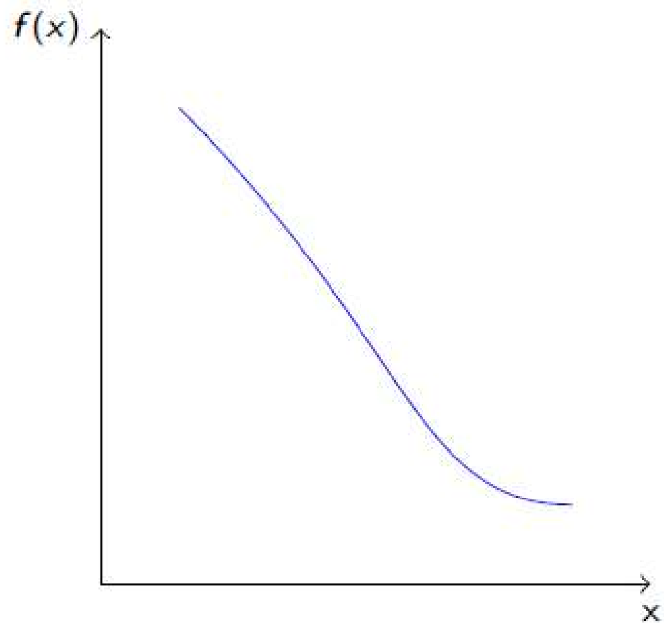
$$f'(c) = \lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h}.$$

- If $f'(x) \geq 0$ for all $x \in D$, we say $f(x)$ is increasing over D. If $f'(x) \leq 0$ for all $x \in D$ we say $f(x)$ is decreasing over D.
- If the inequalities above are strict then the function is strictly increasing/decreasing.

One can show that the first derivative corresponds to the slope of the tangent at c (or any other point in the domain).



(a) $f'(x) \geqslant 0$



(b) $f'(x) \leqslant 0$

- Lost? Review Ch. 6.1, 6.2 and 6.3 to understand first derivatives and increasing/decreasing functions.

Back to business: Convexity and concavity

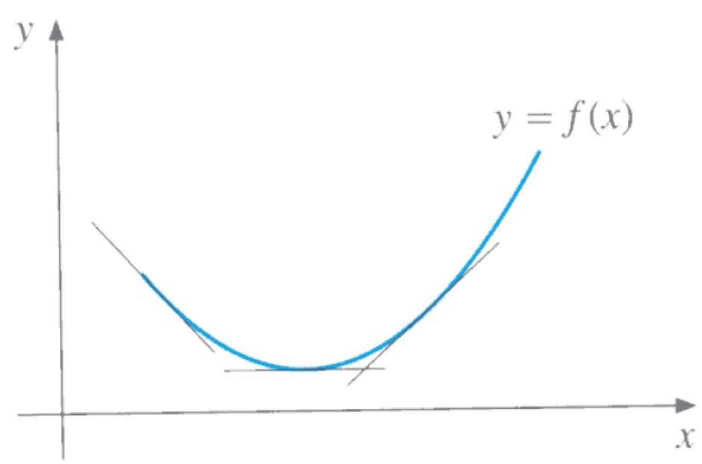
- Convexity and concavity deal with the function's second derivative, that is, the derivative of the first derivative $f'(x)$.
- We denote the second derivative by $f''(x)$.

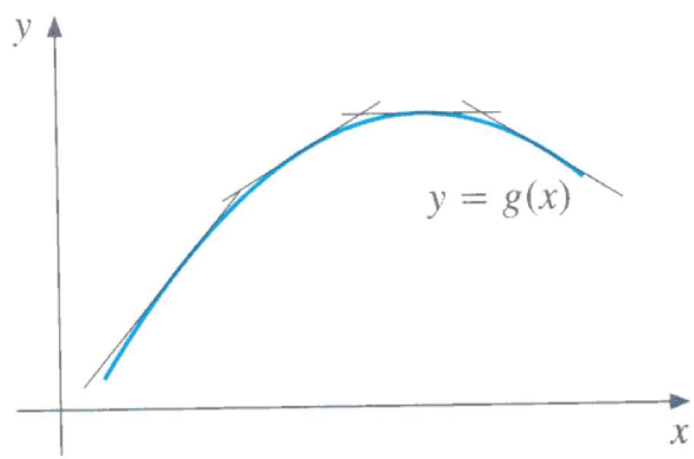
Definition

Given that it is twice continuously differentiable, a function $f(x)$ is convex over D if and only if $f''(x) \geq 0$ for all $x \in D$; concave over D if and only if $f''(x) \leq 0$ for all $x \in D$.

- From the previous definition it follows that if a function is convex, its first derivative is increasing. If the function is concave, then its first derivative is decreasing.

To understand whether a function is convex or concave examine whether the tangent increases or decreases with x.

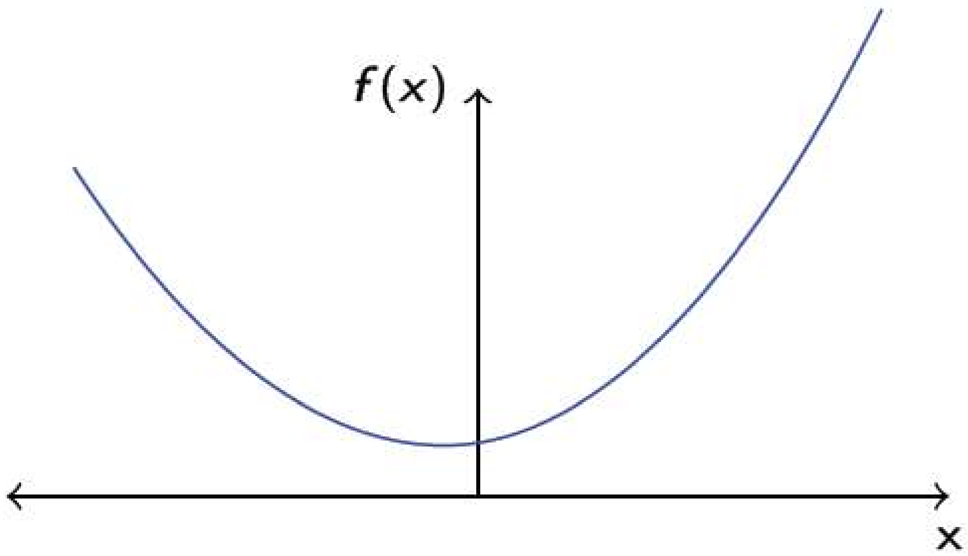




- Can you think of a function that is both convex and concave at the same time? Answer: A linear function $f(x) = a + bx$ since $f'(x) = b$ and $f''(x) = 0$ and thus it satisfies both definitions.

Example 1

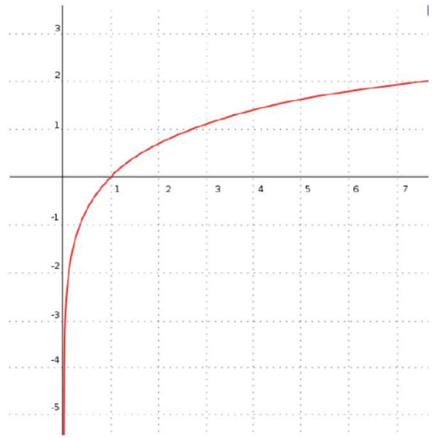
Consider the function $f(x) = 1 + x + x^2$ defined over $\mathbb{R}$:



1) Its first derivative is $f'(x) = 1 + 2x$.
   - Thus it is increasing in the interval $[-\frac{1}{2}, \infty)$ since $x \geq -\frac{1}{2}$ implies $f'(x) \geq 0$.
   - Conversely, it is decreasing in the interval $(-\infty, -\frac{1}{2}]$.
2) Its second derivative is $f''(x) = 2$.

Example 2

Consider the function $f(x) = \ln(x)$ whose domain is $X = \mathbb{R}_{++}$ (that is, all x which are positive.)

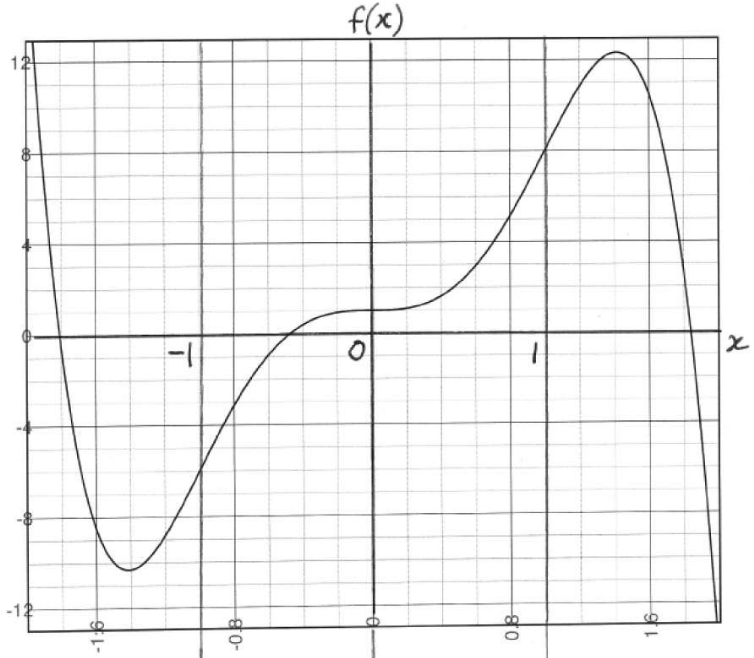


- Its first and second derivatives are $f'(x) = \frac{1}{x} > 0$ and $f''(x) = -\frac{1}{x^2} < 0$.
- Thus the natural logarithm function is increasing and concave.

Example 3

A function can be convex over one domain and concave over another.

- Consider the function $f(x) = 1 + 10x^3 - 3x^5$.
- Its first and second derivatives are $f'(x) = 30x^2 - 15x^4$ and $f''(x) = 60x - 60x^3$.
- The function is **both** convex **and** concave for different domains as the sign of f'' changes with the value of x.
  - To find the boundaries of the domains where the function changes from convex to concave or vice versa, set $f''(x) = 0$:
  - $60x - 60x^3 = 0 \Leftrightarrow 60x(1 - x^2) = 60x(1 - x)(1 + x) = 0$. Therefore $x = 0, +1, -1$.
  - We can see that if $x < -1$, then $f''(x) > 0$ (e.g. $x = -2, f''(-2) = -120 + 480 = 360$). This indicates convexity.
  - Likewise, if $0 < x < 1, f''(x) > 0$, so the function is convex over this domain.
  - And, if $-1 < x < 0$, or $x > 1$, $f''(x) < 0$, so the function is concave over this domain.



Domain:	$x < -1$	$-1 < x < 0$	$0 < x < 1$	$x > 1$
2nd Derivative:	$f'' > 0$	$f'' < 0$	$f'' > 0$	$f'' < 0$
Concave/ Convex?	Convex	concave	Convex	Concave

A more general definition of convexity and concavity (Ch. 8.7)

Definition

A function $f(x)$ is convex if for any two x_1, x_2 and any $t \in [0,1]$, $tf(x_1) + (1-t)f(x_2) \geq f(tx_1 + (1-t)x_2)$. If the inequality holds as $\leq$, then the function is concave.

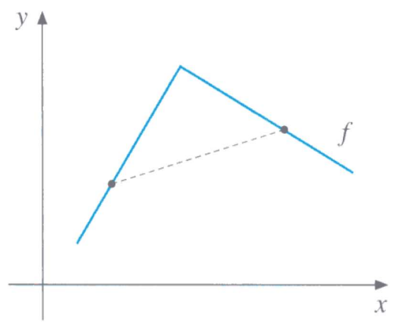
- You will be using this definition in your microeconomics classes a lot!
- Notice: If $f(x)$ is convex, you can write

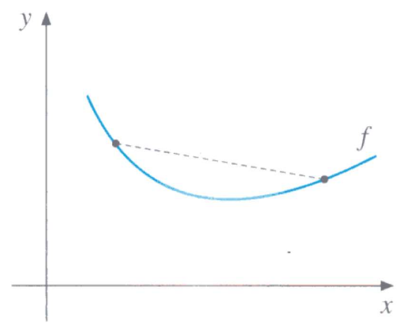
$$\begin{aligned} tf(x_1) + (1-t)f(x_2) &\geq f(tx_1 + (1-t)x_2) \\ \iff -(tf(x_1) + (1-t)f(x_2)) &\leq -f(tx_1 + (1-t)x_2) \\ \iff t(-f(x_1)) + (1-t)(-f(x_2)) &\leq -f(tx_1 + (1-t)x_2). \end{aligned}$$

- Thus, if $f(x)$ is convex, $-f(x)$ is concave, and vice versa.

Definition

A function f is concave (convex) if the line segment that joins any two points on the graph is below (above) the graph, or on the graph.





- The figure on the right is convex.
- The figure on the left is concave even though it is not differentiable! (Do you see now why this is a more general definition?)
- For twice continuously differentiable functions this definition nests our previous definition that was based on the second derivative.

I.2 Extreme points (Ch. 8.1, 8.2 and 8.3)

- Most economic problems involve the notion of finding the “best way” of doing something.
- Examples: Deciding how much to study, how much to exercise, how much of a risky asset to buy or how much fertilizer to use.
- Economic modeling involves setting up the economic problem as a mathematical problem. Solving such a problem usually involves maximizing/minimizing a function.
- Examples: Choose how many hours to study by maximizing utility.
- Now we start studying how to maximize/minimize one-variable functions.
- In mathematical jargon, we study how to find the extreme points of a univariate function.

Definition

The point where a function reaches its maximum value is called a maximum point. Formally, $c \in D$ is a maximum point if $f(x) \leq f(c)$ for all $x \in D$.

Definition

The point where a function reaches its minimum value is called a minimum point. Formally, $c \in D$ is a minimum point if $f(x) \geq f(c)$ for all $x \in D$.

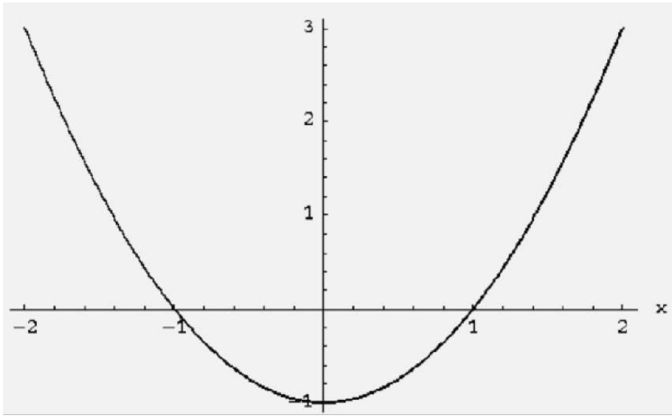
Definition

If $c \in D$ is either a maximum or minimum point, then c is an extreme point or optimal point.

- If the inequalities above are strict, then we refer to the extreme point as a strict maximum or strict minimum.
- Note: 'maximum' is singular, 'maxima' is plural etc.

How do we find extreme points?

- Sometimes it is very easy: we just need to plot the function. Consider the function $f(x) = x^2 - 1$ defined over $\mathbb{R}$:



- By inspection it is obvious that the function has a minimum at $x = 0$. The minimum value corresponds to $f(0) = 0^2 - 1 = -1 \in \mathbb{R}$.
- As the minimum is $0 \in \mathbb{R}$, the extreme point is interior (as opposed to $-\infty$ or $+\infty$ or at the edges of D).

- But sometimes life is not as easy and we cannot rely on graphical inspection to determine the extreme points.
- Go back to the example of $f(x) = x^2 - 1$. Is there anything particularly noticeable at $f(0)$? At that point the slope of the tangent is zero!

Definition

A point $c \in I$ is called a stationary point of the function $f(x)$ if $f'(c) = 0$.

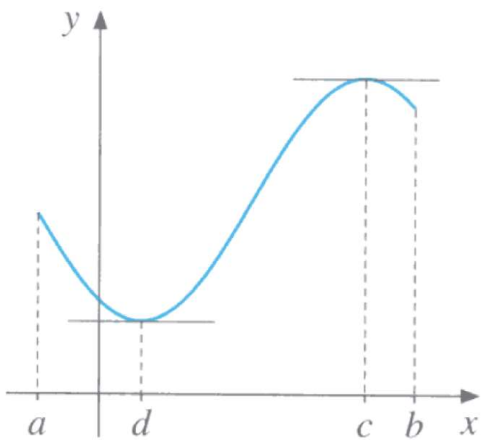
- The following theorem is **one of the most important things** you will learn in mathematical techniques.

Theorem

Suppose the function $f(x)$ is differentiable over an interval I and that c is an interior point of I. Then c is an extreme point in I only if it is a stationary point, that is

$$f'(c) = 0. \qquad \textit{(first-order condition, or FOC)}$$

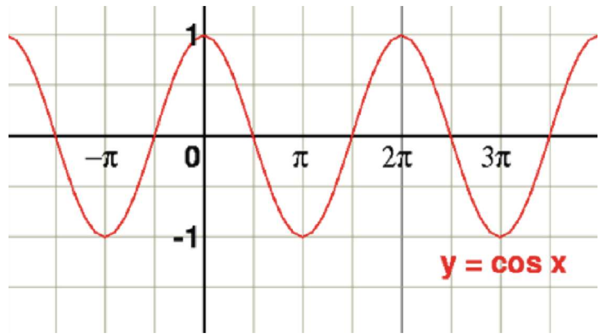
For example, consider the following function:



- There are two stationary points corresponding to a maximum (point c) and a minimum (point d).

Are extreme points unique?

- In our Example 1 there was only one minimum. This is not always the case. Consider the function $f(x) = cos(x)$ defined over $[-\pi, 3\pi]$:

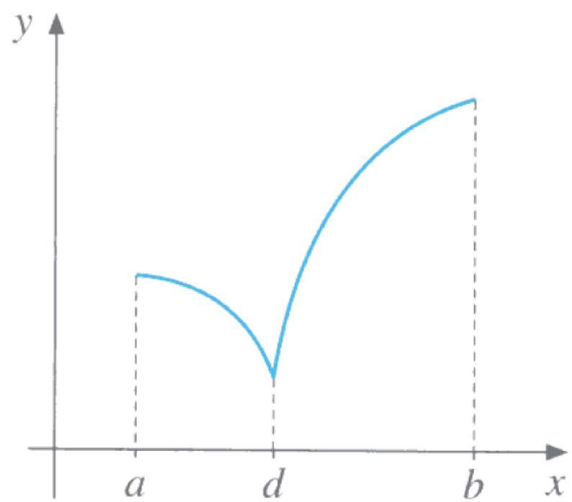


- There are two maxima: at 0 and at 2π.
- The FOC will identify the extreme points.

Is the first-order condition all we need to identify the extreme points?

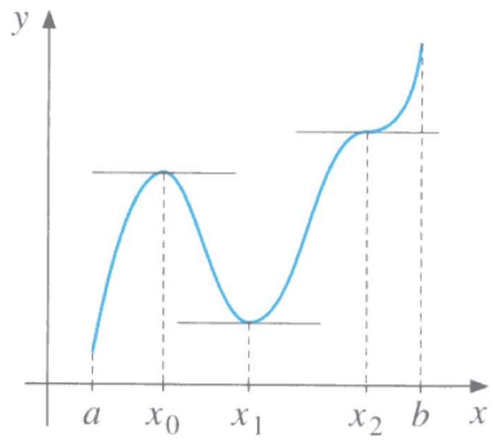
- Sadly, no. We now explore some ways in which the FOC may fail.

1) A function is not differentiable over the whole domain.



- $f(x)$ is defined over $[a, b]$ and it is not differentiable at d.
- The function has a minimum at $d \in [a, b]$ and a maximum at b but it has no stationary points as the function is not differentiable at those points.
- (Notice that the theorem above requires $f(x)$ to be differentiable over its whole relevant domain!)

2) A function has a *local* maximum/minimum but neither is an extreme point.



- The function has three stationary points: x_0, x_1 and x_2 but neither of them is an extreme point over the domain $[a, b]$.
- The minimum of the function corresponds to a and the maximum corresponds to b.
- x_0 is a *local* maximum, x_1 is a *local* minimum, and both are stationary points.
- x_2 is a special case of an *inflection point* (special because it is also a stationary point). Points at which a function changes from being convex to being concave, or vice versa, are called inflection points.

Thus $f'(x) = 0$ is not a *sufficient* condition to identify extreme points.

I.3 Existence of extreme points

- Sometimes we are only interested in knowing whether an extreme point exists rather than finding the point itself.
- This happens mainly when the objective function is not an explicit function or is simply too funky for its derivatives to be computed in a closed form.

Theorem

If f is a continuous function over a closed bounded interval $[x_0, x_1]$, then there exists a point $c \in [x_0, x_1]$ where f has a minimum and a point $c' \in [x_0, x_1]$ where f has a maximum so that

$$f(c) \leq f(x) \leq f(c') \qquad \text{for all } x \in [x_0, x_1].$$

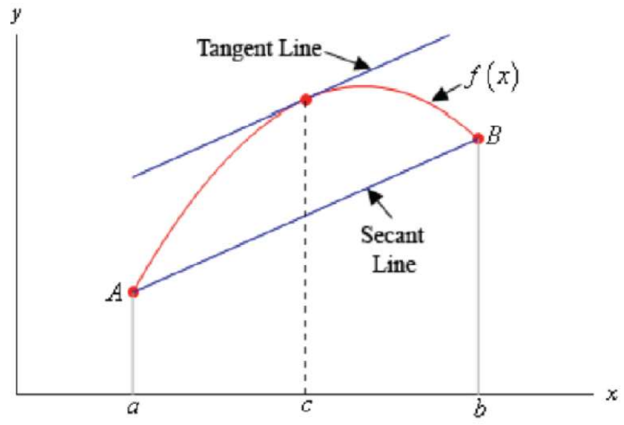
- *This is the so-called extreme value theorem (EVT).*

Theorem

If f is continuous over the closed bounded interval $[x_0, x_1]$ and differentiable in the open interval (x_0, x_1), then there exists at least one interior point $c \in (x_0, x_1)$ such that

$$f'(c) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

- *This is the so-called mean value theorem (MVT).*
- *What happens if the function is such that $f(x_0) = f(x_1)$? Then the theorem ensures that there exists a c such that $f'(c) = 0$. A stationary point! (Rolle's theorem)*



- The slope of the secant between A and B is $\frac{f(b)-f(a)}{b-a}$. (Show this, it is easy.)
- Then c is such that $f'(c) = \frac{f(b)-f(a)}{b-a}$. The theorem ensures that such a c exists.
- Notice that the MVT is more demanding over the function than the EVT: it requires the function to be differentiable everywhere except at the end points of the domain.

Example

- Test the mean value theorem on $f(x) = x^3 - x$ in $[0, 2]$.
- Solution:
  - We find that $\frac{f(2)-f(0)}{2-0} = 3$ and $f'(x) = 3x^2 - 1$.
  - The equation $f'(x) = 3$ has two solutions, $x = \pm 2\sqrt{3}/3$.
  - Since the positive root $x^* = 2\sqrt{3}/3 \in (0, 2)$, we have $f'(x^*) = \frac{f(2)-f(0)}{2-0} = 3$.
  - Thus, the mean value theorem is confirmed in this case.

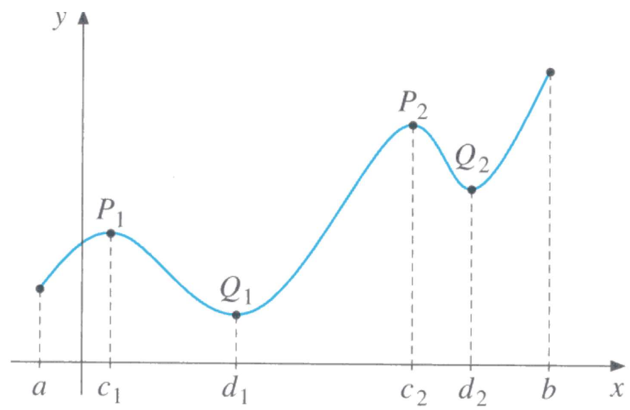
I.4 Local extreme points

- Sometimes it is useful to emphasize that a stationary point is a maximum or a minimum but only over a **given** domain.
- To distinguish these extreme points from those defined over the whole domain of the function we state the following definition:

Definition

A function $f(x)$ has a local maximum/minimum at c if there exists an interval (a, b) such that $c \in (a, b)$ and c is a maximum/minimum for the function defined over the interval (a, b).

Consider the following function defined over $[a, b]$.



- Here c_1, c_2 and b are local maximum points around their respective local intervals.
- Which is the maximum over the entire domain? b

$\Rightarrow$ Every extreme point is a local extreme point but not vice versa.

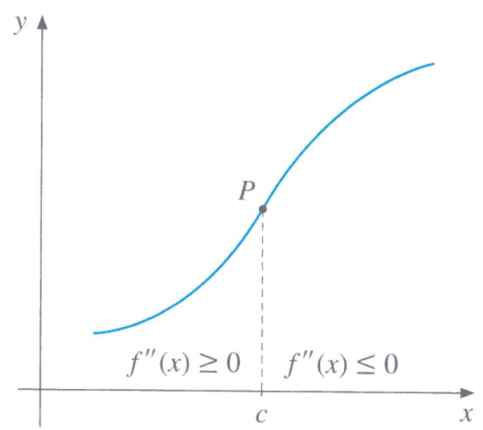
- Moreover, a, d_1 and d_2 are local minima, and d_1 is the minimum.

I.5 Inflection points

- Before moving to sufficient conditions for extreme points, we need to study one more type of point that you may encounter.

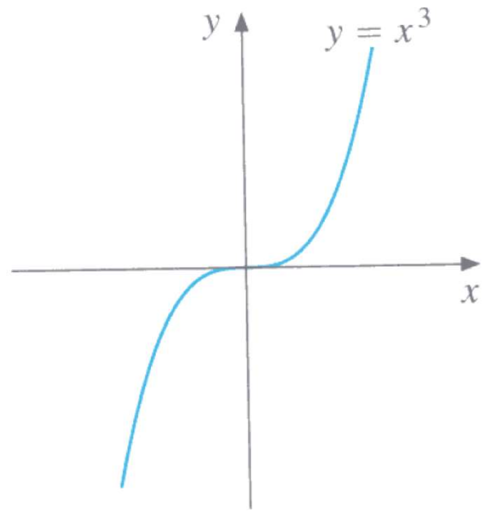
Definition

A point c is called an inflection point for the function f if there exists an interval (a, b) around c such that the function is convex (concave) on (a, c) and concave (convex) on (c, b).



- Here the function is convex to the left of c and concave to the right of c.

- Inflection points are another way in which the FOC may fail to identify a maximum or minimum.



- To identify an inflection point look at the second derivative and check whether the signs flip below and above.
- If c is an inflection point, then $f''(c) = 0$ and $f''(c)$ changes sign for smaller/bigger points on the domain.

Example

Consider the following polynomial:

$$f(x) = \frac{1}{9}x^3 - \frac{1}{6}x^2 - \frac{2}{3}x + 1.$$

The derivatives are

$$f'(x) = \frac{3}{9}x^2 - \frac{2}{6}x - \frac{2}{3}, \qquad f''(x) = \frac{6}{9}x - \frac{2}{6}.$$

Thus $\frac{6}{9}x - \frac{2}{6} = 0 \Rightarrow x = \frac{1}{2}$. Evaluating the second derivative at $x < \frac{1}{2}$ and $x > \frac{1}{2}$, we have

$$f''(x) = \frac{6}{9}x - \frac{2}{6} < 0 \quad \text{for} \quad x < \frac{1}{2},$$

$$f''(x) = \frac{6}{9}x - \frac{2}{6} > 0 \quad \text{for} \quad x > \frac{1}{2}.$$

Example

Show that $f''(x) = 0$ is not sufficient for an inflection point.

Consider the function

$$f(x) = x^4.$$

The derivatives are

$$f'(x) = 4x^3, \qquad f''(x) = 12x^2.$$

Thus $x = 0$ yields $f''(x) = 0$. However, $f''(x) > 0$ for all other x. Thus we need to check two things: that the second derivative is zero (necessary condition) **and** that the function changes from convex (concave) to concave (convex).

I.6 Sufficient conditions for extreme points

- Are there circumstances under which we can ensure that the FOCs are necessary and sufficient? *Yes!*
- The following theorem specifies such conditions.

Theorem

Suppose $f(x)$ is a convex (concave) continuous function over an interval I. If c is a stationary point for f in the interior of I, then c is a minimum (maximum) point for f in I.

- Thus the golden rule is to check whether the function is convex or concave and then simply use the FOC.
- This theorem is the **second most important thing** you'll learn in mathematical techniques!

Example 1

Take

$$f(x) = \ln(x) - x \qquad \text{for} \quad x > 0.$$

The first and second derivatives are

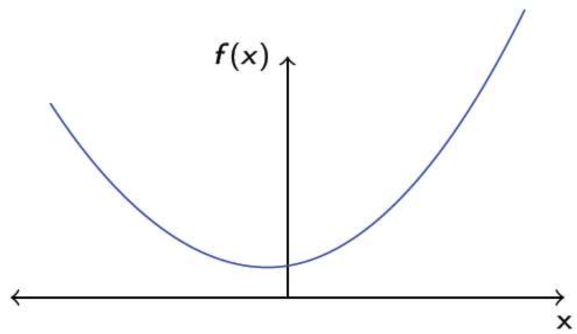
$$f'(x) = \frac{1}{x} - 1, \quad f''(x) = -\frac{1}{x^2}.$$

Thus the function is concave, and the FOC identifies a maximum equal to

$$f'(x) = 0 \Rightarrow \frac{1}{x} - 1 = 0 \Rightarrow x = 1.$$

Example 2

- Recall our function from a previous example. $f(x) = 1 + x + x^2$ defined over $\mathbb{R}$:



- That function was convex and thus the FOC was all we needed to find out that the stationary point $x = -\frac{1}{2}$ was an extreme point. Moreover, since we knew it was convex, we knew it was a minimum (rather than a maximum.)

General conditions for stationary points

- Suppose the function $f(x)$ has a stationary point at $x = c$, that is, $f'(c) = 0$.
- We can assign the point c to one of three categories on the basis of the first nonzero derivative obtained on repeatedly differentiating.
- Let $f^{(j)}(x)$ be the j^{th} derivative (provided it exists) and n the smallest number such that $f^{(n)}(c) \neq 0$. Then c is:

  1. **a local maximum** if n is even and $f^{(n)}(c) < 0$,
  2. **a local minimum** if n is even and $f^{(n)}(c) > 0$,
  3. **an inflection point** if n is odd.

Example 1

Find and classify the stationary point of $f(x) = -(x-1)^4$.
The stationary point c is such that $f'(c) = 0$, and

$$f'(x) = -4(x-1)^3.$$

Therefore, setting

$f'(x) = 0, \quad$ we solve $\quad -4(x-1)^3 = 0, \quad$ therefore, $\quad x = 1, \quad$ i.e. $\; c = 1.$

Differentiating again:

$$f''(x) = -12(x-1)^2, \qquad f''(1) = 0.$$

$$f^{(3)}(x) = -24(x-1), \qquad f^{(3)}(1) = 0.$$

$$f^{(4)}(x) = -24, \qquad f^{(4)}(1) = -24.$$

So the first nonzero derivative is the 4th, therefore $n = 4$ and we see that $f^{(4)}(1) < 0$.
Therefore, $x = 1$ is the value that gives a local maximum.

Example 2

Find and classify the stationary point of $f(x) = (x+1)^3$.
The stationary point c is such that c satisfies $f'(c) = 0$.
Taking the first derivative yields $f'(x) = 3(x+1)^2$.
For $f'(c) = 0$ it must be $c = -1$.

$$f''(x) = 6(x+1), \quad \text{so} \quad f''(-1) = 0.$$

$$f^{(3)}(x) = 6, \quad \text{hence,} \quad f^{(3)}(-1) = 6.$$

Hence, n is odd. So $x = -1$ is an inflection point.

A note on necessary and sufficient conditions

The distinction in logic between a necessary and a sufficient condition is important in both maths and economics.

Suppose there is some event or state of the world labelled X and another labelled Y. Then there are four ways in which a logical or causal connection may run from X to Y.

1) X is **necessary but not sufficient** for Y. We say: Y holds **only if** X holds.

Example

X is the state of being a male, and Y is the state of being a father. Then X is necessary for Y (you can't be a father and not be male). But X is not sufficient for Y (you can be a male without being a father).

2) X is **neither necessary nor sufficient** for Y.

Example

X is the state of being a male, and Y is the state of being a parent. Then X is not necessary for Y (a mother is a parent but not male). X is not sufficient for Y (you can be a male without being a parent).

3) X is **sufficient but not necessary** for Y. We say: Y holds **if** X holds.

Example

X is the state of being a father, and Y is the state of being a male. Then X is sufficient for Y (you cannot be a father without being male). But X is not necessary for Y (you can be male without being a father).

4) X is **both necessary and sufficient** for Y. We say: Y holds **if and only if** X holds. Or: Y holds **iff** X holds.

Example

X is the state of being married, and Y is the state of being a spouse. Being married and being a spouse have the same meaning. So anyone who is married is by definition a spouse, and anyone who is a spouse is by definition married.

Some things we need to know before leaving Topic I

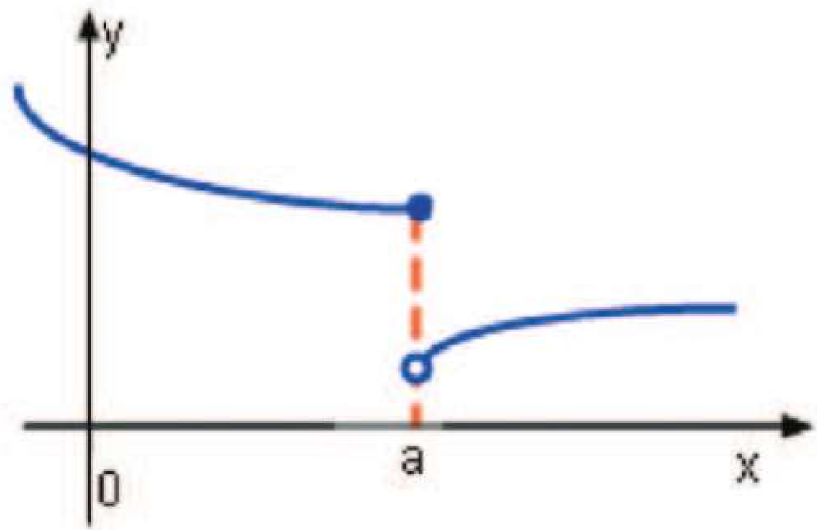
- Before moving on to Topic II we need to formally define some concepts used in our theorems.

1) Continuity
2) Limits
3) L'Hôpital's rule
4) Kinked functions

I. Continuity

Ch. 7.8

- Roughly speaking, a function is continuous if small changes in its argument do not cause big changes in the function's value.
- A counterexample:



- Graphically, a function is continuous if its graph is connected and has no jumps.
- What is the graph of a function $f(x)$? Simply the set of all points $(x, f(x))$ where x is in the domain of f.

- Tip: When you draw a continuous function, there mustn't be any holes and gaps — there should be no need to lift your hand.
- We can formalize the argument above by using limits.

Definition

A function $f(x)$ is continuous at $x = a$ if $\lim_{x \to a} f(x) = f(a)$.

This implicitly requires all of the following:

1. $f(a)$ is defined.
2. $\lim_{x \to a} f(x)$ exists.
3. The limit is equal to $f(a)$.

- Some examples follow to understand how things can go wrong.

Example 1

A case where $f(a)$ is not defined

- Consider the function:

$$f(x) = \frac{x^2 - 16}{4\sqrt{x} - 8} \text{ with domain } x \geq 0.$$

- Notice that when $x = 4$

$$f(4) = \frac{4^2 - 16}{4\sqrt{4} - 8} = \text{“}\frac{0}{0}\text{”}.$$

$\Rightarrow$ Thus $f(4)$ is not defined.

$\Rightarrow$ The function is **not continuous** everywhere.

- Does $\lim_{x \to 4}$ exist?
- We need to find a number A such that

$$\lim_{x \to 4} f(x) = A.$$

- Consider a sequence: $x = 3, 3.9, 3.99, 3.9999, \ldots$

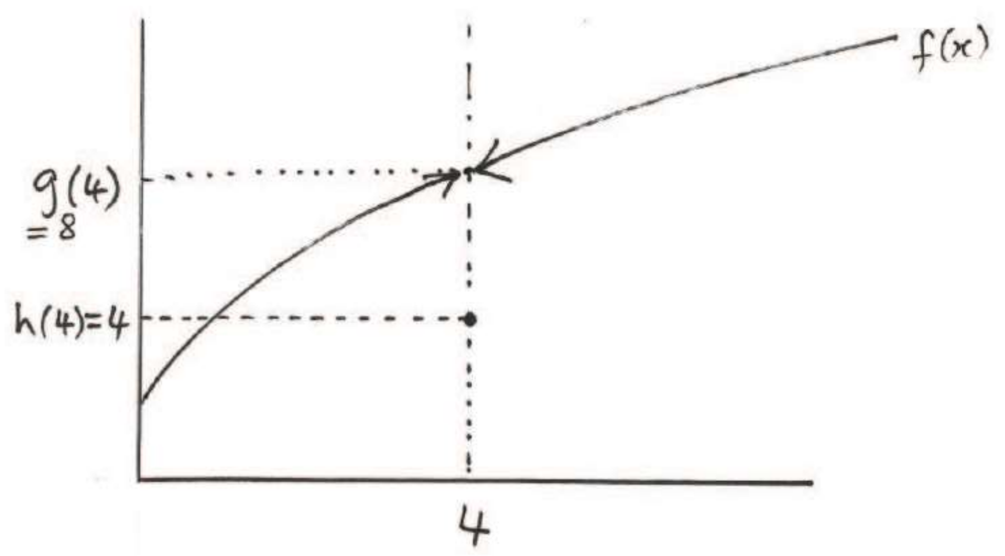
$$\begin{aligned} f(3) &= \frac{9-16}{4\sqrt{3}-8} = 6.53, \\ f(3.9) &= 7.85, \\ f(3.99) &= 7.98, \\ f(3.9999) &= 7.9999. \end{aligned}$$

- Thus the value of the function approaches $A = 8$ as x approaches 4 $\Leftrightarrow \lim_{x \to 4} f(x) = 8$.

- We can redefine the function for it to be continuous.
- How do we do that?

$$g(x) = \begin{cases} \frac{x^2-16}{4\sqrt{x}-8} & \text{if } x \geq 0, x \neq 4, \\ 8 & x = 4. \end{cases}$$

Now the function is continuous everywhere! By construction $\lim_{x\to 4} g(x) = 8$.



Example 2

A case where $f(a)$ is not defined and the limit is not finite

- Consider the function

$$f(x) = \frac{1}{(x+2)^2} \text{ with its domain being the set of real numbers } \mathbb{R}.$$

- Two things go wrong with this function.

1) The function is not defined over the whole domain:

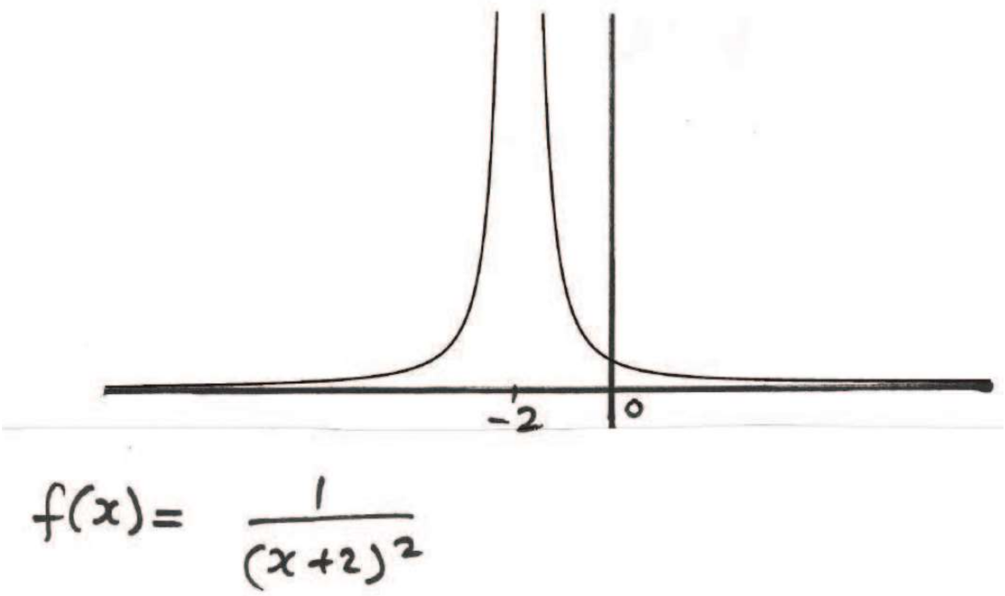
- $f(-2)$ is **not defined** since $f(-2) =$ "$\frac{1}{0}$".

2) Moreover, $\lim_{x \to -2} f(x)$ is not finite.

Consider the sequence: $x = -1.9, -1.999, ..., -1.999999$.

$$\begin{aligned} f(-1.9) &= 100, \\ f(-1.999) &= 1,000,000, \\ f(-1.999999) &= 10^{12}. \end{aligned}$$

- Therefore, as $x \to -2$ we have $f(x) \to \infty$, which is the same as saying that the function grows without bound as $x \to -2$.
- So the function $f(x)$ is **not continuous** at $x = -2$.



- This discontinuity cannot be corrected...

II. Going a bit further on limits

Ch. 7.9

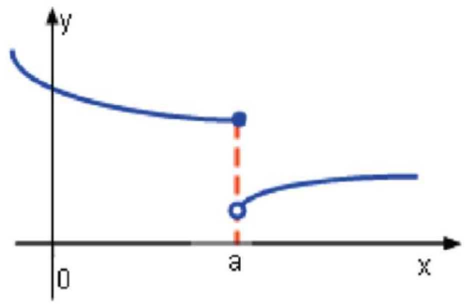
- In school you probably learned that $\lim_{x \longrightarrow a} f(x) = A$ means that we can make $f(x)$ as close to A as we make x sufficiently close to a.
- We can formalize that definition a bit further.

Definition

The function $f(x)$ has a limit A as x tends to a, denoted by $\lim_{x \rightarrow a} f(x) = A$, if for each number $\varepsilon > 0$ there exists a number $\delta > 0$ such that $|f(x) - A| < \varepsilon$ for every x with $0 < |x - a| < \delta$.

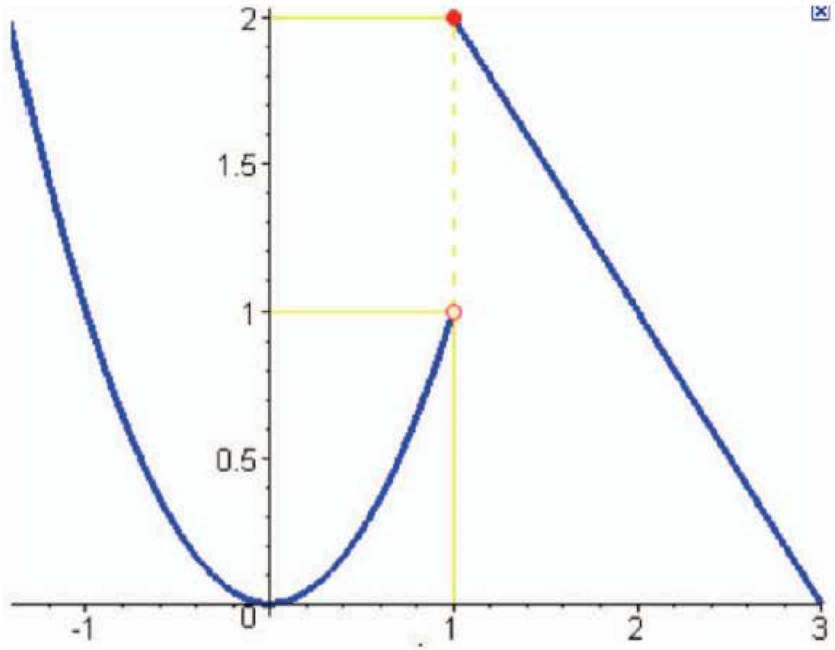
- This is the $\varepsilon - \delta$ definition (more advanced) where ε can be understood as 'error in the measurement of the value at the limit' and δ as 'distance to the limit point.' In plain English, the definition means that no matter how small ε is, we can find a δ that is small enough to be consistent with the value of ε.
- For a great online explanation see the `Khan Academy video` of this definition.

Consider the following discontinuous function:



- It seems that the function has a limit if we take values of x that are smaller than a.
- Mathematically, $f(x) \longrightarrow A$ as $x \longrightarrow a^-$, where the minus in the superscript of a denotes that we are taking values of x that come from the left.
- These are called one-sided limits.

- Depending on whether we are considering values of x coming from the left or the right of a, we call them one-sided limits from below or one-sided limits from above, or left and right limits.



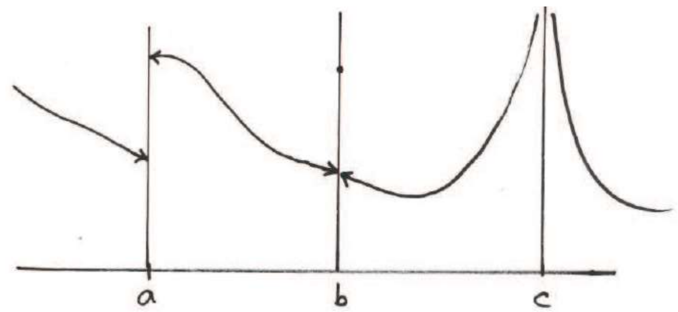
- We denote the limit from below (or left limit) as $\lim_{x \to a^-} f(x) = A$ and that from above (or the right) as $\lim_{x \to a^+} f(x) = B$.

Theorem

A necessary and sufficient condition for the limit at $x = a$ to exist is that the two one-sided limits of f at a exist and are equal.

One further example on discontinuity

This figure



shows a function with three discontinuities, $x = a, x = b, x = c$.

- At $x = a$ the limit from below is different than the limit from above. Therefore the limit as $x \longrightarrow a$ does not exist and thus the function is not continuous at a.
- At $x = b$ the limit exists but it is a jump up. The function is not continuous at $x = b$.
- At $x = c$ the function does not exist and the limit is not finite.

III. L'Hôpital's rule

Ch. 7.12

- Suppose that we want to find

$$\lim_{x \to a} \frac{f(x)}{g(x)}$$

  for two functions f and g, where $\lim_{x \to a} f(x) = 0$ and $\lim_{x \to a} g(x) = 0$. The ratio $\frac{f(a)}{g(a)}$ is indeterminate!

- For example:

$$f(x) = e^x - 1, \quad g(x) = x, \quad a = 0,$$

$$\lim_{x \to 0} (e^x - 1) = 0,$$

$$\lim_{x \to 0} x = 0.$$

- L'Hôpital's rule helps us deal with these situations.

- Suppose both functions f and g are **differentiable** and that $f(a) = g(a) = 0$.
- Then for $x \neq a$, we can write

$$\frac{f(x)}{g(x)} = \frac{[f(x) - f(a)]/(x-a)}{[g(x) - g(a)]/(x-a)}.$$

- Now, letting $x \longrightarrow a$, the right-hand side tends to

$$\frac{f'(a)}{g'(a)}.$$

Thus, we have the rule:

L'Hôpital's rule If $f(a) = g(a) = 0$ and $g'(a) \neq 0$, then

$$\lim_{x \longrightarrow a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

Named after the French mathematician Guillaume François Antoine, Marquis de l'Hôpital (1661-1704).

Example

Let $f(x) = e^x - 1, \quad g(x) = x, \quad a = 0.$

Taking the derivatives yields

$$f'(x) = e^x, \text{ therefore } f'(0) = 1,$$

$$g'(x) = 1, \text{ therefore } g'(0) = 1.$$

By L'Hôpital's rule

$$\lim_{x \to 0} \frac{e^x - 1}{x} = \frac{1}{1} = 1.$$

- Whenever $g'(a) = 0$, the rule does not work as stated above. Moreover, we need that $f'(a)$ and $g'(a)$ both exist.
- A more general form of L'Hôpital's rule (with weaker requirements) is as follows:

Theorem

Suppose f and g are differentiable over an interval (α, β) that contains a, except possibly at a, and suppose that $f(x)$ and $g(x)$ both tend to 0 as x tends to a. If $g'(x) \neq 0$ for all $x \neq a$ in (α, β) and if $\lim_{x \to a} \frac{f'(x)}{g'(x)} = L$, then

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)} = L.$$

This is true whether L is finite, ∞ or $-\infty$.

Example: Cobb-Douglas and CES

- A famous functional form in economics, especially for production functions, is the so-called *Cobb-Douglas* function $AK^{\alpha}L^{1-\alpha}$ where α is a weight parameter with $\alpha \in (0, 1)$. Think of $K > 0$ as the capital input, $L > 0$ as the labour input and $A > 0$ as a technology efficiency parameter.
- Another, more flexible functional form is the so-called *CES* ("constant elasticity of substitution") function

$$F(K, L) = A\left(\alpha K^{-\rho} + (1-\alpha)L^{-\rho}\right)^{-1/\rho},$$

  where $\rho \neq 0$ is a parameter. Note that $-1/\rho$ appears in the exponent so that we cannot set $\rho = 0$.
- Your task: Apply L'Hôpital's rule to $z = \ln\left[F(K, L)/A\right]$ as $\rho \to 0$ to show that $F(K, L)$ converges to the Cobb-Douglas function.

- Solution: We get

$$\begin{aligned} z &= \ln\left(\alpha K^{-\rho} + (1-\alpha)L^{-\rho}\right)^{-1/\rho} \\ &= -\ln\left(\alpha K^{-\rho} + (1-\alpha)L^{-\rho}\right)/\rho \to \text{“}0/0\text{”} \text{ as } \rho \to 0. \end{aligned}$$

- Note that $(\mathrm{d}/\mathrm{d}\rho)K^{-\rho} = (\mathrm{d}/\mathrm{d}\rho)e^{\ln(K^{-\rho})} = (\mathrm{d}/\mathrm{d}\rho)e^{-\rho\ln K} = \left(e^{-\rho\ln K}\right)(-\ln K) = -K^{-\rho}\ln K$. Analogously, it follows $(\mathrm{d}/\mathrm{d}\rho)L^{-\rho} = -L^{-\rho}\ln L$.
- Also note that $(\mathrm{d}/\mathrm{d}\rho)\left[-\ln\left(\alpha K^{-\rho} + (1-\alpha)L^{-\rho}\right)\right] = -\frac{(\mathrm{d}/\mathrm{d}\rho)[\alpha K^{-\rho}+(1-\alpha)L^{-\rho}]}{\alpha K^{-\rho}+(1-\alpha)L^{-\rho}}$ (outer derivative times inner derivative).
- Thus, applying L'Hôpital's rule yields

$$\begin{aligned} \lim_{\rho\to 0} z &= \lim_{\rho\to 0}\left(\frac{\alpha K^{-\rho}\ln K + (1-\alpha)L^{-\rho}\ln L}{\alpha K^{-\rho} + (1-\alpha)L^{-\rho}}\right)/1 \\ &= \alpha\ln K + (1-\alpha)\ln L = \ln\left(K^{\alpha}L^{1-\alpha}\right). \end{aligned}$$

- Hence, $e^z \to K^{\alpha}L^{1-\alpha}$ and the conclusion follows.

Reminder: Logarithms and exponentials

- $\ln(xy) = \ln(x) + \ln(y) \quad \text{and} \quad \ln(x/y) = \ln(x) - \ln(y)$
- $\ln(x^a) = a\ln(x)$
- Special cases: $\ln(1) = 0 \quad \text{and} \quad \ln(e) = 1$
- The derivative of the logarithmic function is

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln(x) = \frac{1}{x}.$$

- Notation: $e^x = \exp(x)$
- $e^{x+y} = e^x e^y$
- Special cases: $e^0 = 1 \quad \text{and} \quad e^{\ln(x)} = x$
- The derivative of the exponential function is

$$\frac{\mathrm{d}}{\mathrm{d}x}e^x = e^x.$$

By the chain rule we therefore have

$$\frac{\mathrm{d}}{\mathrm{d}x}e^{f(x)} = f'(x)e^{f(x)}.$$

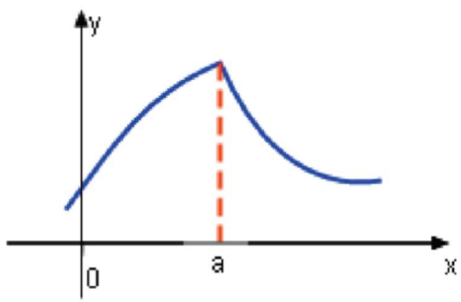
Continuity and differentiability

- One thing you may know and should be reminded of:

Theorem

If f is differentiable at $x = a$, then f is continuous at a.

- But a continuous function may not be differentiable:



(At the kink in a there is no unique tangent.)

IV. Kinked functions

- Recall that the derivative of $f(x)$ at a point a corresponds to

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

- We can extend this definition to one-sided limits:

$$f'(a^+) = \lim_{h \to 0^+} \frac{f(a+h) - f(a)}{h}, f'(a^-) = \lim_{h \to 0^-} \frac{f(a+h) - f(a)}{h}.$$

Definition

A continuous function at $x = a$ is said to have a kink at a if the one-sided limits of the first derivative f' exist and

$$f'(a^+) \neq f'(a^-).$$

(This definition will be important for your microeconomics modules.)

Topic II: Calculus of functions of many variables

EC123 Mathematical Techniques B

Dr Dennis Novy

University of Warwick

Autumn Term 2015

Topic II: Functions of many variables

Topic II's big picture:

- Until now we have only used univariate functions, that is, functions that have only *one* argument: $f(x)$.
- In economics, however, we are often interested in functions with *several* arguments: $f(x_1, x_2, \ldots, x_n)$.
- For example, demand depends not only on the price but also on the consumer's income; production functions depend not only on one input but on several inputs etc.
- Before discussing how to optimize such functions, we will study some key features:

II.1 Definition of functions with many variables

II.2 Level curves

II.3 Partial derivatives

II.4 Elasticities

II.1 Definition of functions with many variables
(Ch. 11.1, 11.5)

Definition

A function f of n variables $x_1, x_2, \ldots, x_n$ with domain D is a rule that assigns a number $f(x_1, x_2, \ldots, x_n)$ to each vector $x_1, x_2, \ldots, x_n$ in D.

Examples with $n = 2$:

- $f(x_1, x_2) = x_1 + x_2$
- $f(x_1, x_2) = (x_1 - x_2)^2$
- Basically, any way you can think of "combining" x_1 and x_2.

Examples with $n = 3$:

- $f(x_1, x_2, x_3) = x_1 + x_2 + x_3$
- $f(x_1, x_2, x_3) = (x_1 - x_2)^2 + \exp(x_3)$
- Basically, any way you can think of "combining" x_1, x_2 and x_3.

Examples

- Consider the function $f(x, y) = 2x + x^2y^3$. What are the values of $f(1, 0)$, $f(0, 1)$ and $f(a+1, b)$?
  1. $f(1, 0) = 2 \cdot 1 + 1^2 \cdot 0^3 = 2$
  2. $f(0, 1) = 2 \cdot 0 + 0^2 \cdot 1^3 = 0$
  3. $f(a+1, b) = 2(a+1) + (a+1)^2b^3 = (a+1)[2 + (a+1)b^3]$
- A very well-known function in economics is the Cobb-Douglas function defined as

$$f(x, y) = x^{\alpha}y^{\beta},$$

where α and β are constants.

II.2 Level curves

(Ch. 11.3)

- We will work with the case of bivariate functions, that is, functions with two arguments ($n = 2$).

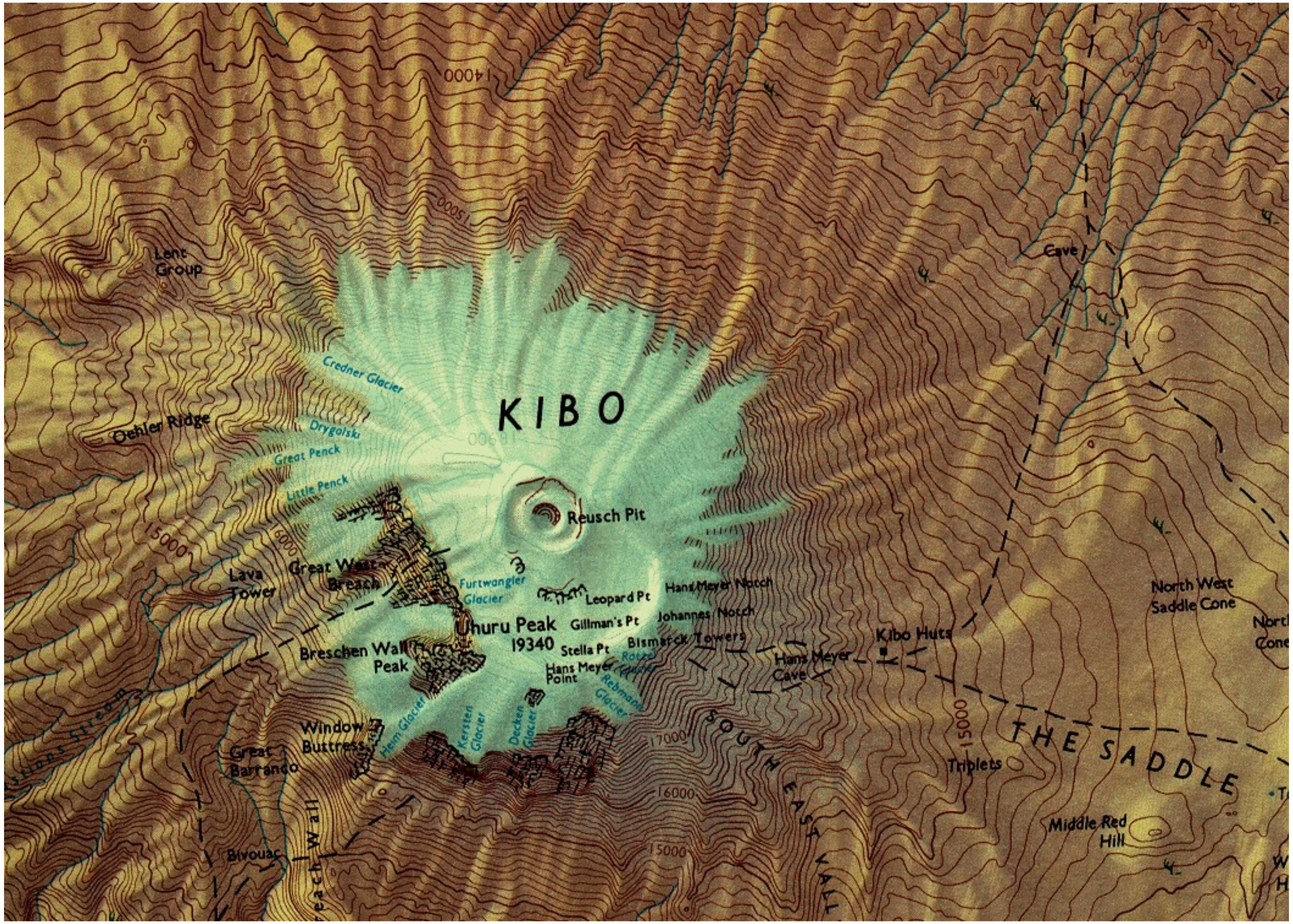
Definition

A level curve c of the function $f(x_1, x_2)$ corresponds to all the pairs (x_1, x_2) satisfying

$$f(x_1, x_2) = c.$$

- In other words, a level curve is the set of combinations of x_1 and x_2 that lead to the same value of the function.
- If we vary c, then we have several level curves for the same function.

- As a visual illustration, think of a map where the map maker includes contours corresponding to 1000 feet above sea level, 2000, 3000, 4000 feet etc. For instance, see Mount Kilimanjaro.

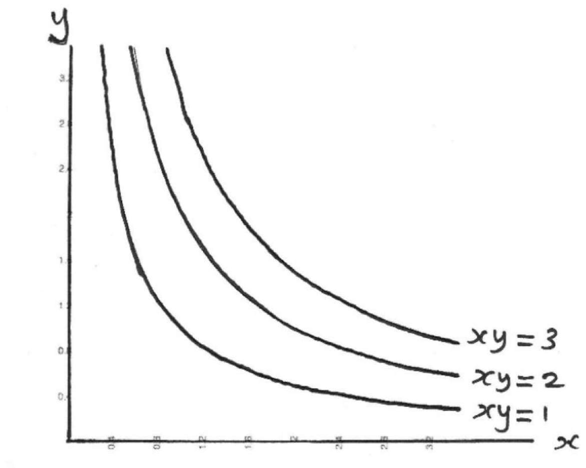

KIBO
Reusch Pit
Uhuru Peak
19340
Leopard Pt
Gillman's Pt
Stella Pt
Hans Meyer Point
Hans Meyer Notch
Johannes Notch
Bismarck Towers
Hans Meyer Cave
Kibo Huts
Furtwangler Glacier
Credner Glacier
Drygalski
Great Penck
Little Penck
Heim Glacier
Kersten Glacier
Decken Glacier
Rebmann Glacier
Great West Breach
Lava Tower
Breschen Wall
Peak
Window Buttress
Great Barranco
Bivouac
Oehler Ridge
Lent Group
Cave
North West Saddle Cone
THE SADDLE
Triplets
Middle Red Hill
SOUTH EAST VALL
14000
15000
16000
17000

Example 1

- Take the function $f(x, y) = xy$.
- The level curve is defined as $xy = c$.
- Suppose $c = 1$, then
  - $x = 1$ and $y = 1$ belong to the level curve c.
  - $x = 2$ and $y = \frac{1}{2}$ belong to the level curve c.
  - $x = 7$ and $y = \frac{1}{7}$ belong to the level curve c.
  - ...
  - More generally, for any x we pick, we can find a y that belongs to the level curve c. So we have lots of such pairs.
  - Since we have lots of (x, y) pairs that map to c, let's see whether we can plot them.

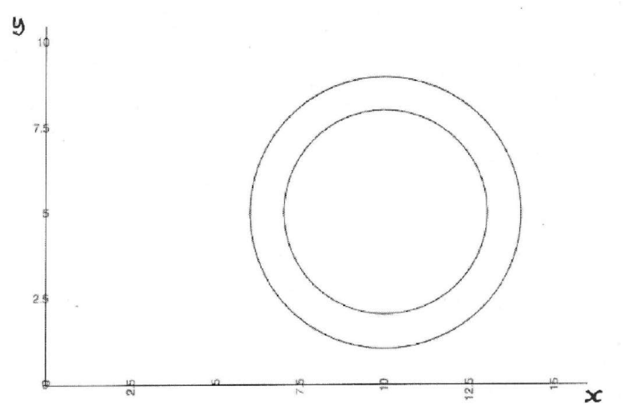
Level curves graphically

- Notice that the bigger x, the smaller y has to be. Thus we expect the level curves for $c = xy$ to have a negative slope when plotted in (x, y) space.



Example 2

- Consider the function $f(x, y) = (x - 19)^2 + (y - 5)^2$.
- Two level curves are:

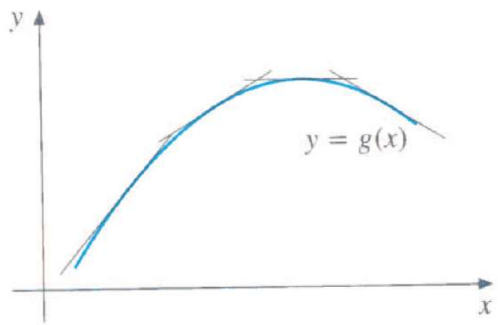


- Level curves have an important application in your microeconomics lectures: indifference curves.
- Indifference curves illustrate all the bundles or combinations of goods that leave the consumer indifferent, i.e., that give him or her the same level of utility.

II.3 Partial derivatives (Ch. 11.2, 11.6)

First-order partial derivatives

- Recall that in a univariate function $f(x)$ the derivative at a point $x_0 \in D$ is
$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$
- Therefore, if we change x_0, then the derivative measures the rate of change in $f(x)$ as its argument changes.



- How can we generalize the same idea of “how much the function changes” when f is multivariate?
- Answer: By focusing on the derivative of f when only **one** of its arguments changes whilst keeping the other arguments fixed.

- This type of derivative is called a "partial" derivative because we only change one argument of the function.
- Example with $n = 2$:
  - Take the function $f(x_1, x_2) = x_1^2 + x_2^3$.
  - There are two partial derivatives:

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 2x_1, \qquad \frac{\partial f(x_1, x_2)}{\partial x_2} = 3x_2^2.$$

  - Notice the rounded symbol ∂ instead of the usual d.
  - $\frac{\partial f(x_1,x_2)}{\partial x_1}$ represents the change in f when x_1 changes whilst x_2 is kept constant, and analogously for $\frac{\partial f(x_1,x_2)}{\partial x_2}$.
- Other notation: $\frac{\partial f(x_1,x_2)}{\partial x_1} \equiv f_1' \equiv f_{x_1}$ and $\frac{\partial f(x_1,x_2)}{\partial x_2} \equiv f_2' \equiv f_{x_2}$, where the symbol $\equiv$ means "is defined as."

Definition

$\frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_i}$ for $i = 1, \ldots, n$ is the partial derivative of f with respect to x_i whilst all other arguments x_j, $j \neq i$ are held constant.

Definition

The first partial derivative of f with respect to x_i at $(x_1, x_2, \ldots, x_n) \in D$ corresponds to

$$\frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}.$$

- This is nothing but the straightforward generalization of the definition in the univariate case.
- If the above limit does not exist, we say that the function is "*not differentiable* at $(x_1, x_2, \ldots, x_n)$."
- Partial derivatives are found using the same rules of differentiation as for ordinary derivatives in univariate functions: the chain rule, the product rule, the quotient rule etc. all apply. Revise those basic rules at home.

Example 1

Let

$$f(x_1, x_2) = x_1^2 + 3x_2^2 + 2x_1 x_2.$$

Then

$$\frac{\partial f}{\partial x_1} \equiv f_{x_1} = 2x_1 + 2x_2,$$

$$\frac{\partial f}{\partial x_2} \equiv f_{x_2} = 6x_2 + 2x_1.$$

It is often useful to write the first derivatives all together as a **vector**. For example, $\begin{bmatrix} f_{x_1} \\ f_{x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 + 2x_2 \\ 6x_2 + 2x_1 \end{bmatrix}$.

Example 2

Let

$$f(x_1, x_2) = x_1^{\frac{1}{2}} x_2.$$

Then

$$\frac{\partial f}{\partial x_1} = \frac{1}{2} x_1^{-\frac{1}{2}} x_2,$$

$$\frac{\partial f}{\partial x_2} = x_1^{\frac{1}{2}},$$

or

$$\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} f_{x_1} \\ f_{x_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} x_1^{-\frac{1}{2}} x_2 \\ x_1^{\frac{1}{2}} \end{bmatrix}.$$

Note that sometimes we will use the following notation: $y = x_1^{\frac{1}{2}} x_2$ and thus $\frac{\partial y}{\partial x_1} = \frac{1}{2} x_1^{-\frac{1}{2}} x_2$, $\frac{\partial y}{\partial x_2} = x_1^{\frac{1}{2}}$.

The chain rule

Suppose you have a 'chained' function $h(x) = f\left(g(x)\right)$. Its derivative with respect to x is

$$h'(x) = f'\left(g(x)\right) \times g'(x).$$

That is, multiply the outer derivative f' by the inner derivative g'.

Example:

$$h(x) = \left(7x^4 + 5x^2\right)^{12}$$

such that

$$\begin{aligned} f(g) &= g^{12}, \\ g(x) &= 7x^4 + 5x^2. \end{aligned}$$

It follows

$$\begin{aligned} f'(g) &= 12g^{11}, \quad g'(x) = 28x^3 + 10x \quad \text{and} \\ h'(x) &= \underbrace{12\left(7x^4 + 5x^2\right)^{11}}_{\text{outer derivative}} \times \underbrace{\left(28x^3 + 10x\right)}_{\text{inner derivative}}. \end{aligned}$$

The product rule

Suppose you have a product of two functions $h(x) = f(x)g(x)$. Its derivative with respect to x is

$$h'(x) = f'(x)g(x) + f(x)g'(x).$$

Example:

$$h(x) = 5x^3 \ln(x)$$

such that

$$\begin{aligned} f(x) &= 5x^3, \\ g(x) &= \ln(x). \end{aligned}$$

It follows

$$\begin{aligned} f'(x) &= 15x^2, \quad g'(x) = \frac{1}{x} \quad \text{and} \\ h'(x) &= \underbrace{15x^2 \ln(x)}_{f'(x)g(x)} + \underbrace{\frac{5x^3}{x}}_{f(x)g'(x)}. \end{aligned}$$

The quotient rule

Suppose you have a ratio of two functions $h(x) = f(x)/g(x)$. Its derivative with respect to x is

$$h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}.$$

Example:

$$h(x) = 5x^3 / \ln(x)$$

such that

$$\begin{aligned} f(x) &= 5x^3, \\ g(x) &= \ln(x). \end{aligned}$$

It follows

$$\begin{aligned} f'(x) &= 15x^2, \quad g'(x) = \frac{1}{x} \quad \text{and} \\ h'(x) &= \frac{15x^2 \ln(x) - 5x^3/x}{[\ln(x)]^2}. \end{aligned}$$

Second-order partial derivatives

- We can differentiate the first-order derivatives again to obtain second-order partial derivatives.
- Take a bivariate function $f(x_1, x_2)$. If we take the derivative of f_{x_1} with respect to x_1 one more time, we have a second-order partial derivative:

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right)$$

with x_2 held constant, and

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right)$$

with x_1 held constant. Careful: the correct notation is $\partial^2 f$ and ∂x_1^2, not ∂f^2 and $\partial^2 x_1$.

Cross-partial derivatives

- But if we can differentiate f_{x_1} with respect to x_1, why not also differentiate it with respect to x_2?
- This is what we will do to determine the second-order cross-partial derivatives:

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)$$

and

$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right).$$

- A pretty famous theorem (Young's theorem) states that these are equal if all the mth-order derivatives are continuous. In that case we say that the function is "continuously differentiable in D." If all partial derivatives up to order m exist, it is called a C^m function.

Hessian matrix

- In microeconomics it will be useful to gather all derivatives together in a *matrix* (which is a rectangular array of numbers.)
- The so-called *Hessian matrix* is a matrix that contains all the second partial derivatives of a multivariate function.

Definition

The Hessian matrix of $y = f(x_1, \dots, x_n)$ corresponds to

$$\begin{bmatrix} f''_{11} & \cdots & f''_{1j} & \cdots & f''_{1n} \\ \vdots & & \vdots & & \vdots \\ f''_{i1} & \cdots & f''_{ij} & \cdots & f''_{in} \\ \vdots & & \vdots & & \vdots \\ f''_{n1} & \cdots & f''_{nj} & \cdots & f''_{nn} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 y}{\partial x_1^2} & \cdots & \frac{\partial^2 y}{\partial x_1 \partial x_j} & \cdots & \frac{\partial^2 y}{\partial x_1 \partial x_n} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial^2 y}{\partial x_i \partial x_1} & \cdots & \frac{\partial^2 y}{\partial x_i \partial x_j} & \cdots & \frac{\partial^2 y}{\partial x_i \partial x_n} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial^2 y}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 y}{\partial x_n \partial x_j} & \cdots & \frac{\partial^2 y}{\partial x_n^2} \end{bmatrix},$$

where $f''_{ij} = f''_{ji} = \frac{\partial^2 y}{\partial x_i \partial x_j}$.

- In the $n = 2$ case the Hessian matrix simply corresponds to

$$\begin{bmatrix} f_{x_1 x_1} & f_{x_1 x_2} \\ f_{x_2 x_1} & f_{x_2 x_2} \end{bmatrix}.$$

- The matrix is named after the German mathematician Ludwig Otto Hesse (1811-1874).

Example 1 (continued)

Let

$$f(x_1, x_2) = x_1^2 + 3x_2^2 + 2x_1 x_2,$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + 2x_2 \quad \text{and} \quad \frac{\partial f}{\partial x_2} = 2x_1 + 6x_2.$$

Thus

$$\frac{\partial^2 f}{\partial x_1^2} = 2, \qquad \frac{\partial^2 f}{\partial x_2^2} = 6,$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = 2 \ \text{ and } \ \frac{\partial^2 f}{\partial x_2 \partial x_1} = 2.$$

The Hessian matrix H is therefore

$$H = \begin{bmatrix} f_{x_1 x_1} & f_{x_1 x_2} \\ f_{x_2 x_1} & f_{x_2 x_2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 6 \end{bmatrix}.$$

Example 2 (continued)

Let

$$f(x_1, x_2) = x_1^{\frac{1}{2}} x_2.$$

From above we know that

$$\frac{\partial f}{\partial x_1} = \frac{1}{2} x_1^{-\frac{1}{2}} x_2, \qquad \frac{\partial f}{\partial x_2} = x_1^{\frac{1}{2}}.$$

Thus

$$\frac{\partial^2 f}{\partial x_1^2} = -\frac{1}{4} x_1^{-\frac{3}{2}} x_2, \qquad \frac{\partial^2 f}{\partial x_2^2} = 0,$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = \frac{1}{2} x_1^{-\frac{1}{2}}.$$

Example 3

Let

$$f(x, y) = x^{\alpha} + xy^{\beta},$$

where α and β are constants. The first derivatives are

$$f_x = \alpha x^{\alpha-1} + y^{\beta},$$

$$f_y = \beta xy^{\beta-1}.$$

The second derivatives are

$$f_{xx} = \alpha(\alpha - 1)x^{\alpha-2},$$

$$f_{xy} = \beta y^{\beta-1},$$

$$f_{yy} = \beta(\beta - 1)xy^{\beta-2}.$$

Example 4: The marginal product

Consider an agricultural production function

$$Y = f(K, L, T).$$

For instance, Y could be the amount of wheat produced, $K > 0$ the invested capital (tractors, combine harvesters etc.), $L > 0$ the labour input (say, the number of hours the farmer works) and $T > 0$ is the amount of land.

Then $\partial Y/\partial K = f_K$ is called the *marginal product of capital*. It indicates how much wheat production changes if capital is changed (holding L and T fixed). Likewise, $\partial Y/\partial L = f_L$ is the marginal product of labour and $\partial Y/\partial T = f_T$ is the marginal product of land.

For example, suppose the Cobb-Douglas function

$$f(K, L, T) = AK^a L^b T^c,$$

where A, a, b and c are positive constants. It follows

$$f_K = AaK^{a-1} L^b T^c > 0.$$

II.4 Elasticities

Definition

An elasticity is a unit-free measure of the change in the function's value when one of its arguments changes.

- Does this sound familiar? This definition seems similar to the notion of a derivative.
- Elasticities are widely used in economic analysis as they answer many interesting questions. For example:
  - How much does demand decrease if we raise the price?
  - How much does production decrease if we diminish one particular input?
  - How much will a consumer buy of a particular product if his or her income increases?

Before talking about derivatives let's start from the very beginning.

- Suppose I have a function $y = f(x)$ and I want to know how y changes when its argument x changes.
- One first educated guess of how to measure such a change could be:

$$\frac{\Delta y}{\Delta x},$$

where Δ means "change" (to be precise, a *discrete* change Δy as opposed to an *infinitesimal* change dy or ∂y).

- What can go wrong with this definition?

- Suppose y is the amount of petrol bought by a motorist and x is the price of petrol.
- Suppose we observe that the price of petrol increases by 10 pence and as a result the motorist demands 30 liters less. Then $\frac{\Delta y}{\Delta x} = \frac{-30}{0.10} = -300$.
- Suppose we observe the same motorist in the U.S. The price of petrol increases by 16 cents (roughly the same as 10 pence at current exchange rates). Then $\frac{\Delta y}{\Delta x} = \frac{-30}{0.16} = -187.5$.
- Can you conclude that in the U.S. the demand for petrol is less sensitive to changes in the price? Of course not.

⇒ Our first-guess measure depends on the units chosen (pence vs. cents). This is arbitrary and therefore not satisfactory.

- Instead we want a unit-free measure. How do we do that?
- Take a function $f(x_1, x_2)$. For example, f is demand, x_1 is the price of petrol and x_2 is the income of the motorist.
- Informally, the elasticity of f with respect to x_1 corresponds to

$$\frac{\frac{\triangle f(x_1,x_2)}{f(x_1,x_2)}}{\frac{\triangle x_1}{x_1}}.$$

  - The numerator $\frac{\triangle f(x_1,x_2)}{f(x_1,x_2)}$ corresponds to the *percentage change* in the function's value.
  - The denominator $\frac{\triangle x_1}{x_1}$ corresponds to the *percentage change* in the argument's value (i.e. in the value of x_1.)
  - Thus the ratio indicates the percentage change in the function's value when one of its arguments (in this case x_1) changes by a certain percentage.
  - Is this what we want? Not quite yet. Why? Because the ratio depends on a discrete change Δ.
- An elasticity removes the problem by making Δ as small as possible. That is, it makes Δ infinitesimal.
- We are now ready to formally define an elasticity.

Definition

For a function $f(x_1, x_2)$ the partial elasticities of f with respect to x_1 and x_2 correspond to

$$El_{x_1} f = \frac{x_1}{f(x_1, x_2)} \frac{\partial f(x_1, x_2)}{\partial x_1}, \qquad El_{x_2} f = \frac{x_2}{f(x_1, x_2)} \frac{\partial f(x_1, x_2)}{\partial x_2}.$$

- The idea is to measure the change in the function's value when its arguments change marginally (i.e. infinitesimally).

Elasticities as logarithmic derivatives

- When x_1 and x_2 are positive variables, then one can show that

$$El_{x_1} f = \frac{x_1}{f(x_1, x_2)} \frac{\partial f(x_1, x_2)}{\partial x_1} = \frac{\partial \ln f}{\partial \ln x_1},$$

$$El_{x_2} f = \frac{x_2}{f(x_1, x_2)} \frac{\partial f(x_1, x_2)}{\partial x_2} = \frac{\partial \ln f}{\partial \ln x_2}.$$

- Why? Rewrite

$$\begin{aligned} \frac{\partial \ln f}{\partial \ln x_1} &= \frac{\partial \ln f}{\partial f} \frac{\partial f}{\partial \ln x_1} = \frac{\partial \ln f}{\partial f} \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial \ln x_1} \\ &= \frac{\partial \ln f}{\partial f} \frac{\partial f}{\partial x_1} \left(\frac{\partial \ln x_1}{\partial x_1} \right)^{-1} = \frac{1}{f} \frac{\partial f}{\partial x_1} \left(\frac{1}{x_1} \right)^{-1} = \frac{x_1}{f} \frac{\partial f}{\partial x_1}. \end{aligned}$$

- See a more detailed proof below.
- It is very common in economics to use variables in logarithmic form, so this version of the elasticity is widely used.

Proof for the univariate case

Use the chain rule and implicit differentiation:

$$\frac{d\ln(y)}{d\ln(x)} = \frac{d\ln(y)}{dx}\frac{dx}{d\ln(x)} = \frac{d\ln(y)}{dy}\frac{dy}{dx}\frac{dx}{d\ln(x)}.$$

Now recall

$$\frac{d\ln(y)}{dy} = \frac{1}{y}.$$

To find

$$\frac{dx}{d\ln(x)}$$

use the fact that $x = e^{\ln(x)}$ by definition of $\ln(x)$. Therefore

$$\frac{dx}{d\ln(x)} = \frac{d\left(e^{\ln(x)}\right)}{d\ln(x)} = e^{\ln(x)} = x.$$

Therefore

$$\frac{d\ln(y)}{d\ln(x)} = \frac{1}{y}\frac{dy}{dx}x = \frac{x}{y}\frac{dy}{dx}.$$

Example 1

Suppose the following demand function:

$$\ln(\text{quantity demanded}) = 10 - 2^{*} \ln(\text{price}).$$

We can rewrite it as

$$\ln(y) = 10 - 2\ln(x).$$

The elasticity of demand with respect to the price (or short: the *price elasticity of demand*) corresponds to

$$El_x y = -2 = \frac{d\ln(y)}{d\ln(x)}.$$

Example 2

Find the price elasticity of demand for the following demand function:

$$D(p) = 8000p^{-1.5}.$$

Taking the logarithm of the equation we obtain

$$\ln(D(p)) = \ln(8000) - 1.5\ln(p).$$

Thus

$$El_p D(p) = \frac{d\ln(D(p))}{d\ln p} = -1.5.$$

The interpretation of the elasticity is as follows. If the price increases by 1 percent, then the quantity demanded decreases by 1.5 percent.

Topic III: Multivariate optimization

EC123 Mathematical Techniques B

Dr Dennis Novy

University of Warwick

Autumn Term 2015

Topic III: Multivariate optimization

Topic III's big picture:

- In Topic I we learned how to solve optimization problems where the objective function was a univariate function.
- In economics, however, we are often interested in functions with several arguments $f(x_1, x_2, \ldots, x_n)$.
- For example, demand depends not only on the price but also on the consumer's income and preferences; production functions depend not only on one input but on several inputs etc.
- So after reviewing these functions (Topic II) we will finally study how to solve optimization problems with several variables.

III.1 The two-variable case: necessary conditions

III.2 The two-variable case: sufficient conditions

III.3 Local extreme points

III.4 Extreme value theorem

III.5 The n-variable case

III.1 The two-variable case: necessary conditions
(Ch. 13.1)

- In the univariate case we learned that an interior point in the domain of a continuously differentiable function could only be an extreme point if it was a stationary point, i.e. if the first derivative of the function was zero at that point.
- The following theorem extends this reasoning to the multivariate case.

Theorem

A differentiable function $z = f(x, y)$ can have a maximum or a minimum at an interior point $(x_0, y_0) \in D$ only if it is a stationary point, that is, if the point (x_0, y_0) satisfies

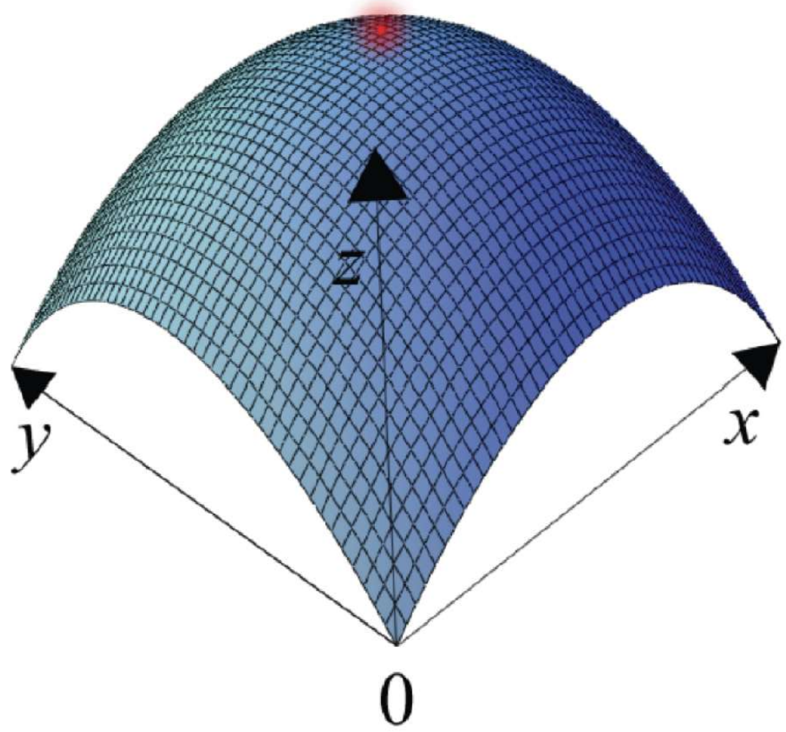
$$f_1'(x_0, y_0) = 0, \qquad f_2'(x_0, y_0) = 0 \quad \textit{(FOCs)}$$

- where FOCs means **first-order conditions**.

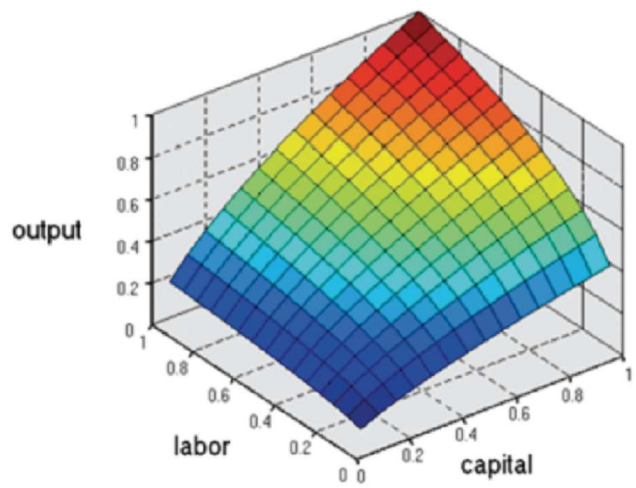
- The logic of this theorem mimics that of the univariate case. To see this note:
  - Given some value y_0 the function $f(x, y)$ can be thought of as a univariate function $g(x) \equiv f(x, y_0)$.
  - This univariate function achieves a maximum or minimum at x_0 only if x_0 is a stationary point, i.e. if $g'(x_0) = 0$.
  - However, $g'(x_0) = f'(x_0, y_0)$ and thus $g'(x_0) = 0 \Rightarrow f_1'(x_0, y_0) = 0$.
  - You can do the same for a function $h(y) \equiv f(x_0, y)$ to conclude that a necessary condition for y_0 given x_0 to be an extreme point is $f_2'(x_0, y_0) = 0$.

Lost? Review the univariate case!

- Graphically $z = f(x, y)$ can be depicted as



- Why do we say the FOC is a necessary condition for an **interior** extreme point?
- Consider a production function $z = f(x, y)$ where z is output, x is capital and y is labour and where $f(x, y) = x^{0.35}y^{0.65}$ (Cobb-Douglas).



- Without restrictions the maximum output is a corner solution where the derivative does not exist.

Example

Suppose $f(x, y) = -2x^2 - 2xy - 2y^2 + 36x + 42y - 158$ has a maximum point. Find it.

- The first partial derivatives are

$$f_1'(x, y) = -4x - 2y + 36, \qquad f_2'(x, y) = -2x - 4y + 42.$$

- The FOCs follow as

$$-4x - 2y + 36 = 0, \qquad -2x - 4y + 42 = 0.$$

- Thus we need to solve this **system of equations**: find the (x, y) pair that satisfies both equations **simultaneously**.
- Rearranging the first equation we get $4x = -2y + 36 \iff -2x = y - 18$.
- Now we can substitute this result into the second equation: $\Rightarrow [y - 18] - 4y + 42 = 0 \iff y = 8$.
- Plugging $y = 8$ back into the first equation to obtain x, we find that the maximum is at $(5, 8)$.

III.2 The two-variable case: sufficient conditions
(Ch. 13.2)

- Recall from the univariate case that an extreme point has to be a stationary point (unless you have a corner solution), but not all stationary points are extreme points.
  - In other words, the FOCs were necessary but not sufficient to find the maximum/minimum of a function.
  - As an illustration we went through a couple of examples
    1. where the function was strictly increasing and thus there were no stationary points (extreme points were at the boundaries),
    2. where the function had several "valleys and peaks" and thus the stationary points were only local maxima or local minima.
- We discussed that a sufficient condition for a stationary point to be an extreme point was that the function be convex/concave.
- We now extend this notion to the multivariate case.

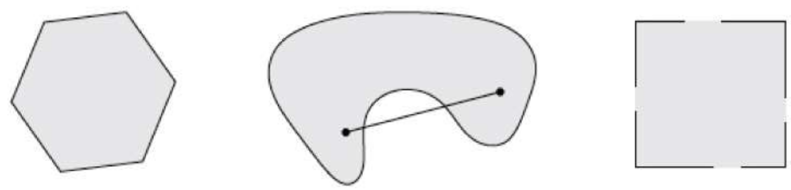
- To define convexity/concavity of a multivariate function, we will need the following definition.

Definition

A set S is convex if for all elements of the set s_1, s_2 as well as $\lambda s_1 + (1 - \lambda)s_2$ for any $\lambda \in [0, 1]$ belong to the set S.

- The expression $\lambda s_1 + (1 - \lambda)s_2$ can also be called a 'weighted average' of s_1 and s_2 (because the weights λ and $1 - \lambda$ add up to 1), or more generally a 'linear combination' (for a linear combination the weights would not necessarily have to add up to 1).
- Note that this definition is about a convex *set*, not a convex function.
- In rough English the meaning of a convex set is that if you draw a straight line (weighted average) that passes through s_1 and s_2, all the points on that line that are located between s_1 and s_2 (as well as s_1 and s_2 themselves) must be part of the set. There must not be any holes or gaps in the set on that stretch.

- Example: One convex set and two non-convex sets.



- Note that S is not necessarily a subset of $\mathbb{R}$. It could be a subset of $\mathbb{R}^n$ and thus an element of S could be a vector.

- We can now define the concavity/convexity of a function.

Definition

A function $f(\mathbf{x}) \equiv f(x_1, x_2)$ defined over a convex set S is concave if

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

for all $\mathbf{x}$ and $\mathbf{y}$ in S and for any $\lambda \in [0, 1]$. If the inequality is reversed $(\leq)$, the function is convex.

- Notice the **bold** notation for $\mathbf{x}$ and $\mathbf{y}$. This means they are vectors.
- Also recall that a function f is concave if and only if $-f$ is convex, and vice versa.
- If for $\lambda \in (0, 1)$ the inequalities are strict ($>$ and $<$ as opposed to $\geq$ and $\leq$), we say that the function is strictly concave or strictly convex.

- The previous definition is hard to check. However, we can relate it to the derivatives of the function.

Theorem

Suppose the function $f(x, y)$ is twice differentiable over a convex set S. Then f is said to be

(a) concave if for all (x, y) in S

$$f''_{11}(x,y) \leq 0, \quad f''_{22}(x,y) \leq 0 \quad \textit{and} \quad f''_{11}(x,y)f''_{22}(x,y) - \left(f''_{12}(x,y)\right)^2 \geq 0,$$

(b) convex if for all (x, y) in S

$$f''_{11}(x,y) \geq 0, \quad f''_{22}(x,y) \geq 0 \quad \textit{and} \quad f''_{11}(x,y)f''_{22}(x,y) - \left(f''_{12}(x,y)\right)^2 \geq 0.$$

- As we will learn when we study matrices, the expressions above on the right correspond to the *determinant* of the Hessian matrix.

Back to sufficient conditions

- Given our definition we are ready to pin down the sufficient conditions for extreme points in the multivariate case.

Theorem

Suppose (x_0, y_0) is an interior stationary point for a twice continuously differentiable function $f(x, y)$ defined over a convex set S.
(a) If f is concave, then (x_0, y_0) is a maximum point for f in S.
(b) If f is convex, then (x_0, y_0) is a minimum point for f in S.

- This theorem is the extension of the one in the univariate case and thus **it is also one of the most important things you will learn in mathematical techniques.**

Example

Find the values of x and y which maximize the function

$$f(x, y) = -2x^2 - y^2 + 4x + 4y - 3.$$

First-order conditions: $\quad f_x = 0 \quad \text{and} \quad f_y = 0$. Thus

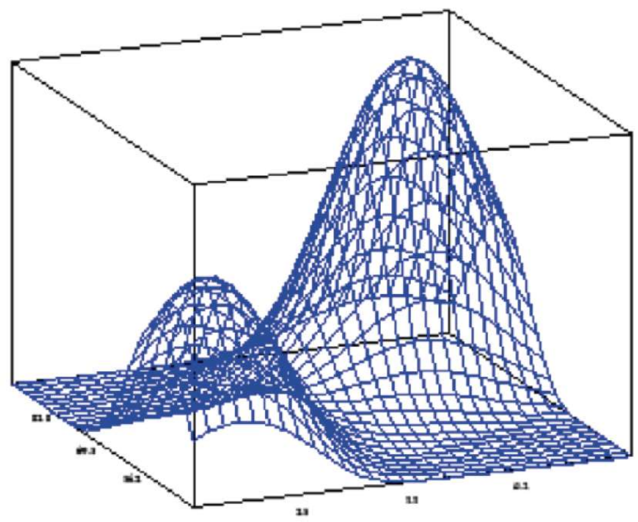
$$f_x = -4x + 4 = 0 \quad \Leftrightarrow \quad x = 1,$$
$$f_y = -2y + 4 = 0 \quad \Leftrightarrow \quad y = 2.$$

Second-order conditions for a maximum: $f_{xx} \leq 0, f_{yy} \leq 0$ and $f_{xx}f_{yy} - f_{xy}^2 \geq 0$. Thus

$$f_{xx} = -4, \quad f_{yy} = -2, \quad f_{xy} = 0,$$
$$f_{xx}f_{yy} - f_{xy}^2 = (-4)(-2) - 0 = 8 > 0.$$

Therefore, the function has a maximum at $x = 1, y = 2$.

- Note that the conditions in the theorem are sufficient but not necessary. That is, if they hold we know *for sure* there is an extreme point, but extreme points can exist even when these conditions do not hold.
- For example:



- There is clearly an interior maximum but the function is not concave over the whole domain.

III.3 Local extreme points
(Ch. 13.3)

- Just as in the univariate case, some extreme points are not maxima nor minima over the whole domain but they are maxima or minima for a subset of the domain. We call those local extreme points.

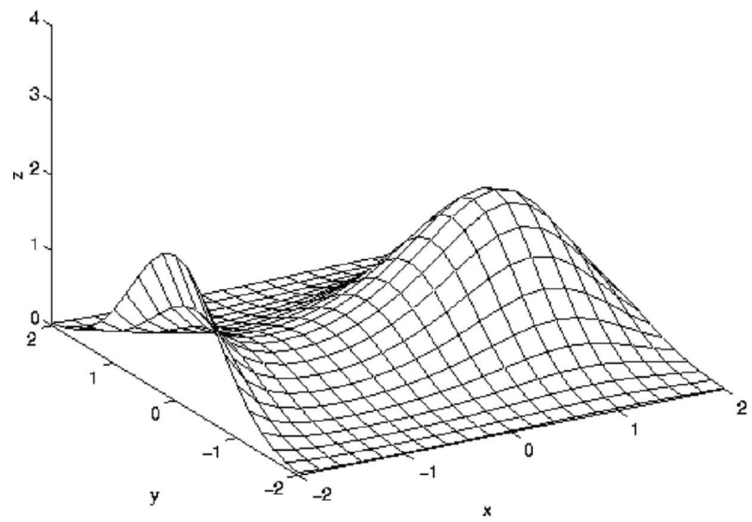
$\Rightarrow$ An extreme point is a local extreme point but a local extreme point is not necessarily an extreme point.

- Because of the theorem stating necessary conditions we know that local extreme points are stationary points.
- A stationary point can be a (local) maximum, a (local) minimum or a *saddle point*.
- We will now study saddle points.

- One type of stationary point that is not an extreme point is the so-called *saddle point*.

Definition

A saddle point (x_0, y_0) is a stationary point with the property that arbitrarily close to (x_0, y_0), there exist points (x, y) with $f(x, y) < f(x_0, y_0)$ and also points with $f(x, y) > f(x_0, y_0)$.



Example

Consider the function $f(x, y) = x^2 - y^2$.

- First we take derivatives: $f_x = 2x$ and $f_y = -2y$.
- Thus we have found our stationary point: $(x_0, y_0) = (0, 0)$.
- To see that $(0, 0)$ is a saddle point note that points very close to $(0, 0)$ holding $y_0 = 0$ fixed are positive:

$$f(x, 0) = x^2 > 0 \quad \text{for all } x.$$

Meanwhile points very close to $(0, 0)$ with $x_0 = 0$ held fixed are negative:

$$f(0, y) = -y^2 < 0 \quad \text{for all } y.$$

Theorem

Suppose $f(x, y)$ is a function with continuous second-order partial derivatives in a domain S, and let (x_0, y_0) be an interior stationary point of f. Write $A = f''_{11}(x_0, y_0)$, $\quad B = f''_{12}(x_0, y_0)$, $\quad C = f''_{22}(x_0, y_0)$.

1. *If $A < 0$ and $AC - B^2 > 0$, then (x_0, y_0) is a (strict) local maximum point.*
2. *If $A > 0$ and $AC - B^2 > 0$, then (x_0, y_0) is a (strict) local minimum point.*
3. *If $AC - B^2 < 0$, then (x_0, y_0) is a saddle point.*
4. *If $AC - B^2 = 0$, then (x_0, y_0) could be a local maximum, a local minimum or a saddle point.*

- (1), (2) and (3) are called (local) second-order conditions.
- Note that $AC - B^2 > 0$ in (1) implies $AC > B^2 \geq 0$, and so $AC > 0$. Thus, if $A < 0$, then also $C < 0$. The condition $C = f''_{22}(x_0, y_0) < 0$ is therefore indirectly included in (1). Analogous for (2).

Example continued

Consider again the function $f(x,y) = x^2 - y^2$.

- The first derivatives are $f_x = 2x$ and $f_y = -2y$. The stationary point is thus $(x_0, y_0) = (0,0)$.
- The second derivatives are $f_{xx} = 2$, $f_{xy} = f_{yx} = 0$ and $f_{yy} = -2$.
- According to the above theorem we have

$$\begin{aligned} & AC - B^2 \\ = \; & f_{xx}(x_0, y_0) f_{yy}(x_0, y_0) - (f_{xy}(x_0, y_0))^2 \\ = \; & 2^*(-2) - 0^2 = -4 < 0. \end{aligned}$$

- Thus (x_0, y_0) is a saddle point.

III.4 The extreme value theorem
(Ch. 13.5)

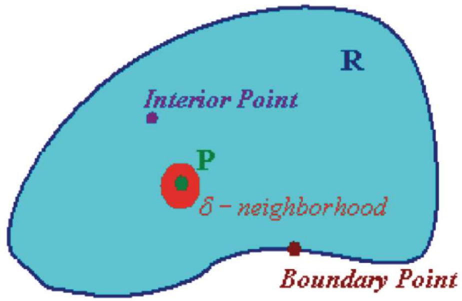
- Recall that the EVT provides sufficient conditions for extreme points to exist. We studied it in the univariate case.
- We now extend the EVT to the multivariate case.
- Before introducing the theorem we will study some further properties of sets that are widely used in economic theory.
- We will study
  1. interior points,
  2. boundary points,
  3. closed and open sets,
  4. bounded sets,
  5. compact sets.

Definition

A point (a, b) is an interior point if there exists a circle centred at the point such that all points strictly inside the circle lie in the set S.

Definition

A point (c, d) is a boundary point of the set S if *every* circle centred at (c, d) contains points in S as well as points lying in its complement (i.e. outside the set S).

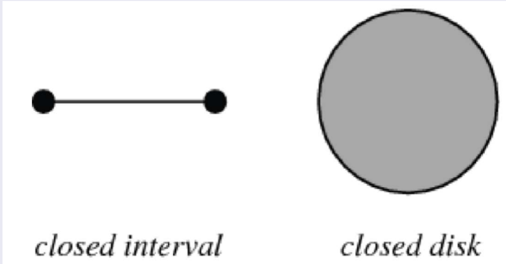


- Note that a boundary point does not necessarily lie in the set.
- Example: The point $(1, 1)$ is a boundary point of the set $\{(x, y)|x < 1, y < 1\}$.

- Therefore we have the following definitions:

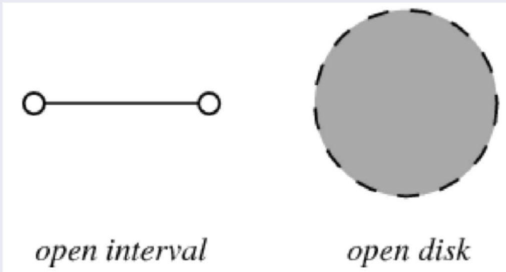
Definition

A set is closed if it contains all its boundary points.



Definition

A set S is open if it consists only of interior points.



- Some sets can be neither open nor closed like $(0, 1]$.

Examples of open sets

1. Let $S_1 = \{(x, y) | x < 1, y < 1\}$.
   For any point in S_1, we can draw a circle around it such that it is completely within S_1. What is the key feature? It is $<$ (as opposed to $\leq$).

2. Let $S_2 = \{(x, y) | x > 0, y > 0\}$.

   This is the set of pairs of strictly positive values of x and y. For any point in S_2, we can draw a circle around it such that it is completely within S_2.

3. Let $S_3 = \{(x, y) | -\infty < x < \infty, -\infty < y < \infty\}$.

   This is the set of pairs of numbers without restriction. (Infinity does not constitute a boundary point so that S_3 only consists of interior points.)

Examples of closed sets

1. $S_4 = \{(x, y) | x \geq 0, y \geq 0\}$
2. $S_5 = \{(x, y) | px + qy \leq m, x \geq 0, y \geq 0\}$. What is the latter set? It is the budget set in consumer theory. Here x, y are quantities of two goods purchased by a consumer, and p, q are the respective prices of the goods, and m is the income of the consumer.
3. In general, if $g(x, y)$ is a continuous function, then given a constant c the sets defined by $g(x, y) \geq c, g(x, y) = c$, $g(x, y) \leq c$ are closed (for example, the set $\{(x, y) | g(x, y) \leq c\}$). The sets defined by $g(x, y) > c$, $g(x, y) < c$ are open (for example, the set $\{(x, y) | g(x, y) < c\}$).

Definition

A set is bounded if it can be contained within a sufficiently large circle.

- For example, the budget set is bounded because m is finite.
- The diagrams we used when showing closed and open sets were all bounded.
- The set $\{(x, y) | x \geq 0, y \geq 0\}$ is not bounded.

Relation between closed and bounded sets

- It may seem that all sets that are bounded are closed and all sets that are closed are bounded...
- But such statements would be incorrect. Why? Here are counter-examples:
  1. The set $\{(x, y)|0 \leq x < 1, 0 \leq y < 1\}$ is bounded but it is not closed.
  2. The set $\{(x, y)|x \geq 1, y \geq 0\}$ is closed but it is not bounded.
- Thus we have the following type of set:

Definition

A set is compact if it is both closed and bounded.

Back to the existence of extreme points

- The EVT goes as follows:

Definition

Let $f(x, y)$ be a continuous function through a nonempty, closed and bounded set S. Then there exist points (a, b),(c, d) in S where f has a minimum and a maximum value, respectively. That is,

$$f(a, b) \leq f(x, y) \leq f(c, d) \quad \text{for all} \quad x, y \in S.$$

- The EVT only tells us when extreme points exist, but it says nothing about how to find those points.
- Moreover, the theorem gives us sufficient conditions, not necessary conditions.

- We now examine a general procedure for finding extreme points of a differentiable function $f(x, y)$ on a closed, bounded (i.e. compact) set S.
- When the set is compact, an extreme point can be either a stationary interior point or a boundary point. Thus the method has to consider all possibilities.

A "scientific recipe" to find extreme points

1. Find all stationary points in the interior of S.
2. Find the largest and smallest values of $f(x, y)$ on the boundary.
3. Compare the values of f obtained in steps 1 and 2.

Example

Find the extreme points and extreme values of $f(x, y)$ in S defined as

$$f(x, y) = x^2 + y^2 + y - 1, \quad S = \{(x, y) | x^2 + y^2 \leq 1\}.$$

(S is compact. Note that S delineates the unit circle.)

- According to the extreme value theorem $f(x, y)$ has both a maximum and a minimum over S.
- Let's follow the recipe.

1. The FOCs are $f_x = 2x = 0$, thus $x = 0$, and $f_y = 2y + 1 = 0$, thus $y = -\frac{1}{2}$.
   Thus there is only one stationary point at $(0, -\frac{1}{2})$.
   The value of the function at the stationary point is $f(0, -\frac{1}{2}) = -\frac{5}{4}$.
2. The boundary of S is a circle given by $x^2 + y^2 = 1$.
   The function is $f(x, y) = (x^2 + y^2) + y - 1$, and on the boundary it is $f(x, y) = 1 + y - 1 = y$. And y can go from -1 to 1 in S (otherwise we would step outside the unit circle). Therefore, the maximum value is $f(0, 1) = 1$ and the minimum value is $f(0, -1) = -1$.
3. Comparing f at these three points we arrive at the following conclusion:
   There is a maximum at $x = 0, y = 1$ with $f = 1$.
   There is a minimum at $x = 0, y = -\frac{1}{2}$ with $f = -\frac{5}{4}$ (note that $-\frac{5}{4} < -1$).
   We therefore have an interior minimum and a maximum on the boundary of the set.

III.5 The n-variable case
(Ch. 13.6)

- No major new insights! We simply restate everything with minor modifications.
- For simplicity we use vector notation: $\mathbf{x} = (x_1, x_2, ..., x_n)$. We represent all these variables by a single symbol (in **bold**) and write $f(\mathbf{x})$.
- Now the set S is not a subset of the two-dimensional (x, y) plane but of $\mathbb{R}^n$, the Euclidean n-dimensional space.

Definition

The Euclidean n-dimensional space $\mathbb{R}^n$ is the set of all possible n-vectors $\mathbf{x} = (x_1, x_2, ..., x_n)$ whose elements are real numbers, $x_i \in \mathbb{R}$.

- For $n = 1, 2$ and 3, $\mathbb{R}^n$ is a line, a plane and a space, respectively. But there is no geometrical interpretation for $n \geq 4$, which is a hyperspace.

Distance in the Euclidean space

- To generalize the concepts of interior and boundary points, closed, open, bounded and compact sets we need to define a concept of distance in the Euclidean space.

Definition

The Euclidean distance between two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ is equal to

$$||\mathbf{x} - \mathbf{y}|| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}.$$

Definition

An open ball with centre $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and radius r is the set of points such that $||\mathbf{x} - \mathbf{a}|| < r$.

- The definitions of interior and boundary points, closed, open, bounded and compact sets can now be immediately extended to the n-variable case if we replace "circle" with "ball."

Example

- Suppose the vectors **x** and **y** have two elements each:

$$\mathbf{x} = (6, 7),$$
$$\mathbf{y} = (2, 4).$$

- Then the Euclidean distance follows as

$$\begin{aligned} ||\mathbf{x} - \mathbf{y}|| &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \\ &= \sqrt{(6-2)^2 + (7-4)^2} \\ &= \sqrt{4^2 + 3^2} = \sqrt{25} = 5. \end{aligned}$$

- The interpretation in two-dimensional space is that a straight line drawn between **x** and **y** is exactly five units long.
  - You can verify this with Pythagoras' famous theorem for right-angled triangles: $a^2 + b^2 = c^2$.
  - If you draw the points $\mathbf{x} = (6, 7)$ and $\mathbf{y} = (2, 4)$ into a two-dimensional diagram, then $a = 6 - 2 = 4$ is the horizontal distance and $b = 7 - 4 = 3$ is the vertical distance in the associated triangle. It follows $c = 5$.

- Now we are ready to restate the theorems for the n-variable case.
- First, the EVT:

Theorem

Let $f(\mathbf{x})$ be continuous on a nonempty, closed and bounded set S in $\mathbb{R}^n$ (written $S \subseteq \mathbb{R}^n$). Then there exist points $\mathbf{c}, \mathbf{d} \in S$ where f has a minimum and maximum, respectively. That is,

$$f(\mathbf{c}) \leq f(\mathbf{x}) \leq f(\mathbf{d}) \quad \text{for all } \mathbf{x} \in S.$$

- Points $\mathbf{c}$ and $\mathbf{d}$ must lie either in the **interior** of S or on the **boundary** of S. If $f(\mathbf{x})$ is differentiable, then any minimum or maximum point in the interior must be a stationary point satisfying the first-order conditions.

- Second, the necessary FOC:

Theorem

Let $f(\mathbf{x})$ be a differentiable function defined on a set $S \subseteq \mathbb{R}^n$ and let $\mathbf{c}$ be an interior point in S. A necessary condition for $\mathbf{c}$ to be a maximum or a minimum is that $\mathbf{c}$ is a stationary point for f. That is, $\mathbf{x} = \mathbf{c}$ satisfies the condition that all the partial first-order derivatives be zero:

$$f_i'(\mathbf{c}) = 0, \qquad i = 1, 2, ..., n.$$

- The generalization of the second-order conditions is considerably more complicated than in the $n = 2$ case. The classification of stationary points as maxima, minima or saddle points depends on the Hessian matrix of second-order partial derivatives.
- The conditions are in terms of determinants of certain sub-matrices of the Hessian. We leave this until later.

One final (useful) result for optimization problems

- A simple result that is often very useful in solving optimization problems is the following:

Maximizing a function is equivalent to maximizing a strictly increasing transformation of that function.

- We illustrate this with an example. Consider the problem of maximizing $f(x) = e^{-x^2}$.
  - The first derivative is $f'(x) = -2xe^{-x^2}$.
  - So the FOC is $f'(x) = 0 \quad \Leftrightarrow \quad x = 0$.
  - The SOC is $f''(x) = 4x^2e^{-x^2} - 2e^{-x^2}$.
  - Evaluation of the SOC at $x = 0$ gives us $f''(x) = -2 < 0$. Therefore $x = 0$ is a maximum.
- Now take a function $F(x)$ where $F'(x) > 0$ for all x (this is an increasing function).
- We can show that maximizing the transformed function $F(f(x))$ is equivalent to maximizing $f(x)$.

- For example, take $F(x) = \ln(x)$ where $x > 0$.
  - We know from the definition of the ln function that $\ln(e^a) = a$ for any real number a.
  - Thus $F(f(x)) = \ln(f(x)) = \ln(e^{-x^2}) = -x^2 = g(x)$.
  - The FOC for maximizing $g(x)$ is $g'(x) = 0 \Leftrightarrow -2x = 0 \Leftrightarrow x = 0$.
  - The SOC is $g''(x) = -2 < 0$. Thus, we have a maximum.
- This monotonically increasing transformation has made the analysis much easier.
- This technique can be used for any n. When n is large it can result in considerable simplification. For example, the function $e^{x^2+2xy^2-y^3}$ has the same extreme points as the function $x^2 + 2xy^2 - y^3$.

Theorem

Suppose $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ *is defined over a set* S *in* $\mathbb{R}^n$ *and let* F *be a function of one variable defined over the range of* f. *Define* g *over* S *by*

$$g(\mathbf{x}) = F(f(\mathbf{x})).$$

Then

(a) *if* F *is increasing and* $\mathbf{c}$ *maximizes (minimizes)* f *over* S, *then* $\mathbf{c}$ *also maximizes (minimizes)* g *over* S.

(b) *if* F *is strictly increasing, then* $\mathbf{c}$ *maximizes (minimizes)* f *over* S *if and only if* $\mathbf{c}$ *maximizes (minimizes)* g *over* S.

Topic IV: Constrained optimization

EC123 Mathematical Techniques B

Dr Dennis Novy

University of Warwick

Autumn Term 2015

Topic IV: Constrained optimization
(Ch. 14)

Topic IV's big picture:

- We already know how to maximize/minimize univariate and multivariate functions.
- We now study how to maximize functions when we impose restrictions over the values the arguments of the function can take. We call this "constrained optimization."
- The topic's outline is:

IV.1 Substituting the constraint

IV.2 The Lagrange multiplier method for two variables

  (1) Necessary conditions
  (2) Sufficient conditions

IV.3 Interpretation of the Lagrange multiplier

IV.4 The general problem

IV.5 Non-linear programming

- We previously considered the problem of unconstrained optimization.
- For example, a two-product firm whose problem is

$$\max \Pi(Q_1, Q_2),$$

  where we choose outputs Q_1 and Q_2 to maximize profits. Here there are no constraints!
- Is this reasonable? Can the firm simply produce as much Q_1 and Q_2 as it wishes given their prices?
- No. For example, the firm may have limited capacity to store its products before delivery to customers, or there may be external constraints limiting production (such as noise reduction regulation or runway capacity constraints in the airline industry).
- Suppose therefore that capacity is limited such that the firm faces the following constraint: $Q_1 + Q_2 = 950$. $\Rightarrow$ Now Q_1 and Q_2 can only be chosen freely to the extent that the constraint is not violated.

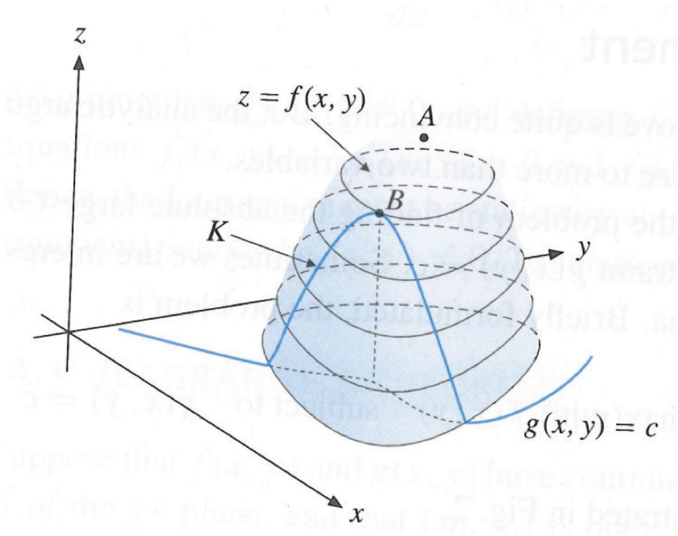
The problem is now one of constrained optimization:

$$\max_{Q_1, Q_2} \Pi(Q_1, Q_2) \qquad \text{subject to} \quad Q_1 + Q_2 = 950.$$

- The constraint means that Q_1 and Q_2 are not independent and they cannot be chosen entirely freely from the domain.
- The existence of a constraint always means that optimum profits are lower (or at least not higher) than they would be without the constraint.

A graphical illustration

- Consider the generic problem of $\max f(x, y)$ subject to $g(x, y) = c$.



IV.1 Substituting the constraint

- How do we deal with this constraint? We will now consider the easiest way to answer this question.
- Consider the constrained maximization

$$\max_{x,y} z = xy \quad \text{with} \quad x, y \in \mathbb{R} \qquad \text{subject to} \quad x + y = 6.$$

- Note that the function $z = xy$ is increasing everywhere.
- Thus the constraint will help us pin down the solution.

- Rewrite the constraint as
$$y = 6 - x.$$
- Then substitute it into the objective function
$$z = xy \quad \Rightarrow z = x(6-x) = 6x - x^2.$$
- Thus our problem reduces to
$$\max_x \left(6x - x^2\right).$$
- And this problem is straightforward to solve, as discussed in previous lectures. To find the maximum set the first derivative to zero: $\frac{dz}{dx} = 0$, i.e.
$$\frac{dz}{dx} = 6 - 2x = 0 \quad \text{so that} \quad x = 3.$$
Hence $y = 6 - x = 3$.
- It is helpful to get used to the following notation with asterisks for the solution:
$$x^* = 3, \qquad y^* = 3.$$

- This method of substituting the constraint is useful and you will be using it in other modules.
- But it is not a general method. We may have a more complicated constraint that cannot be solved for one variable in terms of the other in this way.
  - For instance, consider the constraint $x^3y^2 + xy^5 = 6$. You cannot simply solve for y as a function of x, or vice versa.
- We will now study the general method: the so-called *Lagrange multiplier* method.

IV.2 The Lagrange multiplier method for two variables

- The general set-up of a constrained optimization problem is $\max f(x, y)$ subject to $g(x, y) = c$.

Definition

The Lagrangian of the constrained optimization problem corresponds to

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - c],$$

where λ is an unknown **constant** called the Lagrange multiplier.

- The method is named after Joseph Louis Lagrange (1736-1813), mathematician and astronomer.
- The $\mathcal{L}$ is an auxiliary function that helps us to put the objective function and the constraint together into one equation.
- We are going to discuss now that previously studied concepts (necessary and sufficient conditions) apply in this setting, too.
- The only difference is that instead of f we will now have $\mathcal{L}$.

(1) Necessary conditions

- We find the necessary conditions just as before: by identifying the stationary points of the Lagrangian through the FOCs.
- The recipe is as follows:

1. Write down the problem as $\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - c]$.
2. Take the first-order partial derivatives of $\mathcal{L}$ and set them equal to zero. These first-order derivatives together with the constraint define three equations (first-order conditions):

$$\mathcal{L}_1' \equiv \frac{\partial \mathcal{L}}{\partial x} = f_1'(x, y) - \lambda g_1'(x, y) = 0,$$
$$\mathcal{L}_2' \equiv \frac{\partial \mathcal{L}}{\partial y} = f_2'(x, y) - \lambda g_2'(x, y) = 0,$$
$$\mathcal{L}_3' \equiv \frac{\partial \mathcal{L}}{\partial \lambda} = -[g(x, y) - c] = 0 \text{ or simply } g(x, y) = c.$$

3. These equations can be solved to find x, y and λ.

- Let's redo the previous example with the $\mathcal{L}$ method to check that we actually get the same answer.

Problem:

$$\begin{aligned} \max \quad & xy \\ \text{subject to} \quad & x + y = 6. \end{aligned}$$

The Lagrangian function is

$$\mathcal{L} = xy - \lambda[x + y - 6].$$

Hence the FOCs are

$$\mathcal{L}_x = y - \lambda = 0,$$

$$\mathcal{L}_y = x - \lambda = 0,$$

$$\mathcal{L}_\lambda = -[x + y - 6] = 0 \text{ or simply } x + y = 6.$$

Solve the first two FOCs for λ to obtain $x = y$. From the constraint we then get $x + y = x + x = 6 \Leftrightarrow x^* = 3$, and also $y + y = 6 \Leftrightarrow y^* = 3$. Finally $\lambda^* = x^* = y^* = 3$.

- The Lagrangian approach yields the same solution.

(2) Sufficient conditions

- Just as with unconstrained optimization, the stationary points of the Lagrangian are only (interior) *candidates* for extreme points.
- We need to check that these stationary points are in fact extreme points.
- How? Just as before: by considering the sufficient conditions over $\mathcal{L}$.

Theorem

Suppose (x_0, y_0) is an interior stationary point for the Lagrangian $\mathcal{L}(x, y)$.

(i) *If $\mathcal{L}(x, y, \lambda)$ is concave, then (x_0, y_0) is a maximum.*

(ii) *If $\mathcal{L}(x, y, \lambda)$ is convex, then (x_0, y_0) is a minimum.*

- The sufficient conditions are described in more detail in Chapter 14.5.

- One final detail remains. What happens if our first-order conditions give us several stationary points?
- The answer is simple. We just evaluate the objective function at each of these points and check which one gives us the maximum/minimum value.

Another example

- Consider the following production function with inputs $x > 0$ and $y > 0$ and its associated budget constraint for the input costs:

$$f(x,y) = x^{\frac{2}{3}} y^{\frac{1}{3}}$$

$$\text{subject to } 100x + 100y = 400000.$$

Write down the Lagrangian as

$$\mathcal{L} = x^{\frac{2}{3}} y^{\frac{1}{3}} - \lambda[100x + 100y - 400000].$$

- The FOCs are

$$\begin{aligned}
\mathcal{L}_x &= \frac{2}{3} x^{-\frac{1}{3}} y^{\frac{1}{3}} - 100\lambda = 0, \\
\mathcal{L}_y &= \frac{1}{3} x^{\frac{2}{3}} y^{-\frac{2}{3}} - 100\lambda = 0, \\
\mathcal{L}_\lambda &= -[100x + 100y - 400000] = 0 \iff 100x + 100y = 400000.
\end{aligned}$$

- From the first FOC solve for λ as

$$\lambda = \frac{\frac{2}{3}x^{-\frac{1}{3}}y^{\frac{1}{3}}}{100}$$

  and from the second FOC

$$\lambda = \frac{\frac{1}{3}x^{\frac{2}{3}}y^{-\frac{2}{3}}}{100}.$$

  Set the last two equations equal to each other to obtain

$$y = \frac{1}{2}x.$$

- Substitute this solution back into the constraint to obtain

$$x = 2666.\overline{6} \text{ and } y = 1333.\overline{3}.$$

  Maximum production is therefore roughly 2117 units.

- As to the sufficient conditions, note that $\mathcal{L}$ is continuously differentiable. We obtain

$$\begin{aligned} \mathcal{L}_{xx} &= -\frac{2}{9}x^{-\frac{4}{3}}y^{\frac{1}{3}} < 0, \mathcal{L}_{yy} = -\frac{2}{9}x^{\frac{2}{3}}y^{-\frac{5}{3}} < 0, \\ \mathcal{L}_{xy} &= \mathcal{L}_{yx} = \frac{2}{9}x^{-\frac{1}{3}}y^{-\frac{2}{3}} > 0 \end{aligned}$$

such that the determinant of the Hessian matrix is

$$\mathcal{L}_{xx}\mathcal{L}_{yy} - \mathcal{L}_{xy}^2 = \frac{4}{81}x^{-\frac{2}{3}}y^{-\frac{4}{3}} - \frac{4}{81}x^{-\frac{2}{3}}y^{-\frac{4}{3}} = 0.$$

Hence, $\mathcal{L}$ is concave and we have found a (constrained) maximum.

IV.3 Interpretation of the Lagrange multiplier λ

- Suppose that x^* and y^* solve $\max f(x, y)$ subject to $g(x, y) = c$.
- Assume that $x^*(c)$ and $y^*(c)$. Thus the function can be rewritten as

$$f^*(c) = f^*(x^*(c), y^*(c)).$$

Theorem

If $f^*(c)$ is differentiable, then

$$df^*(c) = \lambda dc \Leftrightarrow \frac{df^*(c)}{dc} = \lambda.$$

- λ represents the rate at which the value of the function changes when the constraint constant c changes. Sometimes λ is therefore called the *shadow price* or *marginal value* of the resource represented by the constraint.

- For example, suppose f is a profit function to be maximized and c is a resource constraint (say, the number of take-off/landing slot pairs a particular airline has at London's congested Heathrow airport). Then λ indicates the approximate increase in profits if the airline gets one additional slot pair.
- As another example, suppose f is the utility function of a child and c is the maximum number of ice creams a mother allows her child to eat per week. Then λ indicates the approximate increase in utility if the child is allowed to eat one additional ice cream per week.
- In our previous example with Q_1 and Q_2, λ indicates the increase in profits if the capacity constraint of 950 is relaxed by one unit.

IV.4 The general problem

- A general problem involves a multivariate function with n variables $f(x_1, \ldots, x_n)$ and m constraints $g_1(x_1, \ldots, x_n) = c_1$, $g_2(x_1, \ldots, x_n) = c_2$, ..., $g_m(x_1, \ldots, x_n) = c_m$.
- Formally, the problem is

$$\max_{x_1,\ldots,x_n} f(x_1, \ldots, x_n) \text{ s.t. } \begin{cases} g_1(x_1, \ldots, x_n) = c_1, \\ \vdots \\ g_m(x_1, \ldots, x_n) = c_m, \end{cases}$$

where s.t. means "subject to."

- We will have one Lagrange multiplier for each constraint. Thus the $\mathcal{L}$ function becomes

$$\mathcal{L}(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m) = f(x_1, \ldots, x_n) - \sum_{j=1}^{m} \lambda_j [g_j(x_1, \ldots, x_n) - c_j].$$

- Since there are n variables, how many FOCs are there? Answer: n (if you also count the partial derivatives w.r.t. λ_j, then there are $n+m$).
- Thus we have $n+m$ unknowns $(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m)$ and $n+m$ equations (the FOCs plus the constraints).
- The rest proceeds exactly as before.
- By the way, the $\sum$ symbol is shorthand for summation.
  - For example, $\sum_{h=1}^{H}\left(5x_h^2\right) = 5x_1^2 + 5x_2^2 + ... + 5x_H^2$.
  - For $H = 3$ you would get $\sum_{h=1}^{3}\left(5x_h^2\right) = 5x_1^2 + 5x_2^2 + 5x_3^2$.
  - Another example: $\sum_{h=1}^{4}(h) = 1 + 2 + 3 + 4 = 10$.
  - h is called the summation index.

IV.5 Non-linear programming

- Sometimes it may be the case that we are interested in **inequality constraints**: $g(x, y) \leq c$ (rather than $g(x, y) = c$.)
- In this case the problem is $\max f(x, y)$ subject to $g(x, y) \leq c$. This is a case of *non-linear programming*.
- Intuitively, we now look for solutions on a broader set.
- Moreover, at the optimum (x^*, y^*) it can be the case that $g(x^*, y^*) < c$ and thus we say that the constraint is slack or non-binding. Meanwhile if $g(x^*, y^*) = c$, we say that the constraint is binding.
- Solving these problems is more complex as they are typically solved on the basis of the so-called Kuhn-Tucker conditions.
- We will not cover the Kuhn-Tucker conditions as they are not as common in undergraduate economics.

- One common type of inequality constraint in economic applications is the *non-negativity constraint*:

$$\max f(x, y) \text{ s.t. } g(x, y) = c, \quad x \geq 0, \quad y \geq 0.$$

- Before trying anything complicated, the best approach is simply to solve the problem as $\max f(x, y)$ s.t. $g(x, y) = c$.
- If it turns out that $x^* \geq 0$ and $y^* \geq 0$, you are done (the constraints are either slack or just about binding).
- If it turns out that $x^* < 0$ or $y^* < 0$, the constraints are violated and thus you must look for maxima or minima on the bounds (these would be corner solutions).

Example

- Consider

$$\max_{x,y} xy + x + 2y \text{ s.t. } 2x + y = m, \quad x \geq 0, \quad y \geq 0, \quad m > 0.$$

- Setting up the Lagrangian yields

$$\mathcal{L} = xy + x + 2y - \lambda[2x + y - m].$$

- The FOCs are

$$\mathcal{L}_1' = y + 1 - \lambda 2 = 0, \quad \mathcal{L}_2' = x + 2 - \lambda = 0.$$

  Adding the constraint (obtained through $\mathcal{L}_\lambda'$) gives us three equations.

- From the FOCs we get $\lambda = (y+1)/2 = x + 2 \iff y = 2x + 3$. Substituting into the constraint yields $2x + (2x+3) = m$ so that $x^* = (m-3)/4$ and $y^* = (m+3)/2$.

- Thus
$$x^* = (m-3)/4, \qquad y^* = (m+3)/2.$$
- If $m > 3$, we would be done because the constraints $x \geq 0$ and $y \geq 0$ would be slack.
- But if $m < 3$, then $x^* < 0$ and we would be in trouble because the non-negativity constraint for x would be violated.
- So what would happen if $m < 3$ but either x, y or both are equal to 0? These are three contingencies.
- To find out we have to proceed in a case-by-case manner and check each of these three contingencies:

(1) Clearly, $(x, y) = (0, 0)$ cannot be an optimum because it would violate the constraint $2x + y = m > 0$.

(2) It could be the case that $y = 0$ and $x > 0$. In that case for the constraint $2x + y = m$ to hold we must have $x = m/2$. The value of the function would be $f = xy + x + 2y = x = m/2$.

(3) The other possible case is $x = 0$ and $y > 0$. In that case for the constraint $2x + y = m$ to hold we must have $y = m$. The value of the function would be $f = xy + x + 2y = 2y = 2m > m/2$.

- Therefore the solution has to be $(x^*, y^*) = (0, m)$ if $m < 3$. But $(x^*, y^*) = ((m-3)/4, (m+3)/2)$ has to be the solution if $m > 3$.
- We say that in the optimum, the constraint $x \geq 0$ is binding and the constraint $y \geq 0$ is slack if $m < 3$. But both constraints are slack if $m > 3$.
- Note that $(x^*, y^*) = (0, m)$ is the solution if $m = 3$.

Topic V: Tools for comparative statics

EC123 Mathematical Techniques B

Dr Dennis Novy

University of Warwick

Autumn Term 2015

Topic V: Tools for comparative statics

Topic V's big picture:

- We already know how to maximize/minimize univariate and multivariate functions.
- The next natural step is to explore the **properties** of our solutions. For example, we want to study how solutions change when some variables or some parameters change.
- By the way, we already know a little bit about this as we have already studied elasticities.
- The topic's outline is:

V.1 Implicit differentiation
V.2 Total derivatives and the chain rule for many variables
V.3 The slope of level curves and the marginal rate of substitution
V.4 Homogeneous functions and homothetic functions
V.5 Differentials
V.6 The envelope theorem
V.7 Approximations

V.1 Implicit differentiation
(Ch. 7.1, 7.2)

- Up to now we have only differentiated **explicit** functions through differentiation and partial differentiation.
- Those derivatives gave us an idea of how the value of the function changed when we changed one of its arguments.
- Now we will study a different type of differentiation: differentiation of functions that are defined **implicitly** by an equation.
- Notice the difference:
    - So far we have had $z = f(x, y)$ and considered, for instance, $\frac{\partial z}{\partial x}$.
    - Now we will have an **equation** such as $c = f(x, y)$ where $c \in \mathbb{R}$. We will be interested in finding out how y changes when x changes such that the equation still holds.
    - What we are saying then is that implicitly y is a function of x.

Why should this be interesting in the first place?

- The FOCs are equations that set $f_x = 0$ and $f_y = 0$.
- The FOCs allow us to identify stationary points (x, y), which under sufficient conditions will be extreme points.
- Usually these extreme points are in fact functions. In simple cases they are numbers but in general they can depend on parameters and the values of the other variables.
- For example, in the consumer problem the demand for good x also depends on how much the consumer consumes of good y, apart from exogenous variables such as income and prices.
- As an example, suppose we want to keep the utility of the consumer constant (think indifference curves). $\Rightarrow$ This defines a certain equation that tells us how much the demand for good y must change if the demand for x changes.
- To answer that question we can use implicit differentiation.
- Also, as we will study in V.3, implicit differentiation will help us with the slopes of the level curves we discussed in Topic II.

- Take a function $z = f(x, y)$.
- For a given $z = c$ with $c \in \mathbb{R}$, the function defines an equation $c = f(x, y)$.
- Now suppose that x changes. Then, in order to maintain the equality $c = f(x, y)$, y must also change.
- Thus there exists a function g such that $y = g(x, c)$.
- For example, take the function $f(x, y) = xy$ over $x > 0, y > 0$ and pick $z = 5$:

$$xy = 5.$$

- Here it is obvious that if we change x, then y must also change.
- Define a function g that captures such change **given** $z = 5$:

$$xg(x) = 5, \quad x > 0,$$

  where g is the **implicit function**.
- The question that interests us is: what is $\frac{dg(x)}{dx} \equiv \frac{dy}{dx}$?

- How do we answer that question?
  1) First rewrite the initial equation in terms of the implicit function g:
  $$xg(x) = 5.$$
  2) Now take the derivative of **both sides** without being explicit about the value of the derivative:
  $$1 \cdot g(x) + x\frac{dg(x)}{dx} = 0,$$
  where we have used the *product rule*. If you don't remember the product rule, look it up pronto.
  3) Now solve for $\frac{dg(x)}{dx}$:
  $$\frac{dg(x)}{dx} = -\frac{g(x)}{x}.$$
- Rewrite this last equation using the simpler notation $\frac{dg(x)}{dx} \equiv y'$:
$$y' = -\frac{y}{x}.$$
- The above equation tells us for any point (x, y) how y will change if x changes marginally.

The second derivative of implicit functions

- Just as we calculated the first derivative, we can also work out the second derivative, $\frac{d^2y}{dx^2} \equiv \frac{d^2g(x)}{dx^2}$.
- We follow the same method as before:
  - Take the derivative of both sides of $g(x) + x\frac{dg(x)}{dx} = 0$:

$$\frac{dg(x)}{dx} + 1 \cdot \frac{dg(x)}{dx} + x \cdot \frac{d^2g(x)}{dx^2} = 0,$$

    where we have again used the product rule.
  - Now solve for $\frac{d^2g(x)}{dx^2}$:

$$\frac{d^2g(x)}{dx^2} = -2\frac{\frac{dg(x)}{dx}}{x}.$$

- Changing the notation to $\frac{d^2g(x)}{dx^2} \equiv y''$ yields

$$y'' = -2\frac{y'}{x} = 2\frac{y}{x^2}.$$

- The last example was fairly straightforward because we knew $g(x)$ to be $g(x) = 5/x$ simply because originally we had $xg(x) = 5$. In other words, we were dealing with an explicit function.
- Substitute $y = 5/x$ to arrive at

$$y' = -\frac{5/x}{x} = -\frac{5}{x^2}, \quad y'' = \frac{10}{x^3}.$$

- Moreover, in this example we could have proceeded directly by taking straight first-order and second-order derivatives over $c = f(x, y)$ in the following way:

$$xy = 5 \Leftrightarrow y = \frac{5}{x} \quad \Rightarrow y' = -\frac{5}{x^2} \quad \Rightarrow y'' = \frac{10}{x^3}.$$

- Our solutions coincide.

- But life is not always as simple and in many cases we don't know $g(x)$.
- For example, consider $f(x, y) = y^3 + 3x^2y = c$ for any $c \in \mathbb{R}$. In this case we cannot work out an explicit solution for $y = g(x)$.
- Thus to find out y' we must use **implicit differentiation**. So we follow the steps above.
  - Take the derivative with respect to x on both sides of $y^3 + 3x^2y = c$ imposing $y = g(x)$:

$$3y^2y' + (6xy + 3x^2y') = 0,$$

    where we have used the *chain rule*. If you don't remember the chain rule, you will be lost here. You need to revise it immediately.
  - Now solve for y':

$$y' = -\frac{2xy}{(y^2 + x^2)}.$$

- For any (x, y) we can now find y'.

Example: A simple macroeconomic model

Consider the following income-expenditure model with two equations.

$$Y = C + I, \tag{1}$$

$$C = c(Y), \tag{2}$$

where Y denotes national income = total expenditure, C denotes consumption expenditure and I denotes investment expenditure.

- Equation (1) is an accounting identity saying that in the economy as a whole, income equals total expenditure.
- Equation (2) is a description of the behaviour of households (consumers) and how they spend. They spend more, the more income they have so that we can write $\frac{dC}{dY} > 0$, or equivalently, $c' > 0$.

This is a very simple model of spending in the economy. For example, it assumes a closed economy with no foreign trade (imports or exports), no government spending and no taxes. The only spending is done by consumers (for consumption) and by firms (for investment).

Question: How does investment spending affect the level of national income? Answer: Find the derivative $\frac{dY}{dI}$.

But the simple derivative of $\frac{dY}{dI}$ in equation (1) does not capture the fact that if I changes Y, then due to equation (2) the change in Y will also affect C, which in turn affects Y. In other words, the simple derivative does not capture the circular nature of the economy described by this model. (Technically speaking, this is a system of simultaneous equations in two unknowns, Y and C.)

First solve for Y in terms of I by eliminating C from equation (1). This can be achieved by substituting (2) into (1), which yields the implicit function

$$Y = c(Y) + I. \tag{3}$$

We can interpret (3) as saying that Y is a function of I. We do not need to know the algebraic form of the consumption function $c(Y)$. We just know that its derivative c' is positive.

We can now differentiate (3) by implicit differentiation:

$$\frac{dY}{dI} = \frac{dC}{dI} + 1.$$

The term $\frac{dC}{dI}$ can be found by the chain rule:
Let z be a function of $y, z(y)$, and y a function of $x, y(x)$. Then we can find the derivative of z with respect to x as

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}.$$

Therefore

$$\frac{dY}{dI} = \frac{dC}{dI} + 1 = \frac{dC}{dY}\frac{dY}{dI} + 1 = c'\frac{dY}{dI} + 1$$

$$\iff [1 - c']\frac{dY}{dI} = 1$$

$$\iff \frac{dY}{dI} = \frac{1}{1 - c'}.$$

The term c' is the so-called *marginal propensity to consume*. It tells us what fraction of an additional pound of income households in this model spend on consumption (say, 80 percent) as opposed to save.
We have derived the simple so-called *Keynesian multiplier* $1/(1 - c')$. As typically $0 < c' < 1$, it is called a multiplier because then $1/(1 - c') > 1$.
You will learn about this in macroeconomics.

V.2 Total derivatives and the chain rule for many variables
(Ch. 12.1, 12.2)

- Suppose we have a bivariate function $z = f(x, y)$. And suppose further that each of its variables is in turn a function of another variable t:

$$x = g(t), \qquad y = h(t).$$

- Then we can imagine z as a function of t alone:

$$z = f(x, y) = f(g(t), h(t)).$$

- The derivative of $f(x, y)$ with respect to t is called the total derivative. We denote it by $\frac{dz}{dt}$.
- To find the total derivative we must use the chain rule for several variables.

Theorem

$$\frac{dz}{dt} = \frac{\partial f(x,y)}{\partial x}\frac{dx}{dt} + \frac{\partial f(x,y)}{\partial y}\frac{dy}{dt}.$$

This is the chain rule for the case of two variables.

- Thus the total derivative considers the contributions of both x and y to the change in z due to a change in t.
- We can generalize the chain rule for functions of n variables that at the same time depend on m additional variables.

Theorem

Suppose that $z = f(x_1, ..., x_n)$ *with* $x_1 = f_1(t_1, ..., t_m), ..., x_n = f_n(t_1, ..., t_m)$. *Then*

$$\frac{dz}{dt_j} = \frac{\partial f(x_1, ..., x_n)}{\partial x_1}\frac{\partial x_1}{\partial t_j} + ... + \frac{\partial f(x_1, ..., x_n)}{\partial x_n}\frac{\partial x_n}{\partial t_j} \text{ for any } j = 1, ..., m.$$

This is the chain rule for the case of n variables.

- Also:
  - Recall that the chain rule for one variable is as follows. For a univariate function $z = f(x)$ where $x = g(t)$, we have $\frac{dz}{dt} = \frac{df(x)}{dx}\frac{dx}{dt}$.
  - Curious as to why the chain rule works? See the argument at the end of chapter 12.1 in the textbook.
- In some cases, the variable t is in fact one of the other variables, say, x. Then we have

$$z = f(x, y) \quad \text{and} \quad y = h(x).$$

- In such cases we can write the function z as

$$z = f(x, h(x)).$$

- Then the total derivative is

$$\frac{dz}{dx} = \frac{\partial z}{\partial x} + \frac{\partial z}{\partial y}\frac{dy}{dx},$$

  or with slightly different notation, $\frac{dz}{dx} = f'_x + f'_y h'$. Note the use of the operators ∂ and d.

Example 1

Consider the function

$$z = \frac{x}{y} \text{ and } x = 1 + t, \quad y = 1 - t.$$

The total derivative is

$$\begin{aligned} \frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \\ &= \left(\frac{1}{y}\right)(1) + \left(-\frac{x}{y^2}\right)(-1) \\ &= \frac{1}{y} + \frac{x}{y^2} = \frac{y+x}{y^2} = \frac{2}{(1-t)^2}. \end{aligned}$$

[Of course, we could have done this just as easily by using the *quotient rule* after substituting for x and y. Try this to check that it gives you the same answer.]

Example 2: Marginal revenue of a firm

Total revenue = price × quantity sold:

$$R = pq.$$

For a firm operating under **imperfect competition** (as opposed to perfect competition), the price depends on the quantity sold through the demand function. We can express this dependency as

$$p = p(q).$$

So we can rewrite total revenue as a function of quantity:

$$R(q) = p(q)q.$$

Profit maximization requires that we choose output such that marginal revenue MR equals marginal cost (see your microeconomics lectures). Thus we need to find an expression for marginal revenue, which is the total derivative of $R(q)$:

$$MR = \frac{dR}{dq} = \frac{\partial R}{\partial q} + \frac{\partial R}{\partial p}\frac{dp}{dq}.$$

Since $R = pq$, it follows $\frac{\partial R}{\partial q} = p$ and $\frac{\partial R}{\partial p} = q$. Therefore

$$MR = \frac{dR}{dq} = p + q\frac{dp}{dq}.$$

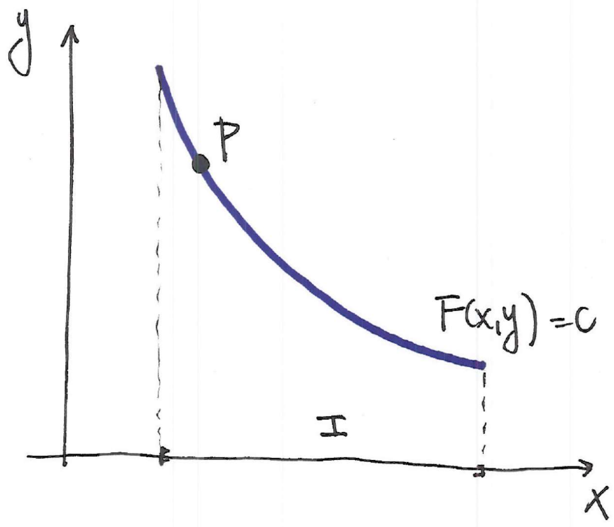
The term $\frac{dp}{dq}$ measures the effect of an increase in output on the price through the demand function.

In **perfect competition** the firm is a **price taker** so that its price is constant and therefore $\frac{dp}{dq} = 0$ and $MR = p$. But in imperfect competition (including the textbook monopoly case) we typically have $\frac{dp}{dq} < 0$.

V.3 The slope of level curves and the marginal rate of substitution
(Ch. 12.3, 12.4, 12.5)

- Remember that a level curve is the set of combinations of x and y that assign a value c to the function $f(x, y)$, that is
$$c = f(x, y).$$
(In consumer theory these are the famous *indifference curves*.)
- What is the slope of a level curve?
  - Remember that in the univariate case the slope of a function is the first derivative of the function with respect to its argument.
  - Level curves are defined on an x, y plane. Thus we need a trick to turn this into a one-variable case. That trick is an implicit function.
- We know that for the function $f(x, y) = c$, y is an implicit function of x. Put differently, we know that there exists a function g that maps x to y for any given c, that is $y = g(x)$.
- Thus f can be rewritten as
$$f(x, g(x)) = c.$$

- What is the slope of the level curve? A graph will help us out:



- The slope is given by the derivative of the implicit function $y = g(x)$, that is $\frac{dy}{dx}$.

- Thus we can take the derivative of $f(x, g(x)) = c$ with respect to x.
- Using the chain rule we have

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0$$

$$\iff \frac{dy}{dx} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}}.$$

- We then have the following theorem:

Theorem

Let a level curve c be defined by $c = f(x, y)$. Then the slope of the level curve corresponds to

$$y' = \frac{dy}{dx} = -\frac{f_1'(x, y)}{f_2'(x, y)}.$$

- Note that we could have obtained a similar result by making x an implicit function of y and then taking the derivative with respect to y.

Example

- Find the slope of the level curve described by $xy = 5$.
- Answer:
  - Write $f(x, y) = xy$. Then $f_1' = y$ and $f_2' = x$.
  - Hence

$$y' = \frac{dy}{dx} = -\frac{f_1'(x, y)}{f_2'(x, y)} = -\frac{y}{x}.$$

Another example

- For the curve given by

$$x^3 + x^2 y - 2y^2 - 10y = 0$$

  find the general slope and its value at the point $(x, y) = (2, 1)$.
- Answer:
  - Let $f(x, y) = x^3 + x^2 y - 2y^2 - 10y$. Then $f_1' = 3x^2 + 2xy$ and $f_2' = x^2 - 4y - 10$.
  - Hence

$$y' = -\frac{3x^2 + 2xy}{x^2 - 4y - 10}.$$

  - Note that $(2, 1)$ is indeed a point on the level curve because $f(2, 1) = 8 + 4{*}1 - 2{*}1 - 10{*}1 = 0$.
  - For $(2, 1)$ we obtain $y' = 8/5$.

The marginal rate of substitution

- The slope of a level curve has a very neat interpretation. It tells us how much y has to change in response to a change in x so that the function f maintains its value c.
- In the context of consumer theory, the slope tells us how much more of the good y the household has to consume if less of good x has become available in order to keep the level of utility constant at value c.
- In consumer theory we call this:

Definition

The marginal rate of substitution (MRS) of y for x corresponds to $MRS_{yx} \equiv \frac{f_1'(x,y)}{f_2'(x,y)} = -\frac{dy}{dx}$.

- Note that in this definition there is no minus sign directly after the $\equiv$ symbol so that MRS_{yx} is typically a positive number.

V.4 Homogeneous functions and homothetic functions

(Ch. 12.6, 12.7)

- We now study two classes of functions: *homogeneous* and *homothetic* functions.
- We study them because they have several interesting properties that make comparative statics convenient.
- Furthermore, they often appear in producer theory (less frequently in consumer theory).
- Note that such functions are called homogen**eous**, not homogenous.

Definition of homogeneous functions

Definition

A function of n variables $(x_1, x_2, ..., x_n)$ is homogeneous of degree $k \in \mathbb{R}$ if for any $t > 0$

$$f(tx_1, tx_2, ..., tx_n) = t^k f(x_1, x_2, ..., x_n).$$

- That is, if each independent variable is multiplied by the same positive number t, the function changes by a factor of t^k.
- Intuitively, it is easiest to set $t = 2$. In producer theory this would mean doubling all inputs.

Some examples of homogeneous functions

1) Consider a function of two variables $f(x, y) = x + y$. To see that this is a homogeneous function we multiply each variable by t:

$$f(tx, ty) = tx + ty = t(x + y) = tf(x, y).$$

Thus the function is homogeneous of degree 1 with $k = 1$.

2) Consider the Cobb-Douglas function

$$Q = AK^{\alpha}L^{\beta},$$

where $A > 0$, $0 < \alpha < 1$, $0 < \beta < 1$ are parameters.

$$f(tK, tL) = A(tK)^{\alpha}(tL)^{\beta} = t^{\alpha+\beta}AK^{\alpha}L^{\beta} = t^{\alpha+\beta}f(K, L).$$

The function is homogeneous of degree $\alpha + \beta$. When Q is production and K is capital and L is labour, we say that the function has

- $\alpha + \beta = 1$ **constant returns to scale** (CRS),
- $\alpha + \beta > 1$ **increasing returns to scale** (IRS),
- $\alpha + \beta < 1$ **decreasing returns to scale** (DRS).

3) Consider the function $f(x,y) = 1 + xy + y^2$. This function is not homogeneous. This is because

$$f(tx,ty) = 1 + t^2xy + t^2y^2 \neq t^k f(x,y) \quad \text{for any } k.$$

4) Consider the function $f(x,y) = \frac{y}{x}$. This function is homogeneous of degree 0 since

$$f(tx,ty) = \frac{ty}{tx} = \frac{y}{x}.$$

5) Consider a function $f(x,y) = x^2 + xy + y^2$. As before we multiply all variables by t:

$$\begin{aligned} f(tx,ty) &= (tx)^2 + (tx)(ty) + (ty)^2 \\ &= t^2x^2 + t^2xy + t^2y^2 \\ &= t^2(x^2 + xy + y^2) \\ &= t^2 f(x,y). \end{aligned}$$

The function is homogeneous of degree $k = 2$.

Properties of homogeneous functions

- We will review six interesting properties of homogeneous functions.
- Let $f(x, y)$ be homogeneous of degree k. Then the following statements are true:

1. f'_x and f'_y are both homogeneous of degree $k - 1$.
2. $f(x, y) = x^k f(1, \frac{y}{x}) = y^k f(\frac{x}{y}, 1) \quad \text{for all} \quad x > 0, y > 0.$
3. $x^2 f''_{xx} + 2xy f''_{xy} + y^2 f''_{yy} = k(k-1)f.$

The proofs of these properties are straightforward. See chapter 12.6 in the textbook (but you don't have to know these proofs).

Euler's theorem

- The fourth (and perhaps the most important of all) property of homogeneous functions is Euler's theorem.

Theorem

If $f(x, y)$ is homogeneous of degree k, then

$$xf'_x(x, y) + yf'_y(x, y) = kf(x, y).$$

- In other words, Euler's theorem links the partial derivatives to the level of the function.
- The general n-variable version of Euler's theorem is:

Theorem

If $f(x_1, x_2, ..., x_n)$ is homogeneous of degree k, then

$$\sum_{i=1}^{n} x_i f'_i = kf.$$

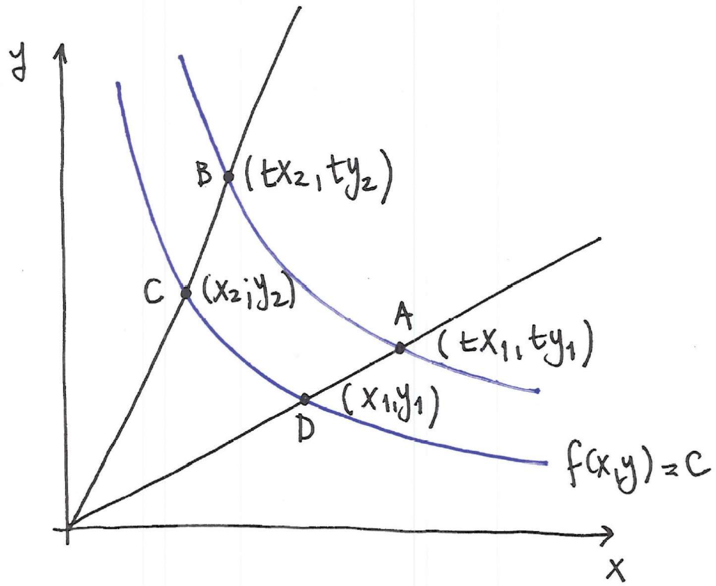
- An important application of Euler's theorem is in production theory.
  - For a Cobb-Douglas production function with two factors of production, $Q(K, L)$, Euler's theorem implies

$$KQ'_K + LQ'_L = (\alpha + \beta)Q.$$

  - That is, capital $\times$ the marginal product of capital $+$ labour $\times$ the marginal product of labour $=$ the returns-to-scale parameters $\times$ output.

Level curves of homogeneous functions

- The fifth property of homogeneous functions concerns the slope of their level curves. It is quite useful.
- Basically, when the function is homogeneous, if you know one level curve, you know them all.
- To see this define a level curve c for the function $f(x, y)$, $c = f(x, y)$.
  - Suppose that the point (x_0, y_0) belongs to the level curve. Thus it must be the case that
$$f(x_0, y_0) = c.$$
  - Since the function is homogeneous of degree k, we know that if we multiply each variable by t we get
$$f(tx_0, ty_0) = t^k f(x_0, y_0) = t^k c.$$
  - Hence, if we want to know the points that belong to a level curve $t^k c$, then all we have to do is multiply the points of the level curve c by t.



y
B
(tx_2, ty_2)
C
(x_2, y_2)
A
(tx_1, ty_1)
(x_1, y_1)
D
$f(x,y) = c$
x

Partial elasticities and Euler's theorem

- The sixth and last interesting property of homogeneous functions we examine relates the partial elasticities to the degree of homogeneity.
- From Euler's theorem we know that if a function $f(x_1, x_2, ..., x_n)$ is homogeneous of degree k, then

$$x_1 f_1' + x_2 f_2' + ... + x_n f_n' = kf.$$

- Dividing each term by f yields

$$\frac{x_1}{f} f_1' + \frac{x_2}{f} f_2' + ... + \frac{x_n}{f} f_n' = k.$$

  Therefore

$$El_1 f + El_2 f + ... + El_n f = k,$$

  where $El_1 f$ stands for the partial elasticity of f with respect to x_1 etc.
- Thus the sum of the partial elasticities is equal to the degree of homogeneity k.

Example: Demand function

Let a demand function for a certain good be expressed as

$$q = p^{-0.3} y^{0.7},$$

where q is the quantity demanded, p is the price of the good and y is the income of households.

The partial elasticities follow as

$$El_p q = -0.3,$$

$$El_y q = 0.7.$$

The degree of homogeneity is equal to the sum of its partial elasticities: $-0.3 + 0.7 = 0.4$.

Homothetic functions

- We now briefly review a broader class of functions: homothetic functions.
- For that purpose we first define a class of sets called *cones*.

Definition

A set $K \subseteq \mathbb{R}^n$ is called a cone if for any $\mathbf{x} \in K$ and $t > 0$, then $t\mathbf{x} \in K$.

Definition

Let f be a function of n variables $\mathbf{x} = (x_1, \ldots, x_n)$ defined over a cone K. Then f is called homothetic if for all $\mathbf{x}, \mathbf{y} \in K$

$$f(\mathbf{x}) = f(\mathbf{y}) \Rightarrow f(t\mathbf{x}) = f(t\mathbf{y})$$

for any $t > 0$.

- Think about a consumer with homothetic preferences. If she is indifferent between two bundles, then it must also be the case that she is indifferent between the two bundles scaled by t. For example, if this consumer is indifferent between 2 litres of soda and 3 litres of juice, she is also indifferent between 20 litres of soda and 30 litres of juice.
- One can prove that homogeneous functions of any degree k are homothetic, but the converse is not true.
- Example: Consider the function $f(x, y) = xy + 1$. This function is not homogeneous but nevertheless homothetic. Suppose $f(x_0, y_0) = f(x_1, y_1)$. It follows $x_0 y_0 = x_1 y_1$. Then for $t > 0$

$$\begin{aligned} f(tx_0, ty_0) &= (tx_0)(ty_0) + 1 \\ &= t^2 x_0 y_0 + 1 \\ &= t^2 x_1 y_1 + 1 \\ &= (tx_1)(ty_1) + 1 = f(tx_1, ty_1). \end{aligned}$$

Since both $f(x_0, y_0) = f(x_1, y_1)$ and $f(tx_0, ty_0) = f(tx_1, ty_1)$, the function is homothetic.

- Consider the logarithmic function

$$f(x,y) = a\ln(x) + b\ln(y),$$

  where a and b are parameters.
- It is not homogeneous because

$$\begin{aligned} f(tx,ty) &= a\ln(tx) + b\ln(ty) \\ &= a(\ln(t) + \ln(x)) + b(\ln(t) + \ln(y)) \\ &= (a+b)\ln(t) + a\ln(x) + b\ln(y) \\ &= (a+b)\ln(t) + f(x,y) \neq t^k f(x,y). \end{aligned}$$

- However, it is homothetic.
  - Suppose $f(x_0, y_0) = f(x_1, y_1)$.
  - Then use the result from above to show that for $t > 0$

$$\begin{aligned} f(tx_0, ty_0) &= (a+b)\ln(t) + f(x_0, y_0) \\ &= (a+b)\ln(t) + f(x_1, y_1) = f(tx_1, ty_1). \end{aligned}$$

V.5 Differentials
(Ch. 12.9)

- Derivatives (partial and total) describe the change of a function when one of its arguments changes.
- We now study how to measure the change of the function when **many** of its arguments change.
- Consider a function of two variables $z = f(x, y)$ where f has continuous partial derivatives everywhere.
- Consider small changes in the variables x and y denoted by dx and dy where $dx \in \mathbb{R}$ and $dy \in \mathbb{R}$.

Definition

The differential of $f(x, y)$ at (x, y) denoted by dz or df corresponds to

$$dz = f'_x(x, y)dx + f'_y(x, y)dy.$$

- It is easy to see that this definition makes sense.

Consider small changes in the *independent* variables Δx and Δy. Let the resulting change in the *dependent* variable be Δz. Then

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y)$$

$$= f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y) + f(x, y + \Delta y) - f(x, y)$$

$$= \frac{f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y)}{\Delta x}\Delta x + \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}\Delta y$$

For infinitesimal changes dx and dy (i.e. in the limit as $\Delta x \longrightarrow 0$ and $\Delta y \longrightarrow 0$), the two ratios are the partial derivatives.
Therefore

$$\Delta z \approx dz = f'_x dx + f'_y dy.$$

- Finally, we define differentials for the case of functions of n variables.

Definition

The differential of $f(x_1, \ldots, x_n)$ at $(x_1, \ldots, x_n)$ denoted by dz or df corresponds to

$$dz = f'_{x_1} dx_1 + \ldots + f'_{x_n} dx_n .$$

- This is the straightforward generalization of the two-variable case.

Example

Consider the function

$$z = x^2 y^3.$$

The first-order partial derivatives are

$$f'_x = 2xy^3 \quad \text{and} \quad f'_y = 3x^2 y^2.$$

The differential is then

$$dz = 2xy^3 dx + 3x^2 y^2 dy.$$

Rules for differentials

Let $f(x, y)$ and $g(x, y)$ be differentiable functions and let a and b be parameters. Denote the differential of any function as d. Then

1. $d(af + bg) = adf + bdg$,
2. $d(fg) = gdf + fdg$,
3. $d\left(\frac{f}{g}\right) = \frac{gdf - fdg}{g^2}$ for $g \neq 0$,
4. if $z = g(f(x, y)) \Rightarrow dz = g'(f(x, y))df$,

where the last statement is the chain rule for differentials.

Another example

Find an expression for dz in terms of dx and dy for the following function:

$$z = Ax^a + By^b.$$

Solution:

$$\begin{aligned} dz &= Ad(x^a) + Bd(y^b) \\ &= Aax^{a-1}dx + Bby^{b-1}dy. \end{aligned}$$

Yet another example

Find an expression for dz in terms of dx and dy for the following function:

$$z = \ln\left(x^2 + y\right).$$

Solution:

$$\begin{aligned} dz &= d \ln\left(x^2 + y\right) \\ &= \frac{d\left(x^2 + y\right)}{x^2 + y} \\ &= \frac{2x\,dx + dy}{x^2 + y}. \end{aligned}$$

V.6 The envelope theorem

(Ch. 13.7)

- Many economic problems involve parameters, that is, numbers that represent variables that are fixed from the point of view of the optimization process.
- For example, tax rates in consumer and producer theory are taken as given numbers.
- One interesting question to ask is how the functions representing the extreme points change when these parameters change.
- For example, how does the production of a certain product adjust when the government changes the taxes on inputs (for instance, an increase in fuel duty or an increase in the departure tax for aviation)?
- To answer this sort of question we use the (pretty famous) *envelope theorem*.

- Consider the problem of

$$\max_x f(x, r) \quad \text{where } r \text{ is a parameter.}$$

- Denote the value of x that maximizes the function by x^*. If we change r, then x^* might also change so that we write $x^*(r)$.
- Then we have the following definition.

Definition

The value function is the value of a function when evaluated at its optimum. That is,

$$f^*(r) \equiv f(x^*(r), r).$$

- How does the value function change when r changes? Using the chain rule we obtain

$$\frac{df^*(r)}{dr} = f_1'(x^*(r), r)\frac{dx^*(r)}{dr} + f_2'(x^*(r), r).$$

- The last equation says that r affects the value function $f(x^*(r), r)$ in two ways:
  1. indirectly through its influence on $x^*(r)$,
  2. directly through r.
- But $x^*(r)$ is an extreme point and therefore a stationary point and thus $f_1'(x^*(r), r) = 0$. This is the key idea.
- Therefore

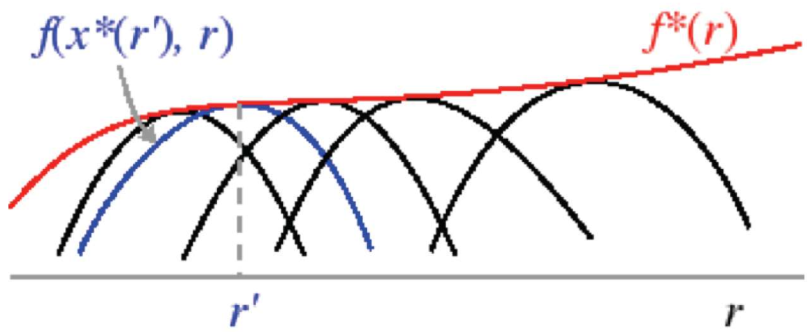
$$\frac{df^*(r)}{dr} = f_2'(x^*(r), r).$$

- In other words, only the direct effect matters!
- This is the **envelope theorem**.

- Why is it called the "envelope" theorem?
- Take a given point x_0. Then there is a curve K_{x_0} representing the function $y = f(x_0, r)$ in r, y space. How many functions like this are there? Many! In fact, as many as there are x's.
- At the same time, we have the value function $y = f^*(r)$, which also maps into r, y space.
- How do these K_x curves relate to the value function? Two key insights:

  1) By definition of the value function for any x, it must be that $f(x, r) \leq max_x f(x, r) = f^*(r)$. Thus, none of the K_x curves can lie above the value function.
  2) For each r there is at least one $x^*(r)$. Thus we know that K_x will touch the curve of $y = f^*(r)$ in the point $(x^*(r), f(x^*(r)) = (x^*(r), f^*(x^*(r), r))$ and will have the same tangent in that point.

A graphic illustration



Example

- Suppose a firm sells x units of a commodity and has revenue $R(x) = rx$, where r is the price that the firm takes as given. Suppose that the firm's cost function is $C(x) = x^2$.
- Profits are then given by

$$\pi(x, r) = R(x) - C(x) = rx - x^2.$$

- Find the optimal choice x^* as a function $x^*(r)$. Then verify $\frac{d\pi^*(r)}{dr} = \pi'_2(x^*(r), r)$.
  - The quadratic profit function has a maximum when $\pi'_1 = r - 2x = 0$. Thus $x^*(r) = r/2$.
  - Then maximum profits can be expressed as a function of r as $\pi^*(r) = rx^* - (x^*)^2 = r(r/2) - (r/2)^2 = r^2/4$.
  - It follows $d\pi^*/dr = r/2$.
  - Using the envelope theorem is more direct because since $\pi'_2(x, r) = x$, it follows $\frac{d\pi^*}{dr} = \pi'_2(x^*(r), r) = x^*(r) = r/2$.

Another example

- Recall this example from the topic on constrained optimization:

$$\max_{x,y} xy + x + 2y \text{ s.t. } 2x + y = m, \quad x \geq 0, \quad y \geq 0, \quad m > 0.$$

We found the solution

$$x^* = \frac{(m-3)}{4}, \qquad y^* = \frac{(m+3)}{2}.$$

For simplicity, assume $m > 3$ such that $x^* > 0, y^* > 0$.

- Then we can rewrite the objective function as a value function:

$$\begin{aligned} f^*(m) &= \frac{(m-3)}{4}\frac{(m+3)}{2} + \frac{(m-3)}{4} + 2\frac{(m+3)}{2} \\ &= \frac{m^2-9}{8} + \frac{(m-3)}{4} + 2\frac{(m+3)}{2}. \end{aligned}$$

- The direct effect follows as

$$\frac{df^*(m)}{dm} = \frac{m}{4} + \frac{1}{4} + 1,$$

and the indirect effect can be ignored since it equals zero.

The general version of the envelope theorem

The general version of the envelope theorem considers the case where there are n variables denoted in vector notation as $\mathbf{x}$ and k parameters denoted in vector notation as $\mathbf{r}$.

Theorem

If $f^*(\mathbf{r}) = \max_{\mathbf{x}} f(\mathbf{x}, \mathbf{r})$ *and if* $\mathbf{x}^*(\mathbf{r})$ *is the value of* $\mathbf{x}$ *that maximizes* $f(\mathbf{x}, \mathbf{r})$, *then*

$$\frac{df^*(\mathbf{r})}{dr_j} = \left[\frac{\partial f(\mathbf{x}, \mathbf{r})}{\partial r_j}\right]_{\mathbf{x}=\mathbf{x}^*(\mathbf{r})} \qquad \textit{for any } j = 1, ..., k,$$

where the subscript of the square brackets means "evaluated at $\mathbf{x}^*(\mathbf{r})$*."*

V.7 Approximations

(Ch. 7.4, 7.5, 7.6, 12.8)

- One last tool that can be used in comparative statics (and also optimization) is an approximation to a function.
- Sometimes the functions we are working with (for instance, objective functions or first-order conditions) are very complicated (highly non-linear).
- One way to try to simplify matters is to use a *linear approximation* to the function.
- We will now discuss how to construct such linear approximations (and more involved approximations) and when it is a good idea to use them.

Linear approximations to univariate functions

Definition

Assume the function $f(x)$ is differentiable at x_0. A linear approximation to $f(x)$ at $x = x_0$ corresponds to the tangent line at $x = x_0$. That is

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0),$$

where you should recall that the tangent y of $f(x)$ at x_0 corresponds to

$$y(x) = f(x_0) + f'(x_0)(x - x_0).$$

How did we derive the tangent?

- Remember that a *linear function* in general can be written as $y(x) = \alpha + \beta x$.
- The tangent at a point x_0 is a linear function defined by $y(x) = \alpha + f'(x_0)x$. (Recall that the derivative $f'(x_0)$ is the slope of the tangent.)
- We need to find α, which is the intercept of this linear function.
- We know that the point $(x_0, f(x_0))$ belongs to the tangent. Thus it must be the case that

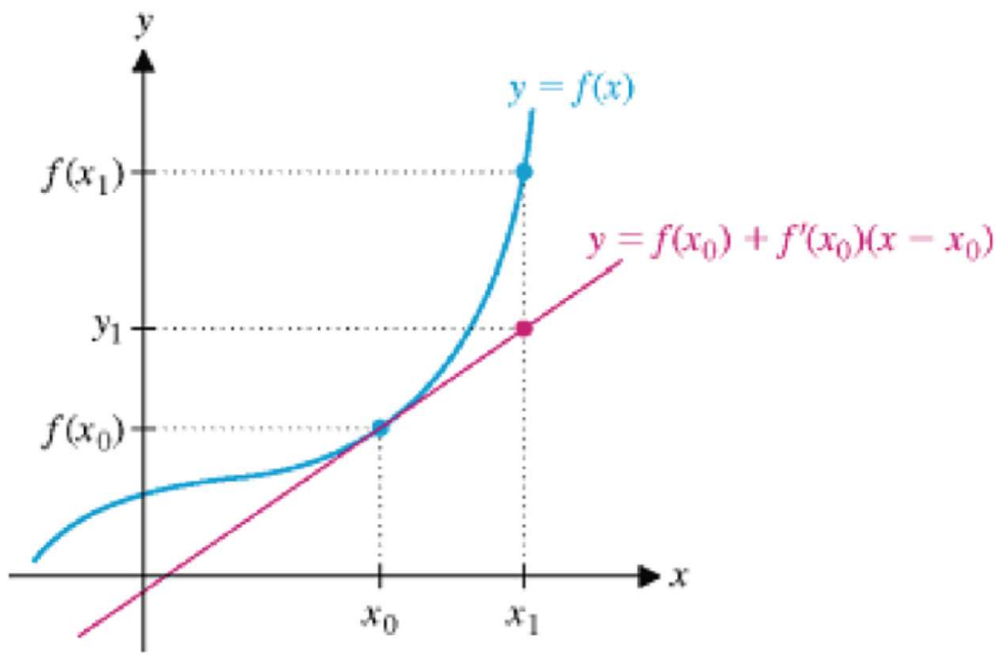
$$f(x_0) = \alpha + f'(x_0)x_0 \quad \Longleftrightarrow \quad \alpha = f(x_0) - f'(x_0)x_0.$$

- Plugging this back into the expression for the tangent yields

$$y(x) = f(x_0) - f'(x_0)x_0 + f'(x_0)x = f(x_0) + f'(x_0)(x - x_0).$$

- And we are done.

Graphical illustration



Example

- Consider the function $f(x) = 5x^3 - 7x$ and form the linear approximation at $x_0 = 4$.
- Then $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$ implies

$$f(x) \approx f(4) + f'(4)(x-4) = \left(5 \cdot 4^3 - 7 \cdot 4\right) + \left(15 \cdot 4^2 - 7\right)(x-4).$$

- Suppose we want to evaluate the function at $x = 5$.
  - The true value of the function is $f(5) = 5 \cdot 5^3 - 7 \cdot 5 = 590$. The approximated value is $f(x) \approx \left(5 \cdot 4^3 - 7 \cdot 4\right) + \left(15 \cdot 4^2 - 7\right)(5-4) = 525$. The *approximation error* is $525 - 590 = -65$.
- Suppose we want to evaluate the function at $x = 6$.
  - The true value of the function is 1038, and the approximated value is 838. The error is $838 - 1038 = -200$.
- It is typically (but not always) the case that the further away from the approximation point we evaluate the function, the less precise the approximation becomes.

Linear approximations to multivariate functions

- The same idea of a linear approximation applies when we are dealing with multivariate functions.
- Recalling that **bold** notation denotes vectors, we have the following definition.

Definition

The linear approximation to $z = f(\mathbf{x}) = f(x_1, ..., x_n)$ around $\mathbf{x}_0 = (x_1^0, ..., x_n^0)$ is given by

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + f_1'(\mathbf{x}_0)(x_1 - x_1^0) + f_2'(\mathbf{x}_0)(x_2 - x_2^0) + ... + f_n'(\mathbf{x}_0)(x_n - x_n^0).$$

Quadratic approximations

- We can also consider approximating the function through a quadratic equation rather than a linear equation.
- A quadratic polynomial near x_0 corresponds to a function $y(x) = A + B(x - x_0) + C(x - x_0)^2$.
- To find the parameters A, B and C we impose the restrictions that at $x = x_0$, $f(x)$ and $y(x)$ should have the same value as well as the same first and second derivatives. That is, we impose $f(x_0) = y(x_0)$, $f'(x_0) = y'(x_0)$ and $f''(x_0) = y''(x_0)$.
- Inserting $x = x_0$ into the expression for $y(x)$ yields $A = y(x_0)$. Note that $y'(x) = B + 2C(x - x_0)$ and $y''(x) = 2C$. Inserting $x = x_0$ it follows $B = y'(x_0)$ and we also have $C = \frac{1}{2}y''(x_0)$. Therefore:

Definition

The quadratic approximation to $f(x)$ around $x = x_0$ is

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2.$$

Example

- Find the quadratic approximation to

$$f(x) = \sqrt[3]{x}$$

  around $x_0 = 1$.
- Answer:
  - Note that $\sqrt[3]{x} = x^{\frac{1}{3}}$. Then it is easy to see that $f'(x) = \frac{1}{3}x^{-\frac{2}{3}}$ and $f''(x) = \frac{1}{3}\left(-\frac{2}{3}\right)x^{-\frac{5}{3}} = -\frac{2}{9}x^{-\frac{5}{3}}$.
  - It follows that $f'(1) = \frac{1}{3}$ and $f''(1) = -\frac{2}{9}$.
  - Note that $f(1) = 1$ and use the quadratic approximation formula to obtain

$$\sqrt[3]{x} \approx 1 + \frac{1}{3}(x-1) - \frac{1}{9}(x-1)^2.$$

  - For example, with this result we find $\sqrt[3]{1.03} \approx 1.0099$ as an approximation compared to the true value $\sqrt[3]{1.03} = 1.00990163$. Not too bad!

- Generalizing the idea behind the quadratic approximation leads to the following definition.

Definition

A Taylor approximation of order n to $f(x)$ around $x = x_0$ is

$$f(x) \approx f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n,$$

where $i! = 1 \cdot 2 \cdot \ldots \cdot (i-1) \cdot i$ is the *factorial* for any $i \in \mathbb{N}$.

- Are you wondering why it is just $\approx$ and not $=$? Because this is just an approximation that is not exactly equal to the function.
- The approximation tends to be better
  1. the higher the order of the polynomial, or
  2. the less funky the original function is at x_0.

- Taylor, however, gave a precise answer to what the error is when one uses an approximation. He calculated the difference between the function and its approximation. The resulting formula was called *Taylor's formula*.
- Taylor's formula is studied in more advanced modules.
- Exercise for you at home: Form the second-order Taylor approximation to the function $f(x) = \ln(x) + 5x$ around $x_0 = 2$. Then compare the true and the approximated values of the function for $x = 2.1$ and $x = 2.2$.
- Typical application in macroeconomics: Suppose you study economic growth, for instance the famous Solow model. You want to understand what happens to economic growth when the government increases its spending (say, more spending on schools, hospitals etc.)
  - You first solve the model for its equilibrium ("steady state").
  - You then 'shock' the system with the change in government policy.
  - Since the system is complicated, researchers typically use a Taylor approximation around the steady state to understand the dynamics (i.e. how the economy will reach a new steady state).

Topic VI: Matrix and vector algebra

EC123 Mathematical Techniques B

Dr Dennis Novy

University of Warwick

Autumn Term 2015

Topic VI: Matrix and vector algebra

Topic VI's big picture:

- We are starting completely different material: matrix algebra.
- We will study matrices and a whole bunch of operations that can be performed with them.
- We will apply them to solve systems of equations—many economic problems can be thought of as solving a system of equations.
- Matrices are widely used in econometrics.
- By the way, the world of *algebra* tells us how objects (numbers, matrices) and operators are combined. The world of *calculus* tells us how objects change.
- The topic outline is as follows:

VI.1 Definition of a matrix
VI.2 Systems of equations
VI.3 Operations with matrices
VI.4 Vectors

VI.1 Definition of a matrix
(Ch. 15.2)

Definition

A matrix is a rectangular array of numbers which are considered an entity. Each matrix consists of a series of m rows and n columns with $m, n \geq 1$. A matrix $\mathbf{A}$ then corresponds to

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

where each a_{ij} is called an element of the matrix.

- For each element a_{ij}, $i = 1, \dots, m$ denotes a particular row and $j = 1, \dots, n$ denotes a particular column. Thus $a_{\text{row,column}}$.
- The elements in a matrix can be pretty much anything: variables, functions, parameters etc.

Definition

The dimension of a matrix $\mathbf{A}$ corresponds to its numbers of rows and columns. In a matrix with m rows and n columns, we say the matrix has dimension $(m \times n)$, or it is of order $(m \times n)$.

- Example 1: If

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 10 \\ 2 & 0 & 8 \end{bmatrix},$$

then $\mathbf{A}$ is a matrix of dimension (2×3).

- Example 2:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 9 & 0 & 0 \\ 5 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is of dimension (3×5).

Definition

A vector is a matrix where $m = 1$ or $n = 1$. Thus a vector is a special case of a matrix with just one row or just one column.

- Example 1:

$$\begin{bmatrix} 1 \\ 7 \\ 0 \end{bmatrix}$$

  When a matrix has just one column, we say it is a column vector of dimension m (with $m = 3$ in this example.)

- Example 2:

$$\begin{bmatrix} 3 & 4 & 0 & 1 \end{bmatrix}$$

  When a matrix has just one row, we say it is a row vector of dimension n (with $n = 4$ in this example.)

Notation for matrices and elements

- We usually denote matrices by **bold** upper-case letters like $\mathbf{A}$ or $\mathbf{B}$.
- Sometimes we also denote a matrix by its elements, i.e.

$$\mathbf{A} = [a_{ij}]_{m \times n},$$

where we make the matrix dimension explicit.

- We usually denote vectors by a **bold** lower-case letter such as $\mathbf{a}$ or $\mathbf{b}$.
- Sometimes we refer to a particular element of a matrix by referring to its position in the matrix:

$$a_{ij} = [\mathbf{A}]_{ij}.$$

- For example, the element $a_{12} = [\mathbf{A}]_{12}$ from the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 10 \\ 3 & 0 & 8 \end{bmatrix}$$

corresponds to 2.

- A null matrix or zero matrix is a matrix where $a_{ij} = 0$ for all i, j, that is, a matrix with only zeros. We denote such matrix by $\mathbf{0}_{m,n}$. For example,

$$\mathbf{0}_{3,2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

is a null matrix of dimension (3×2).

- A square matrix is a matrix where $n = m$. The elements a_{ii} for all $i = 1, \ldots, n$ constitute the main diagonal of the matrix. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 7 & 1 \\ 9 & 3 & 8 \\ 4 & 5 & 6 \end{bmatrix}$$

is a square matrix of dimension 3 with the main diagonal in red.

- A triangular matrix is a square matrix that has only zeros either above or below the main diagonal. If the non-zeros are below the diagonal, the matrix is *lower triangular*. If the non-zeros are above the diagonal, the matrix is *upper triangular*. Here is an example of a lower triangular matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 10 \end{bmatrix}.$$

- An identity matrix is a square matrix where $a_{ij} = 0$ for all $i \neq j$ and $a_{ij} = 1$ for all $i = j$. We denote such matrix by $\mathbf{I}_m$. It therefore has the following structure:

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

- For example,

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is an identity matrix of dimension 3 (it would be redundant to say (3×3) because we know it is a square matrix).

VI.2 Systems of equations

Definition

A system of equations is a collection of equations.

- We have already dealt with systems of equations in one of the examples of implicit functions.
  - Consider this simple example:

$$Y = C + I,$$
$$C = C(Y),$$

    where Y is income, C is consumption and I is investment.
  - This system has two equations, one representing the income accounting identity and the other describing the behaviour of consumption.
  - The interesting thing about this particular system is that the two equations are interrelated: to know income you must know how consumption behaves, which itself depends on income as indicated by the second equation. You will deal with such systems in macroeconomics.

Definition

A system of m equations and n unknowns corresponds to

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, \\ &\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m. \end{aligned}$$

Here $(a_{11}, a_{12}, \ldots, a_{mn})$ are called the coefficients of the system and $(b_1, \ldots, b_m)$ are constants on the right-hand side. All are real numbers. The x's are the system's variables.

- Notice the notation in the coefficient subscripts: a_{ij} is the coefficient of the jth variable in the ith equation. Thus $a_{\text{equation, variable}}$.
- Many coefficients can be zero.

Now the matrices become handy!

Definition

The system of m equations and n variables

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ &\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m, \end{aligned}$$

can be written in matrix notation as $\mathbf{Ax} = \mathbf{b}$ where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

A simple example

Consider the system

$$
\begin{aligned}
3x_1 + 5x_2 + 16x_3 &= 12, \\
12x_1 + 15x_2 + 7x_3 &= 11.
\end{aligned}
$$

It can be written in matrix notation as $\mathbf{Ax} = \mathbf{b}$ where

$$
\mathbf{A} = \begin{bmatrix} 3 & 5 & 16 \\ 12 & 15 & 7 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 12 \\ 11 \end{bmatrix}.
$$

Motivation:

- We care about matrix notation because it will help us to solve systems of equations. You will soon see how.
- This is also crucial for econometrics, say, OLS estimation.

VI.3 Operations with matrices
(Ch. 15.2)

We will now study how to operate with matrices. In particular, we will study

1. equality,
2. addition,
3. multiplication by scalar,
4. multiplication,
5. transposition,
6. determinants,
7. inverse matrices,

and we will see some special matrices that have interesting properties given certain operations.

(1) Equality

Definition

If $\mathbf{A} = [a_{ij}]_{m \times n}$ and $\mathbf{B} = [b_{ij}]_{m \times n}$ are two $m \times n$ matrices, then $\mathbf{A}$ and $\mathbf{B}$ are said to be equal if $a_{ij} = b_{ij}$ for all i and all j. In that case we can write $\mathbf{A} = \mathbf{B}$.

- Thus two matrices are equal if all their elements are equal.
- If there is at least one different element, we say the matrices are different and write $\mathbf{A} \neq \mathbf{B}$.

(2) Addition

Definition

If $\mathbf{A} = [a_{ij}]_{m \times n}$ and $\mathbf{B} = [b_{ij}]_{m \times n}$ are two $m \times n$ matrices, then the sum of $\mathbf{A}$ and $\mathbf{B}$ corresponds to the matrix $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]_{m \times n}$.

- Thus the sum of matrices requires the summation element by element.
- Example:

$$\begin{bmatrix} 2 & 7 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 4 & 6 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 6 & 13 \\ 4 & 7 \end{bmatrix}.$$

- As a consequence, can we sum matrices of different dimensions? No.

(3) Multiplication by scalar

Definition

If α is a real number (a scalar) and $\mathbf{A} = [a_{ij}]_{m \times n}$, then $\alpha\mathbf{A} = [\alpha a_{ij}]_{m \times n}$.

- Thus multiplying a matrix by a scalar implies multiplying each of its elements by the scalar.
- Example: Let $\alpha = 2$ and $\mathbf{A} = \begin{bmatrix} 6 & 6 \\ 4 & 3 \end{bmatrix}$. Then $\alpha\mathbf{A} = \begin{bmatrix} 12 & 12 \\ 8 & 6 \end{bmatrix}$.
- Note that this definition implies that subtraction of two matrices is analogous to addition in the following sense: $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B} = [a_{ij} - b_{ij}]_{m \times n}$.

Rules for matrix addition and scalar multiplication

- Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be $(m \times n)$ matrices and let α, β be real numbers. Then

1. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$ (associative law),
2. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ (commutative law),
3. $\mathbf{A} + \mathbf{0} = \mathbf{A}$,
4. $\mathbf{A} - \mathbf{A} = \mathbf{0}$,
5. $(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$,
6. $\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$.

- Note that all these rules are analogous to those for ordinary algebra.

(4) Multiplication
(Ch. 15.3, 15.4)

- Up to now all operations have been straightforward. However, matrix multiplication is more challenging.
- Matrix multiplication involves multiplying and adding row and column elements.
- Suppose **A** and **B** correspond to

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}. \text{ Then } \mathbf{AB} = ?$$

- To compute **AB** we must **multiply each row of A with each column of B** to obtain individual multiplication elements as a result.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad \mathbf{AB} = \begin{pmatrix} ? & ? \\ ? & ? \end{pmatrix}.$$

1. The element $[\mathbf{AB}]_{11}$ corresponds to the multiplication and subsequent addition of the first row of **A** and the first column of **B**:

$$[\mathbf{AB}]_{11} = a_{11}b_{11} + a_{12}b_{21}.$$

2. The element $[\mathbf{AB}]_{12}$ corresponds to the multiplication and subsequent addition of the first row of **A** and the second column of **B**:

$$[\mathbf{AB}]_{12} = a_{11}b_{12} + a_{12}b_{22}.$$

3. Equivalently, we have

$$[\mathbf{AB}]_{21} = a_{21}b_{11} + a_{22}b_{21}.$$

4. Finally, we have

$$[\mathbf{AB}]_{22} = a_{21}b_{12} + a_{22}b_{22}.$$

Example 1

- Let **A** and **B** correspond to

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 7 & 5 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 6 & 4 \\ 2 & 1 \end{bmatrix}.$$

- To compute **AB** we need to go element by element:
  1. $\mathbf{AB}_{11} = 3 \cdot 6 + 2 \cdot 2 = 22,$
  2. $\mathbf{AB}_{12} = 3 \cdot 4 + 2 \cdot 1 = 14,$
  3. $\mathbf{AB}_{21} = 7 \cdot 6 + 5 \cdot 2 = 52,$
  4. $\mathbf{AB}_{22} = 7 \cdot 4 + 5 \cdot 1 = 33.$

- Thus

$$\mathbf{AB} = \begin{bmatrix} 22 & 14 \\ 52 & 33 \end{bmatrix}.$$

Definition

Suppose $\mathbf{A} = [a_{ij}]_{m\times n}$ and $\mathbf{B} = [b_{ij}]_{n\times p}$. Then, the product matrix $\mathbf{C} = \mathbf{AB}$ is an $(m \times p)$ matrix, $\mathbf{C} = [c_{ij}]_{m\times p}$, whose element in the ith row and the jth column is the **inner product**

$$c_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj}$$

of the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$.

- Note that the summation sign means $\sum_{r=1}^{n} a_{ir} b_{rj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}$ as you let the summation index r run from 1 to n.
- This definition extends our previous examples: there is no need for the matrices to be square matrices so that they can be multiplied.
- In fact, all we need is that the **number of columns in the first matrix is equal to the number of rows in the second matrix**:

$$(m \times n)(n \times p).$$

- Notice that the resulting matrix has the dimension $m \times p$.

Example 2

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 2 \\ 2 & 3 & 1 \\ 4 & -1 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3 & 2 \\ 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

- How many columns does **A** have? 3. How many rows does **B** have? 3. Therefore we can multiply them.
- The resulting matrix will be of dimension (3×2).
- The result is

$$\mathbf{AB} = \begin{bmatrix} -1 & 2 \\ 8 & 5 \\ 11 & 8 \end{bmatrix}.$$

- Question: If we have found **AB**, can we also find **BA**? No because $\mathbf{B}_{3\times 2}$, $\mathbf{A}_{3\times 3}$ and thus the 2 columns of **B** do not match the 3 rows of **A**.
- Thus, matrix multiplication is not commutative. If we can find **AB**, this does not necessarily mean that we can find **BA**.

Example 3

- Let

$$\mathbf{A} = \begin{bmatrix} 4 & 3 & 2 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix},$$

where **A** has dimension (1×4) and **B** has dimension (4×1). Therefore **AB** is a scalar: $\mathbf{AB} = 4 + 6 + 6 + 4 = 20$.

- Once more, is **BA** defined in this case? Yes because (4×1) is **conformable** with (1×4):

$$\mathbf{BA} = \begin{pmatrix} 1\cdot 4 & 1\cdot 3 & 1\cdot 2 & 1\cdot 1 \\ 2\cdot 4 & 2\cdot 3 & 2\cdot 2 & 2\cdot 1 \\ 3\cdot 4 & 3\cdot 3 & 3\cdot 2 & 3\cdot 1 \\ 4\cdot 4 & 4\cdot 3 & 4\cdot 2 & 4\cdot 1 \end{pmatrix} = \begin{pmatrix} 4 & 3 & 2 & 1 \\ 8 & 6 & 4 & 2 \\ 12 & 9 & 6 & 3 \\ 16 & 12 & 8 & 4 \end{pmatrix}.$$

Example 4

- Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & -2 \\ -1 & 1 \end{bmatrix},$$

  where **A** is (2×2) and **B** is (2×2). Therefore **AB** is also a square matrix (2×2), equal to:

$$\mathbf{AB} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

- Thus if $\mathbf{AB} = \mathbf{0}$, this does not necessarily mean that $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$.
- In fact, note that

$$\mathbf{BA} = \begin{bmatrix} 4 & 8 \\ -2 & -4 \end{bmatrix}.$$

Example 5

- Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 8 & 4 \\ 4 & -4 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 6 & 6 \\ 5 & -5 \end{bmatrix}.$$

  Then

$$\mathbf{AB} = \begin{bmatrix} 16 & -4 \\ -16 & 4 \end{bmatrix}, \quad \mathbf{AC} = \begin{bmatrix} 16 & -4 \\ -16 & 4 \end{bmatrix}.$$

- Thus, if $\mathbf{AB} = \mathbf{AC}$, this does not necessarily mean that $\mathbf{B} = \mathbf{C}$.
- Also note that $\mathbf{AB}$ is generally not the same as $\mathbf{BA}$. That is, there is a difference between *premultiplication* and *postmultiplication*. When we write $\mathbf{AB}$, we say that $\mathbf{B}$ is premultiplied by $\mathbf{A}$, or $\mathbf{A}$ is postmultiplied by $\mathbf{B}$.

Example 6: Multiplying by the identity matrix

- Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where **B** is an identity matrix. Then

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix} = \mathbf{A}.$$

- This example illustrates that the identity matrix $\mathbf{I}_n$ is like '1' in the case of real numbers ($\alpha \cdot 1 = \alpha$ for any scalar α). As a consequence, for any $\mathbf{A}_{m \times n}$ and $\mathbf{I}_n$ we have

$$\mathbf{AI} = \mathbf{A}.$$

Example 7: A system of equations

- Take the system

$$x_1 + 2x_2 = 2,$$
$$2x_1 - x_2 = 4.$$

- How many variables are there in this system? Two: x_1 and x_2.
- There are four coefficients: $a_{11} = 1$, $a_{12} = 2$, $a_{21} = 2$ and $a_{22} = -1$.
- We can rewrite this system of equations in matrix form as $\mathbf{Ax} = \mathbf{b}$ with

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

$\Rightarrow$ The solution to this system of equations will depend on the properties of **A**.

Example 8: Idempotent matrices

- A square matrix $\mathbf{A}$ is said to be an idempotent matrix if $\mathbf{A}^2 = \mathbf{A}$.
- For example, the following matrix is idempotent:

$$\mathbf{A} = \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix}.$$

- Verify this at home.

Rules for matrix multiplication

- $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ (associative law),
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ (left distributive law),
- $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ (right distributive law).

Some final notes:

1) We denote the multiplication of a square matrix n times by itself as $\mathbf{A}^n$ and refer to it as the nth power of the matrix. For example, $\mathbf{A}^3 = \mathbf{AAA}$.
2) There is no such thing as matrix division.

(5) Transposition
(Ch. 15.5)

Definition

Let $\mathbf{A}$ be an $(m \times n)$ matrix. The transpose of $\mathbf{A}$ corresponds to the $(n \times m)$ matrix whose first column is the first row of $\mathbf{A}$, whose second column is the second row of $\mathbf{A}$ and so on. Thus

$$\text{If } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \text{ then } \mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}.$$

We denote the transpose by $\mathbf{A}'$ (sometimes it is also denoted by $\mathbf{A}^T$).

Examples

- If
$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix}, \text{ then } \quad \mathbf{A}' = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}.$$
- If
$$\mathbf{A} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \text{ then } \quad \mathbf{A}' = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}.$$

1. $(\mathbf{A}')' = \mathbf{A}$,
2. $(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'$,
3. $(\alpha\mathbf{A})' = \alpha\mathbf{A}'$ where α is a scalar,
4. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

- Easy! The only rule that may look a bit more mysterious is the last one.

Definition

A matrix $\mathbf{A}$ is said to be symmetric if it is equal to its transpose $\mathbf{A} = \mathbf{A}'$.

- For example, the identity matrix and a square null matrix are symmetric.
- As another example, the following matrix is symmetric:

$$\mathbf{A} = \begin{bmatrix} 7 & 5 \\ 5 & 23 \end{bmatrix}.$$

- Note that only square matrices can be symmetric.

(6) The determinant of a matrix
(Ch. 16.1, 16.2, 16.4, 16.5)

- Every square matrix has an associated scalar called its *determinant*.
- We start with the $n = 2$ case.

Definition

Let $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ be a square two-dimensional matrix. Its determinant, denoted by $|\mathbf{A}|$, corresponds to $|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$.

- If $n = 1$ then $\mathbf{A} = a_{11}$ and $|\mathbf{A}| = a_{11}$.
- In some textbooks the determinant is denoted by $\det(\mathbf{A})$.

Example

- Let

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 7 & 5 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 6 & 4 \\ 2 & 1 \end{bmatrix}.$$

- The determinants are

$$|\mathbf{A}| = 3 \cdot 5 - 2 \cdot 7 = 1, \quad |\mathbf{B}| = 6 \cdot 1 - 4 \cdot 2 = -2.$$

Minors and cofactors

- When $n \geq 3$ we try to break down the determinant of the big matrix into the determinants of smaller (less-dimensional) matrices. The determinants of these smaller matrices (or sub-matrices) are called minors.
- Consider the (3×3) matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

- Its determinant is given by

$$|\mathbf{A}| = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix},$$

where $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = M_{11}$ is the minor corresponding to a_{11} etc.

Example 1

- Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 4 & 3 & 7 \\ 2 & 1 & 2 \\ 0 & 2 & 3 \end{pmatrix}.$$

- Applying the definition of the determinant for the case of $n = 3$ yields

$$\begin{aligned} |\mathbf{A}| &= a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13} \\ &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= 4\begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} - 3\begin{vmatrix} 2 & 2 \\ 0 & 3 \end{vmatrix} + 7\begin{vmatrix} 2 & 1 \\ 0 & 2 \end{vmatrix} \\ &= 4(3-4) - 3(6-0) + 7(4-0) \\ &= -4 - 18 + 28 = 6. \end{aligned}$$

- Consider the determinant of the following matrix:

$$\mathbf{H} = \begin{pmatrix} f''_{11} & f''_{12} \\ f''_{21} & f''_{22} \end{pmatrix},$$

where the elements of the matrix are second-order partial derivatives of a continuously differentiable bivariate function $z = f(x, y)$.

- The determinant is given by

$$|\mathbf{H}| = f''_{11} f''_{22} - f''_{12} f''_{21} = f''_{11} f''_{22} - (f''_{12})^2,$$

where we used Young's theorem for the last equality.

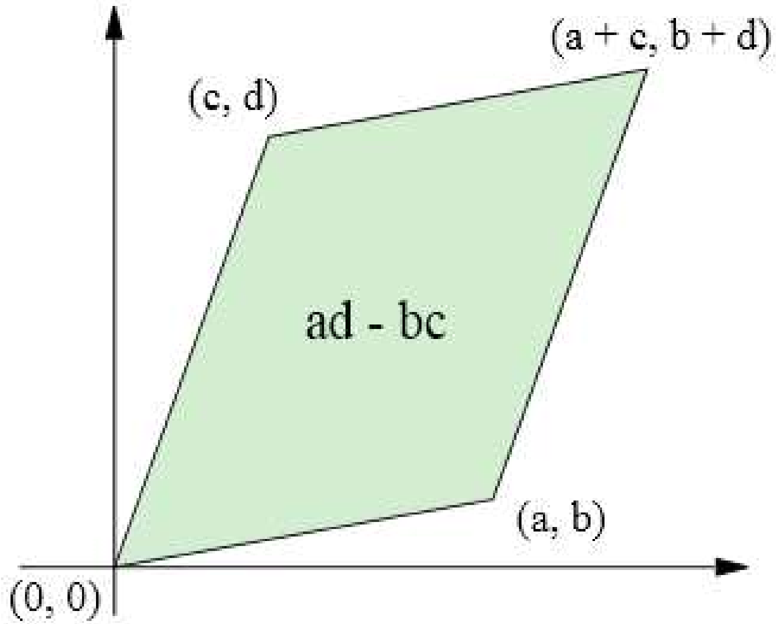
- This formula should look familiar.
- Thus, in the bivariate case convexity/concavity is a restriction over the determinant of the Hessian matrix $\mathbf{H}$.

Graphical interpretation as a parallelogram

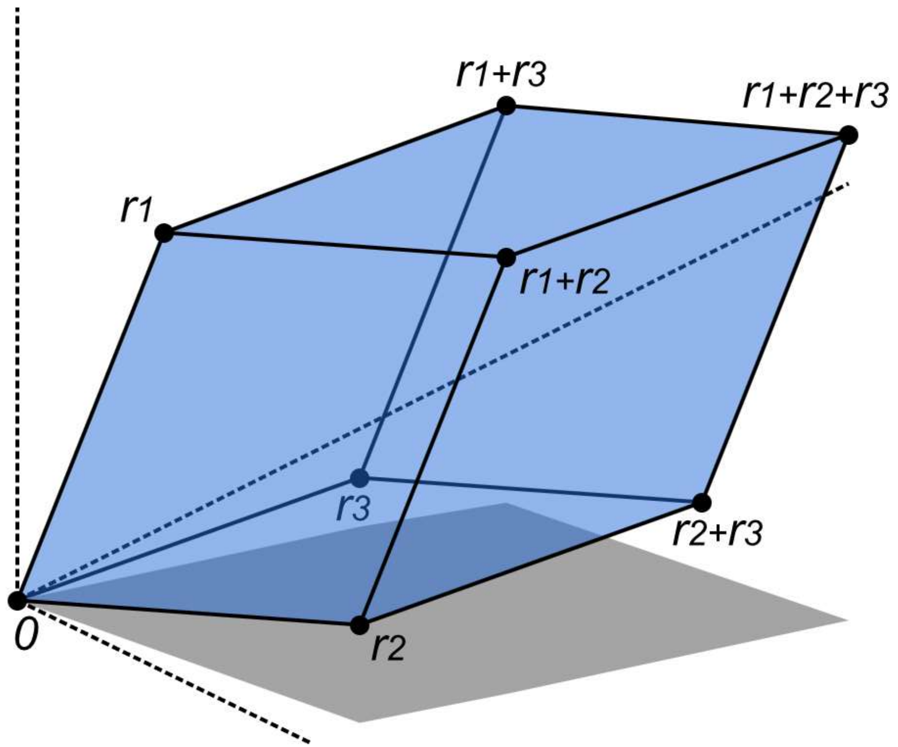
- Consider the determinant

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - ac.$$

- Think of (a, b) as one vector and (c, d) as another. Interpret a and c as values on the horizontal axis, b and d as values on the vertical axis.
- The determinant represents the area formed by the parallelogram given by these two vectors (picture source: Wikimedia Commons).



- In three-dimensional space we get a similar picture for a parallelepiped formed by the rows constructed from the vectors r_1, r_2 and r_3, where the determinant represents the volume of the parallelepiped (picture source: Wikimedia Commons).



Definition

Let $\mathbf{A}$ be an $(n \times n)$ matrix. The determinant $|\mathbf{A}|$ corresponds to

$$|\mathbf{A}| = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{ij}C_{ij} + \ldots + a_{in}C_{in},$$

where C_{ij} corresponds to the cofactor of the element a_{ij} for all i, j defined as

$$C_{ij} = (-1)^{i+j} \begin{vmatrix} a_{11} & \cdots & a_{1,j-1} & a_{1j} & a_{1,j+1} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2,j-1} & a_{2j} & a_{2,j+1} & \cdots & a_{2n} \\ \vdots & & \vdots & \vdots & \vdots & 0 & \vdots \\ a_{i1} & \cdots & a_{i,j-1} & a_{ij} & a_{i,j+1} & \cdots & a_{in} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & a_{nj} & a_{n,j+1} & \cdots & a_{nn} \end{vmatrix}$$

where you have to delete the ith row and the jth column. Thus, the cofactor C_{ij} consists of the minor M_{ij} multiplied by $(-1)^{i+j}$.

Example of a cofactor

- Take the matrix

$$\begin{pmatrix} 3 & 0 & 0 & 7 \\ 2 & 1 & 3 & 1 \\ 4 & 1 & 4 & 2 \\ 9 & 3 & 4 & 0 \end{pmatrix}.$$

- The cofactor of element a_{23}, C_{23}, corresponds to

$$C_{23} = (-1)^{2+3} \begin{vmatrix} 3 & 0 & 0 & 7 \\ 2 & 1 & 3 & 1 \\ 4 & 1 & 4 & 2 \\ 9 & 3 & 4 & 0 \end{vmatrix} = (-1)^{2+3} \begin{vmatrix} 3 & 0 & 7 \\ 4 & 1 & 2 \\ 9 & 3 & 2 \end{vmatrix}$$

- There is a small typo in the previous equation. Can you spot it?

- Recall our previous example:

$$\mathbf{A} = \begin{pmatrix} 4 & 3 & 7 \\ 2 & 1 & 2 \\ 0 & 2 & 3 \end{pmatrix}.$$

- Applying the general definition based on the first row yields

$$|\mathbf{A}| = \underbrace{4(-1)^{1+1}\begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix}}_{C_{11}} + \underbrace{3(-1)^{1+2}\begin{vmatrix} 2 & 2 \\ 0 & 3 \end{vmatrix}}_{C_{12}} + \underbrace{7(-1)^{1+3}\begin{vmatrix} 2 & 1 \\ 0 & 2 \end{vmatrix}}_{C_{13}}$$

$$= 4\begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} - 3\begin{vmatrix} 2 & 2 \\ 0 & 3 \end{vmatrix} + 7\begin{vmatrix} 2 & 1 \\ 0 & 2 \end{vmatrix} = 6.$$

- The solution is exactly the same as previously obtained.

- Note that the general formula does not tell us which particular row of cofactors to pick. Check at home that in the last example we would have obtained the same solution by picking the row of cofactors C_{21}, C_{22} and C_{23} or the row of cofactors C_{31}, C_{32} and C_{33} with the corresponding elements as multiplicative factors.
- Actually, we could have even picked a *column j* and we would have arrived at the same solution

$$|\mathbf{A}| = a_{1j}C_{1j} + a_{2j}C_{2j} + ... + a_{ij}C_{ij} + ... + a_{nj}C_{nj},$$

where the cofactors are defined in an analogous way. (We will not discuss why this is the case. If you are curious, see a general definition of determinants in chapter 16.3 of the textbook.)

$\Rightarrow$ When solving an exercise in this context, choose the row or column wisely.

- For example, which column or row would you pick here?

$$|\mathbf{A}| = \begin{vmatrix} 3 & 1 & 2 \\ 4 & 1 & 2 \\ 9 & 0 & 0 \end{vmatrix}$$

Rules for determinants

Let **A** be an $(n \times n)$ matrix. Then:

1. If all the elements in a row (or column) of **A** are zero, then $|\mathbf{A}| = 0$.
2. $|\mathbf{A}'| = |\mathbf{A}|$.
3. If all the elements in a **single** row or column of **A** are multiplied by a number α, the determinant is $\alpha|\mathbf{A}|$.
4. If two rows (or two columns) of **A** are interchanged, the determinant changes sign but its absolute value remains the same.
5. If two of the rows (or columns) of **A** are proportional, then $|\mathbf{A}| = 0$.
6. The value of the determinant of **A** is unchanged if a multiple of one row (or one column) is added to a different row (or column) of **A**.
7. The determinant of the product of two $n \times n$ matrices **A** and **B** is the product of the determinants of each of the matrices: $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|$.
8. If α is a real number, then $|\alpha\mathbf{A}| = \alpha^n|\mathbf{A}|$.

- Finally, it is important to note that $|\mathbf{A} + \mathbf{B}| \neq |\mathbf{A}| + |\mathbf{B}|$ in general.
- **Very important:** Take pen and paper and play with each of these rules at home.

Example for rule 5

- "If two of the rows (or columns) of $\mathbf{A}$ are proportional, then $|\mathbf{A}| = 0$."
- This rule will be important for your econometrics lectures (especially once you start dealing with actual data).
- Consider the determinant of the following matrix:

$$\begin{vmatrix} a_{11} & a_{12} \\ \beta a_{11} & \beta a_{12} \end{vmatrix} = a_{11}\beta a_{12} - a_{12}\beta a_{11} = 0.$$

- The fact that this determinant is zero is usually not advantageous because—as we will discuss now—it implies that we will not be able to find the "inverse matrix."

Definition

Let $\mathbf{A}$ be an $(n \times n)$ matrix. Then the inverse matrix $\mathbf{A}^{-1}$ of $\mathbf{A}$ is the matrix that solves

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n.$$

- Notice that $\mathbf{A}$ is a square matrix.
- Example: Consider the matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ -\frac{3}{2} & \frac{1}{2} \end{bmatrix}.$$

  Question: Is $\mathbf{B}$ the inverse of $\mathbf{A}$?
- We know that if $\mathbf{B}$ is the inverse of $\mathbf{A}$, then it must be the case that

$$\mathbf{AB} = \mathbf{I}_2 \quad \text{and} \quad \mathbf{BA} = \mathbf{I}_2.$$

Thus

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{3}{2} & \frac{1}{2} \end{bmatrix} \\ &= \begin{bmatrix} 1 \cdot 1 + 0 \cdot \left(-\frac{3}{2}\right) & 1 \cdot 0 + 0 \cdot \frac{1}{2} \\ 3 \cdot 1 + 2 \cdot \left(-\frac{3}{2}\right) & 3 \cdot 0 + 2 \cdot \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2. \end{aligned}$$

Equivalently

$$\mathbf{BA} = \begin{bmatrix} 1 & 0 \\ -\frac{3}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2.$$

Thus, the answer is yes. $\mathbf{B}$ is indeed the inverse of $\mathbf{A}$, that is, $\mathbf{B} = \mathbf{A}^{-1}$.
Of course, it is also true that $\mathbf{A}$ is the inverse of $\mathbf{B}$, that is, $\mathbf{A} = \mathbf{B}^{-1}$.

Finding the inverse

- Suppose we want to find the inverse of the following matrix:

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

- We know that the inverse must be such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Thus

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \underbrace{\begin{pmatrix} x & y \\ z & w \end{pmatrix}}_{\text{unknown } A^{-1}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}.$$

- By multiplying we get the following system of equations:

$$ax + bz = 1, \qquad ay + bw = 0,$$
$$cx + dz = 0, \qquad cy + dw = 1.$$

- Solving the system for x, y, z and w (four equations in four unknowns) we get

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{|\mathbf{A}|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

- Make sure to check this carefully at home.
- Example: Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix}.$$

- The inverse is

$$\mathbf{A}^{-1} = \frac{1}{(4-2)} \begin{pmatrix} 4 & -1 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -1/2 \\ -1 & 1/2 \end{pmatrix}.$$

The general formula for the inverse

Theorem

Any square matrix $\mathbf{A}_{n\times n}$ *with a determinant* $|\mathbf{A}| \neq 0$ *has a unique inverse* $\mathbf{A}^{-1}$*. This is given by*

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\,\mathrm{adj}(\mathbf{A}),$$

where $\mathrm{adj}(\mathbf{A})$ *is the adjoint of* $\mathbf{A}$ *corresponding to*

$$\mathrm{adj}(\mathbf{A}) = (\mathbf{C}^{+})' = \begin{pmatrix} C_{11} & \dots & C_{k1} & \dots & C_{n1} \\ C_{12} & \dots & C_{k2} & \dots & C_{n2} \\ \vdots & & \vdots & & \vdots \\ C_{1n} & \dots & C_{kn} & \dots & C_{nn} \end{pmatrix},$$

where C_{ij} *is the cofactor of element* a_{ij} *for all* i, j *and* $\mathbf{C}^{+}$ *denotes the matrix of cofactors.*

Example

- Take the following matrix $\mathbf{A} = \begin{pmatrix} 2 & 3 & 4 \\ 4 & 3 & 1 \\ 1 & 2 & 4 \end{pmatrix}$.
- Its determinant is $|\mathbf{A}| = -5 \neq 0$ so that the inverse exists.
- Let's calculate the cofactors:

$$C_{11} = (-1)^{1+1} \begin{vmatrix} 3 & 1 \\ 2 & 4 \end{vmatrix} = 10, \quad C_{12} = (-1)^{1+2} \begin{vmatrix} 4 & 1 \\ 1 & 4 \end{vmatrix} = -15,$$

$$C_{13} = (-1)^{1+3} \begin{vmatrix} 4 & 3 \\ 1 & 2 \end{vmatrix} = 5, \quad C_{21} = (-1)^{2+1} \begin{vmatrix} 3 & 4 \\ 2 & 4 \end{vmatrix} = -4,$$

$$C_{22} = 4, \quad C_{23} = -1, \quad C_{31} = -9, \quad C_{32} = 14, \quad C_{33} = -6.$$

- Now we form the adjoint of $\mathbf{A}$:

$$\text{adj}(\mathbf{A}) = \begin{pmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{pmatrix} = \begin{pmatrix} 10 & -4 & -9 \\ -15 & 4 & 14 \\ 5 & -1 & -6 \end{pmatrix}.$$

- Notice the transposition of the C_{ij} elements (hence $(\mathbf{C}^{+})'$ in the formula above and not $\mathbf{C}^{+}$).

- Finally we obtain

$$\begin{aligned}\mathbf{A}^{-1} &= \frac{1}{|\mathbf{A}|}\,\mathrm{adj}(\mathbf{A}) \\ &= \frac{1}{-5}\begin{pmatrix} 10 & -4 & -9 \\ -15 & 4 & 14 \\ 5 & -1 & -6 \end{pmatrix} \\ &= \begin{pmatrix} -2 & 4/5 & 9/5 \\ 3 & -4/5 & -14/5 \\ -1 & 1/5 & 6/5 \end{pmatrix}.\end{aligned}$$

- You can now check the result by showing that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_3$. Go through this carefully at home.

Properties of inverse matrices: Existence

- Caution! Not all square matrices have an inverse.
- For example, the following matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 11 & 0 \end{bmatrix}$$

do not have inverses.

Definition

A matrix that has an inverse is called nonsingular. When a matrix has no inverse, we say it is a singular matrix.

- When is a matrix nonsingular?

Theorem

A necessary and sufficient condition for the matrix $\mathbf{A}_{n\times n}$ to be nonsingular is $|\mathbf{A}| \neq 0$.

- For example,

$$|\mathbf{A}| = \begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} = 1 \cdot 4 - 2 \cdot 2 = 0$$

and

$$|\mathbf{B}| = \begin{vmatrix} 1 & 0 \\ 11 & 0 \end{vmatrix} = 1 \cdot 0 - 0 \cdot 11 = 0.$$

- The matrices **A** and **B** have determinants equal to zero and are thus singular.

Properties of inverse matrices: Uniqueness

- One nice property of inverse matrices is that if they exist, there is only one of them.

Theorem

If $\mathbf{A}$ is an invertible matrix, then its inverse $\mathbf{A}^{-1}$ is unique.

Proof by contradiction:

- Suppose $\mathbf{A}$ has two different inverses $\mathbf{B}$ and $\mathbf{C}$. Then it must be the case that

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n \quad \text{and} \quad \mathbf{AC} = \mathbf{CA} = \mathbf{I}_n.$$

Then

$$\mathbf{CAB} = \mathbf{C}(\mathbf{AB}) = \mathbf{C} \quad \text{and} \quad \mathbf{CAB} = (\mathbf{CA})\mathbf{B} = \mathbf{B}.$$

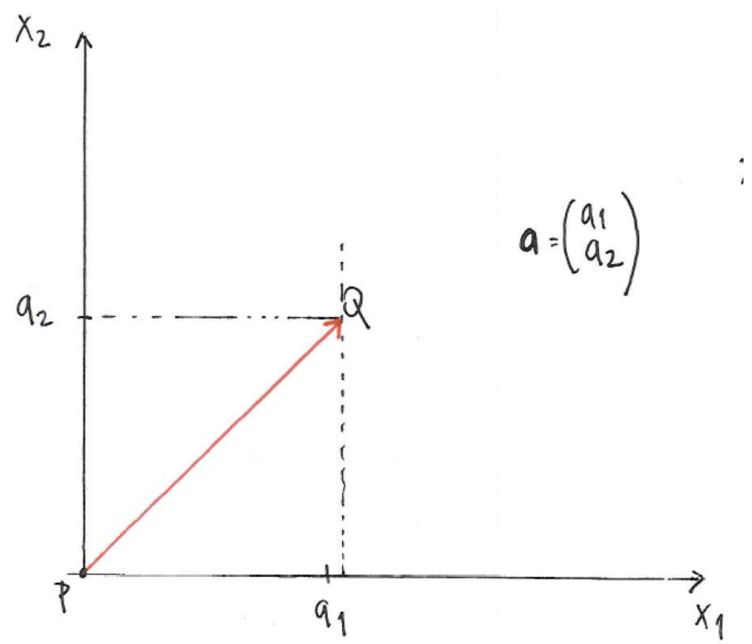
Therefore $\mathbf{B} = \mathbf{C}$. $\mathbf{B}$ and $\mathbf{C}$ cannot be different.

- Let $\mathbf{A}$ and $\mathbf{B}$ be invertible $n \times n$ matrices. Then
  1. $\mathbf{A}^{-1}$ is invertible and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$,
  2. $\mathbf{AB}$ is invertible and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$,
  3. the transpose $\mathbf{A}'$ is invertible and $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$,
  4. $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$ whenever α is a number $\neq 0$.

VI.4 Vectors
(Ch. 15.7, 15.8)

- Recall that a vector is nothing but a matrix with one column or one row.
- An example of a *column vector* is $\begin{pmatrix} 7 \\ 13 \\ 4 \end{pmatrix}$. An example of a *row vector* is $(7, 13, 4)$. The transpose of the column vector is the row vector, that is $\begin{pmatrix} 7 \\ 13 \\ 4 \end{pmatrix}' = (7, 13, 4)$.
- In a vector $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ we denote each a_i as a component (or coordinate). We say that the vector has dimension n.
- All feasible operations from matrix algebra apply.

- Vectors (unlike matrices) can be represented graphically when they are of dimension 2 or 3. For example, consider the vector $\mathbf{a} = (a_1, a_2)$.



⇒ We can think of vectors as representing a movement (and thus a distance) between the point P and the point Q in the x_1, x_2 plane.

Definition

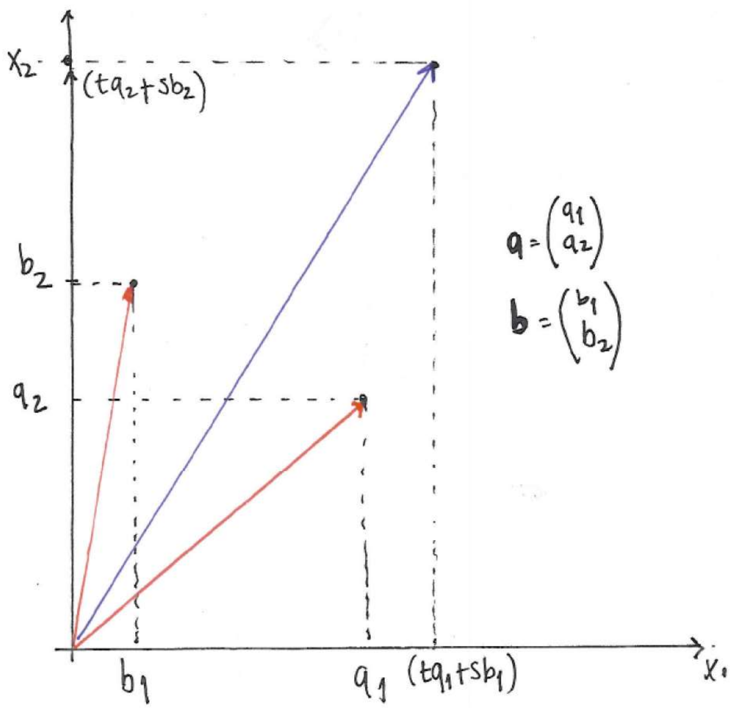
Let $\mathbf{a}$ and $\mathbf{b}$ be two n-vectors and let t and s be real numbers. Then the vector $t\mathbf{a} + s\mathbf{b}$ given by

$$t\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + s\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} ta_1 + sb_1 \\ ta_2 + sb_2 \\ \vdots \\ ta_n + sb_n \end{pmatrix}$$

is called a linear combination of $\mathbf{a}$ and $\mathbf{b}$.

- Example: Let $\mathbf{a} = (1,2,3)', \mathbf{b} = (4,5,6)', t = s = 0.5$.Thus $t\mathbf{a} + s\mathbf{b} = 0.5(1,2,3)' + 0.5(4,5,6)' = (0.5,1,1.5)' + (2,2.5,3)' = (2.5,3.5,4.5)'$. As t and s are exactly 0.5, the linear combination $t\mathbf{a} + s\mathbf{b}$ is the simple average of $\mathbf{a}$ and $\mathbf{b}$ for each component.

- Graphically



- Note that a linear combination can also be defined with matrices but this cannot usually be represented by a graph.

Inner product

Definition

The inner product of two n-dimensional vectors $\mathbf{a}$ and $\mathbf{b}$ corresponds to

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n.$$

- Example: Let

$$\mathbf{a} = (1, 2, 3)', \quad \mathbf{b} = (4, 5, 6)'.$$

Thus the inner product is given by

$$\mathbf{a} \cdot \mathbf{b} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32.$$

Rules for the inner product

If $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are n-vectors and α is a scalar, then

1. $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$,
2. $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$,
3. $(\alpha\mathbf{a}) \cdot \mathbf{b} = \mathbf{a} \cdot (\alpha\mathbf{b}) = \alpha(\mathbf{a} \cdot \mathbf{b})$,
4. $\mathbf{a} \cdot \mathbf{a} > 0 \quad \Leftrightarrow \quad \mathbf{a} \neq 0$.

- However, the inner product does not have a straightforward extension to more general matrices. The analogue is the so-called *Kronecker product* (not covered in this module.)

Length of a vector

- We now formally define the length of a vector in n-space.

Definition

If $\mathbf{a} = (a_1, \ldots, a_n)$, we define the length (or norm) of the vector $\mathbf{a}$, denoted by $||\mathbf{a}||$, as

$$||\mathbf{a}|| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{a_1^2 + a_2^2 + ... + a_n^2}.$$

- The length of a vector measures the distance between the *origin* $(0, \ldots, 0)$ and $\mathbf{a}$. Why?

Definition

The distance between to vectors $\mathbf{x}$ and $\mathbf{y}$ corresponds to

$$||\mathbf{x} - \mathbf{y}|| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2}.$$

- Recall the definition of the Euclidean distance.

Orthogonality

- One last property of vectors we consider is *orthogonality* (you will use this property a lot in econometrics).

Definition

Two vectors $\mathbf{a}$ and $\mathbf{b}$ are said to be orthogonal if

$$\mathbf{a} \cdot \mathbf{b} = 0.$$

- In other words, the inner product of $\mathbf{a}$ and $\mathbf{b}$ is zero.
- Sometimes orthogonality is denoted as follows: $\mathbf{a} \cdot \mathbf{b} = 0 \iff \mathbf{a} \perp \mathbf{b}$.

Topic VII: Systems of linear equations

EC123 Mathematical Techniques B

Dr Dennis Novy

University of Warwick

Autumn Term 2015

Topic VII: Systems of linear equations

Topic VII's big picture:

- We now study how to solve systems of *linear* equations given our knowledge of matrix algebra.
- Systems of equations are very common in economic and econometric models.
- The topic outline is as follows:

VII.1 Definition of a system of equations

VII.2 Solving systems of equations by matrix inversion

VII.3 Cramer's rule

VII.4 Existence and uniqueness of solutions

VII.5 Linear dependence and the rank of a matrix

- Some of this material is not covered in the textbook. But it is fine if you simply study these slides.

VII.1 Definition of a system of equations

Definition

A system of equations is a collection of equations.

- Remember the simple macroeconomic example from Topic VI:

$$Y = C + I,$$
$$C = C(Y),$$

  where Y is income, C is consumption and I is investment.
- This system has two equations, one capturing the national income accounting identity and the other describing the behaviour of consumption.

Definition

A system of m equations and n unknowns corresponds to

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m. \end{aligned}$$

Here $(a_{11}, a_{12}, \dots, a_{mn})$ are called the coefficients of the system and $(b_1, \dots, b_m)$ are constants on the right-hand side. All are real numbers. The x's are the system's variables.

- Notice the notation in the coefficient subscripts: a_{ij} is the coefficient of the jth variable in the ith equation. Thus $a_{\text{equation, variable}}$.
- Many coefficients can be zero.

Example 1

Consider the system

$$\begin{aligned} x_1 + x_2 - 2x_3 &= 2, \\ -x_2 + 2x_3 &= -1. \end{aligned}$$

- There are three variables in this system: x_1, x_2 and x_3. Thus $n = 3$.
- There are six coefficients: $a_{11} = 1$, $a_{12} = 1$, $a_{31} = -2$, $a_{21} = 0$, $a_{22} = -1$, $a_{23} = 2$. To see this note that the second equation can be rewritten as

$$0x_1 + (-1)x_2 + 2x_3 = -1.$$

Example 2

Consider the system

$$x_1 + 2x_2 = 2,$$
$$2x_1 - x_2 = 4.$$

- There are two variables in this system: x_1 and x_2. Thus $n = 2$.
- There are four coefficients: $a_{11} = 1$, $a_{12} = 2$, $a_{21} = 2$ and $a_{22} = -1$.

Definition

The system of m equations and n variables

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m, \end{aligned}$$

can be written in matrix notation as $\mathbf{Ax} = \mathbf{b}$ where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

- Writing the system of equations in matrix notation is not only shorter but it will help us to study its properties.
- The dimensions of the matrices are $(m \times n)$ for $\mathbf{A}$, $(n \times 1)$ for $\mathbf{x}$ and $(m \times 1)$ for $\mathbf{b}$.
- Note that this is a system of *linear* equations.
- We will not deal with systems of *non-linear* equations. But those can often be approximated by linear systems through a technique called *linearization* (essentially this works through the kind of linear approximation that we have already come across).

VII.2 Solving systems of equations by matrix inversion
(Ch. 16.6)

- Recall that a system of equations can be represented in matrix form as

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

  for a square matrix $\mathbf{A}_{n \times n}$, $\mathbf{x}_{n \times m}$ and $\mathbf{b}_{n \times m}$. Note that $m = 1$ in most applications.
- Remember that $|\mathbf{A}| \neq 0$ is a necessary and sufficient condition for the inverse of $\mathbf{A}$ to exist. Note that we are dealing with a *square* matrix $\mathbf{A}$ now. Otherwise we wouldn't be able to solve the system by matrix inversion.
- If the inverse of $\mathbf{A}$ exists, we can premultiply the equation by $\mathbf{A}^{-1}$ and get

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{I}_n\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$
$$\Leftrightarrow \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}.$$

- Thus the solution to the system of equations is given by $\mathbf{x}^*$ equal to $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$. [A common mistake is to write $\mathbf{b}\mathbf{A}^{-1}$ instead of $\mathbf{A}^{-1}\mathbf{b}$.]

Example

- Consider the following system of equations:

$$\begin{aligned} 3x_1 + x_2 &= 5, \\ 2x_1 + 2x_2 &= 4. \end{aligned}$$

- We can write it in matrix form as

$$\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}.$$

- The inverse of **A** is

$$\mathbf{A}^{-1} = \frac{1}{4} \begin{pmatrix} 2 & -1 \\ -2 & 3 \end{pmatrix} = \begin{pmatrix} 1/2 & -1/4 \\ -1/2 & 3/4 \end{pmatrix}.$$

- Thus

$$\begin{aligned} \mathbf{x}^* &= \begin{pmatrix} 1/2 & -1/4 \\ -1/2 & 3/4 \end{pmatrix} \begin{pmatrix} 5 \\ 4 \end{pmatrix} \\ &= \begin{pmatrix} (1/2) \cdot 5 - (1/4) \cdot 4 \\ -(1/2) \cdot 5 + (3/4) \cdot 4 \end{pmatrix} = \begin{pmatrix} 3/2 \\ 1/2 \end{pmatrix}. \end{aligned}$$

VII.3 Cramer's rule

(Ch. 16.8)

- Cramer's rule is a recipe to find a solution for a system of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n, \end{aligned}$$

where there are n equations in n unknowns and where the matrix representation corresponds to

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

- The recipe is as follows.

Theorem

The system of n equations and n unknowns has a unique solution if and only if $\mathbf{A}$ is nonsingular. The solution is

$$x_1 = \frac{D_1}{|\mathbf{A}|}, \quad x_2 = \frac{D_2}{|\mathbf{A}|}, \quad \dots, \quad x_n = \frac{D_n}{|\mathbf{A}|},$$

where

$$D_j = C_{1j}b_1 + C_{2j}b_2 + \dots + C_{nj}b_n$$

and C_{ij} is the cofactor of a_{ij}.

- The convenient feature of Cramer's rule is that it avoids the computation of $\mathbf{A}^{-1}$.

Example

- Let

$$2x + y = 1, \quad x - y + z = 0, \quad 2y - z = 3.$$

- The determinant of **A** is

$$|\mathbf{A}| = \begin{vmatrix} 2 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 2 & -1 \end{vmatrix} = -1 \neq 0.$$

- Let's calculate D_1 (remember that $D_j = C_{1j}b_1 + C_{2j}b_2 + \ldots + C_{nj}b_n$):

$$\begin{aligned} D_1 &= C_{11}b_1 + C_{21}b_2 + C_{31}b_3 \\ &= \begin{vmatrix} -1 & 1 \\ 2 & -1 \end{vmatrix} \cdot 1 - \begin{vmatrix} 1 & 0 \\ 2 & -1 \end{vmatrix} \cdot 0 + \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} \cdot 3 \\ &= (1-2) \cdot 1 - (-1-0) \cdot 0 + (1-0) \cdot 3 = 2. \end{aligned}$$

- We can do the same kind of calculation to find D_2 and D_3:

$$D_2 = C_{12}b_1 + C_{22}b_2 + C_{32}b_3 = -5,$$

$$D_3 = C_{13}b_1 + C_{23}b_2 + C_{33}b_3 = -7.$$

- Thus, applying Cramer's rule yields

$$x = \frac{D_1}{|\mathbf{A}|} = \frac{2}{-1} = -2, \quad y = \frac{D_2}{|\mathbf{A}|} = \frac{-5}{-1} = 5, \quad z = \frac{D_3}{|\mathbf{A}|} = \frac{-7}{-1} = 7.$$

- Double-check that you get the same result with the solution method involving $\mathbf{A}^{-1}$.

VII.4 Existence and uniqueness of solutions

Definition

A solution of a linear system of equations is an assignment of values to the variables $x_1, x_2, \ldots, x_n$ such that each of the equations is satisfied. The set of all possible solutions is called the solution set.

- Recall our previous example:

$$x_1 + 2x_2 = 2,$$
$$2x_1 - x_2 = 4.$$

- We can solve for x_1 in the first equation and plug it into the second equation:

$$\begin{aligned} x_1 &= 2 - 2x_2 \\ &\Rightarrow 2(2 - 2x_2) - x_2 = 4 \Leftrightarrow x_2 = 0. \end{aligned}$$

Therefore $x_1 = 2 - 2x_2 = 2 - 0 = 2$. Thus the only solution is $(x_1, x_2) = (2, 0)$.

Three possibilities

A linear system may behave in any one of three possible ways:

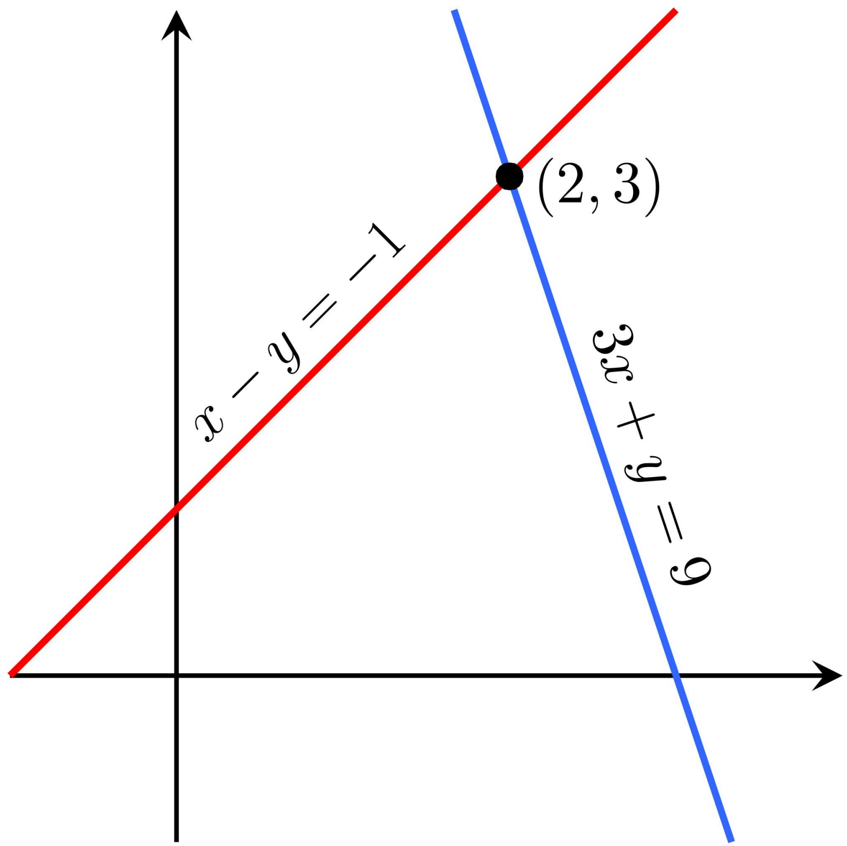
1. The system has infinitely many solutions.
2. The system has a single unique solution.
3. The system has no solution.

A graphical illustration

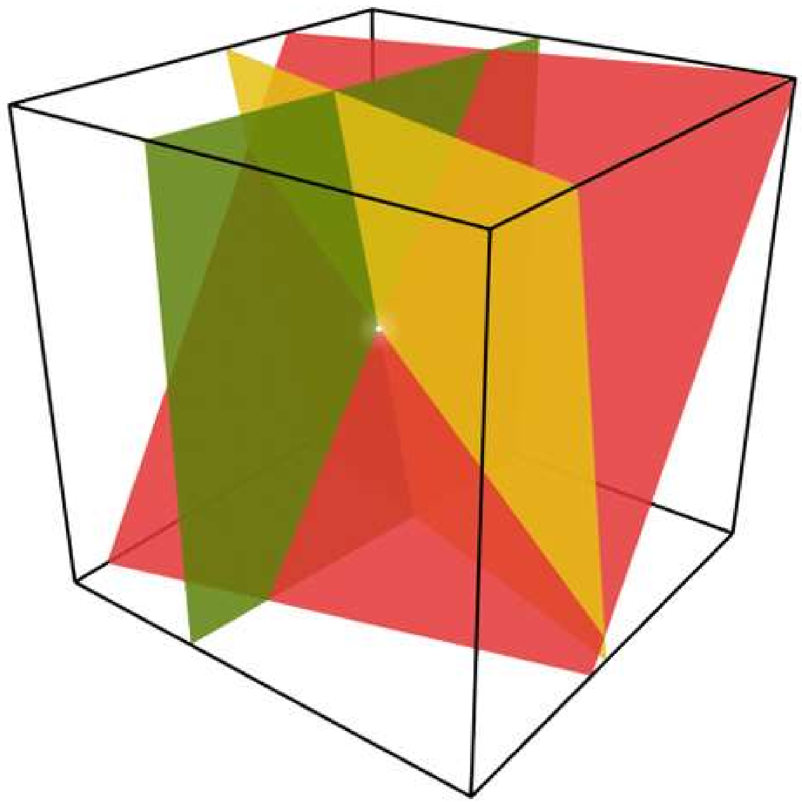
Source: Wikipedia

- For a system involving two variables ($n = 2$), each linear equation determines a line on the two-dimensional plane.
- Because a solution to a linear system must satisfy all of the equations, the solution set is the intersection of these lines. It is therefore either a line (infinitely many solutions), a single point (a unique solution) or an empty set (no solution).

- The following graph is an example of a system in two variables, x and y, with a unique solution.



- A linear system in three variables ($n = 3$) determines a collection of three planes.
- Thus the solution set may be a plane (infinitely many solutions), a line (infinitely many solutions), a single point (a unique solution) or an empty set (no solution).
- The following graph is an example of a system in three variables with a unique solution.



Definition

A system of equations is consistent if it has at least one solution. It is inconsistent otherwise.

- As an example, consider the system

$$\begin{aligned} 2x_1 - 3x_2 &= 1, \\ 4x_1 - 6x_2 &= 1. \end{aligned}$$

- There are two variables in this system and four coefficients.
- However, this system does not have a solution. In particular, note that $(x_1, x_2) = (0, 0)$ would not be a solution.
- This system is therefore inconsistent.
- Graphically, this system would be represented by two parallel lines.
- Note that the associated coefficient matrix $\mathbf{A}$ would be singular because its second row is twice the first row.

VII.5 Linear dependence and the rank of a matrix

Definition

The equations of a system are linearly independent if none of the equations can be derived algebraically from the others. If an equation can be derived from other equations of the system, these equations are linearly dependent.

- When the equations are linearly independent, each equation contains new information about the variables.
  - For example, the equations

$$\begin{aligned} 3x_1 + 2x_2 &= 6, \\ 6x_1 + 4x_2 &= 12 \end{aligned}$$

    are not independent. If scaled by a factor of two, the first equation is the same as the second. The two equations would produce identical graphs. The two rows of the corresponding coefficient matrix $\mathbf{A}$ would therefore not be independent, and $\mathbf{A}$ would be singular.
  - In contrast to the example on the previous slide, the solution to this system would be a line (infinitely many solutions) because graphically, the two equations would be the same (not parallel lines).

- As a more complicated example, the equations

$$\begin{aligned} x_1 - 2x_2 &= -1, \\ 3x_1 + 5x_2 &= 8, \\ 4x_1 + 3x_2 &= 7 \end{aligned}$$

are not independent because the third equation is the sum of the other two.

  - Indeed, any one of these equations can be derived from the other two, and any one of the equations can be removed without affecting the solution set.
  - We therefore say that one of the three equations is *redundant*. That is, no information would be lost by removing one equation.

- The important insight for the solution of systems of n linear equations in n variables is as follows.
  - If all equations of the system are linearly independent, then this is a necessary and sufficient condition for the *nonsingularity* of the corresponding coefficient matrix $\mathbf{A}$, where $\mathbf{A}$ is a square matrix of dimension $(n \times n)$.
  - Recall that a square matrix is nonsingular if $|\mathbf{A}| \neq 0$. Also recall that a matrix $\mathbf{A}$ has an inverse $\mathbf{A}^{-1}$ if and only if $\mathbf{A}$ is nonsingular.
  - Since we can solve the system $\mathbf{Ax} = \mathbf{b}$ by matrix inversion as long as $\mathbf{A}^{-1}$ exists, for the purpose of solving such systems the concept of linear independence is therefore very important.

The rank of a matrix

- Even though we have only discussed the concept of independence with regard to square matrices, it is equally applicable to any $(m \times n)$ matrix where $m \neq n$.
- If the maximum number of linearly independent rows that can be found in such a matrix is r, the matrix is said to be of rank r. (The rank also tells us the maximum number of linearly independent *columns* in that matrix.)
- The rank of an $(m \times n)$ matrix can be at most m or n, whichever is smaller.
- By definition, an $(n \times n)$ nonsingular square matrix $\mathbf{A}$ has n linearly independent rows (or columns). Consequently it must be of rank n. Conversely, an $(n \times n)$ matrix having rank n must be nonsingular.

- As an example, consider the (4×4) matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 1 & 3 \\ -1 & -2 & 1 & 0 \\ 0 & 0 & 2 & 2 \\ 3 & 6 & 2 & 5 \end{bmatrix}.$$

- Instead of the rows let's consider the columns.
- You can see that the second column is twice the first column. The first and second columns are thus linearly dependent.
- In addition, the fourth column equals the sum of the first and the third. The first, third and fourth columns are thus linearly dependent.
- Therefore, we can pick at most two columns that are linearly independent together. So the rank of $\mathbf{A}$ is $r = 2$.
  - Note that it is not always obvious to spot linear dependence by simple visual inspection.
  - The rank of a matrix can be established more formally with the *Gaussian elimination method*. We will not cover this method here.

Topic VIII: Interest rates, present value and difference equations

EC123 Mathematical Techniques B

Dr Dennis Novy

University of Warwick

Autumn Term 2015

Topic VIII: Interest rates, present value and difference equations

Topic VIII's big picture:

- We now study some basic aspects in the mathematics of finance and the economics of growth.
- We will also come across the notion of difference equations.
- We will explicitly consider the movement of variables *over time*. We call this dynamic analysis (as opposed to static analysis).
- The topic outline is as follows:

VIII.1 Interest rates

VIII.2 Continuous compounding

VIII.3 Present value

VIII.4 Difference equations

VIII.1 Interest rates
(Ch. 10.1)

- Most people have a savings account, credit card debt or mortgage debt.
- A crucial feature of all these financial products is the interest rate payment which you receive over a given period (in case you own financial assets) or which you have to pay (in case you have debt).
- Suppose a principal (or capital) of S_0 yields interest at the rate of p percent per period (for example one year).
- After t periods it will have increased to the amount

$$S_t = S_0(1+r)^t,$$

where $r = p/100$ is the interest rate.

- The formula assumes that the interest is added to the principal at the end of each period (this process is called *interest compounding*, see below).

Example 1

- A deposit of £5000 is put into a savings account earning interest at the annual rate of 9 percent, with interest paid quarterly.
- Question: How much will there be in the account after eight years?
- Answer: The so-called periodic rate is the annual interest rate divided by the number of interest periods per year. It is therefore given by $r/n = 0.09/4 = 0.0225$ in this example, where n is the number of periods per year and 32 is the total number of periods.
- Our above formula then gives

$$5000\,(1 + 0.0225)^{32} \approx 10190.52.$$

- The initial deposit will have more than doubled.

Example 2

- How long will it take for the £5000 in the previous example to increase to £15000?
- We know that after t quarterly payments, the account will grow to $5000\left(1+0.0225\right)^{t}$. Thus

$$\begin{aligned} 5000\left(1+0.0225\right)^{t} &= 15000, \text{ or} \\ 1.0225^{t} &= 3. \end{aligned}$$

- To find t we take the natural logarithm of each side:

$$\begin{aligned} t\ln\left(1.0225\right) &= \ln(3) \quad \text{because} \quad \ln\left(a^{b}\right) = b\ln(a) \\ \Leftrightarrow \quad & t = \frac{\ln(3)}{\ln(1.0225)} \approx 49.37. \end{aligned}$$

- Thus it takes about 49.37 quarterly periods, i.e. a bit over 12 years, before the value of the account increases to £15000.

VIII.2 Continuous compounding

(Ch. 10.2)

- Compound interest arises when interest is added to the principal so that from that moment on, the interest that has been added also itself earns interest. This addition of interest to the principal is called compounding.
- For example, a bank account with an initial principal of £1000 and 20 percent interest per year would have a balance of
  - £1200 at the end of the first year ($1200 = 1000(1.2)^1$),
  - £1440 at the end of the second year ($1440 = 1200(1.2)^1 = 1000(1.2)^2$),
  - £1728 at the end of the third year ($1728 = 1440(1.2)^1 = 1000(1.2)^3$) and so on.
- Compound interest may be contrasted with *simple interest*, where interest is not added to the principal (there is no compounding).
- Compound interest is standard in finance and economics, and simple interest is used infrequently.

- Above we considered an example where interest payments were added to the principal at the end of every year. We also considered an example where interest payments were added at the end of every quarter.
- What happens if interest payments are added much more frequently, say, every week, day, hour, minute, second, millisecond etc.?
- In the limit where the time period becomes infinitesimal, we call this process continuous compounding of interest.
- The formula

$$S_t = S_0 e^{rt} = S_0 \exp(rt)$$

  shows how much a principal S_0 will have increased after t years if the annual interest rate is r and if there is continuous compounding of interest. Note that $\exp(rt)$ is just different notation for e^{rt}.
- By the way, this formula is very general. It also applies to other growth processes, for example population growth or GDP growth, where S_0 would be the initial level and r would be the growth rate.

- Suppose the sum of £5000 is invested given an annual interest rate of 9 percent. What is the balance after eight years with continuous compounding?
    - Answer:
$$5000e^{0.09 \cdot 8} = 5000e^{0.72} \approx 10272.$$
    - Note that it is $0.09 \cdot 8$ in the exponent, not $1.09 \cdot 8$.
- Here is an example with *negative* growth. The population of Japan is currently about 127 million and it shrinks by about 0.28 percent every year. How big will the population be in 20 years from now?
    - Answer:
$$127,000,000e^{-0.0028 \cdot 20} = 127,000,000e^{-0.056} \approx 120,083,470.$$
    - Thus, the population is projected to shrink by about 7 million over that period.

Growth rates

- Note that if $S_t = S_0 e^{rt}$, then
$$S_t' = S_0 r e^{rt} = r S_t.$$
- It follows
$$\frac{S_t'}{S_t} = r,$$
where S_t' / S_t is the growth rate of S_t (in continuous notation).
- The growth rate of the principal is therefore constant at r.
- By the way, in discrete notation the growth rate would be written as $\Delta S / S$.

VIII.3 Present value

(Ch. 10.3)

- The sum of £1000 in your hands today is worth more than £1000 at some future date.
- The reason is that you can invest the £1000 today and start earning interest payments so that in the future you will have more than £1000.
- Suppose the interest rate is 11 percent. Then
  - after one year you will own £1000$(1.11) =$£1110,
  - after six years you will own £1000$(1.11)^6 \approx$£1870 etc.
- It follows that given the interest rate of 11 percent, £1000 today has the same value as £1110 next year or £1870 in six years.

- Expressed differently, we can say that the present value of £1110 due for payment in one year's time from now is £1000.
- Since £1000 is less than £1110, we also speak of £1000 as the present discounted value (or PDV) of £1110 next year.
- The ratio £1000/£1110 $= 1/1.11 \approx 0.9009$ is called the discount factor, whose reciprocal 1.11 is one plus the discount rate. Thus the discount rate is equal to the interest rate of 11 percent.

A more general formula

- Suppose an amount K is due for payment in t years after the present date.
- What is the present value A when the interest rate is r? In other words, how much money must be deposited today in order to have the amount K after t years?
- If interest is paid annually, then
  - an amount A will have increased to $A(1+r)^t$ after t years so that we need $A(1+r)^t = K$.
  - Thus $A = K(1+r)^{-t}$.
  - Here the annual discount factor is $(1+r)^{-1}$, and $(1+r)^{-t}$ is the discount factor appropriate for t years.
- If interest is compounded continuously, then
  - an amount A will have increased to Ae^{rt} after t years so that we need $Ae^{rt} = K$.
  - Thus $A = Ke^{-rt}$.

Example

- Find the present value of \$100,000 which is due for payment in 15 years if the interest rate is 6 percent compounded (*i*) annually, (*ii*) continuously.
- Solution:

  (i) The present value is $\$100{,}000(1+0.06)^{-15} \approx \$41{,}726.51$.

  (ii) The present value is $\$100{,}000e^{-0.06\cdot 15} \approx \$40{,}656.97$.

- As expected, the present value with continuous compounding is smaller because given a certain interest rate, capital increases most rapidly in that case.

VIII.4 Difference equations
(Ch. 10.8)

- Many variables that economists study such as income, consumption and savings are recorded at fixed time intervals, for example each quarter or each year.
- We are often interested in modelling how these variable change over time. That is, we are interested in *dynamic* analysis.
- Equations that relate such variables at different discrete moments of time are called difference equations.
- Their continuous counterparts are called *differential equations* but we will not consider those here.

Definition

Let $t = 0, 1, 2, \ldots$ denote different discrete time periods or moments of time. We call $t = 0$ the **initial period**. If $x(t)$ is a function defined for $t = 0, 1, 2, \ldots$ we use $x_0, x_1, x_2, \ldots$ to denote the values of $x(t)$ at $t = 0, 1, 2, \ldots$

A simple example

- A simple example of a *first-order difference equation* is

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

  where a is a constant.
- This is a first-order equation because it relates the value of a function in period $t+1$ to the value of the function in only the previous period t (as opposed to also period $t-1$ etc.)
- Suppose x_0 is given. Repeatedly applying the previous equation gives

$$\begin{aligned} x_1 &= ax_0, \\ x_2 &= ax_1 = a^2 x_0, \\ x_3 &= ax_2 = a^2 x_1 = a^3 x_0 \text{ and so on.} \end{aligned}$$

- In general, we can therefore write

$$x_t = x_0 a^t.$$

Another example

- Let K_t denote the balance in an account at the beginning of period t given an interest rate r.
- Then the balance at time $t+1$ is $K_{t+1} = K_t + rK_t = (1+r)K_t$.
- Hence, K_t satisfies the difference equation

$$K_t = K_0(1+r)^t.$$

- We already know this from the beginning of this topic.

Example: A multiplier-accelerator model of economic growth

- This is the first *dynamic* macroeconomic model we consider in this module.
- Let Y_t denote national income, I_t total investment and S_t total savings.
- Suppose that savings are proportional to national income, and that investment is proportional to the change in income from period t to $t+1$. Then we can write

$$\begin{array}{ll} (i) & S_t = \alpha Y_t, \\ (ii) & I_{t+1} = \beta(Y_{t+1} - Y_t), \\ (iii) & S_t = I_t. \end{array}$$

- The last equation is an equilibrium condition, stating that savings have to equal investment in each period.
- Here α and β are positive constants and we assume $\beta > \alpha > 0$.
- Now deduce a difference equation determining the path of Y_t, given Y_0, and solve it.

Solution to the model

- From (i) and (iii), $I_t = \alpha Y_t$, and so $I_{t+1} = \alpha Y_{t+1}$.
- Inserting this into (ii) yields $\alpha Y_{t+1} = \beta(Y_{t+1} - Y_t)$, or $(\alpha - \beta) Y_{t+1} = -\beta Y_t$.
- Thus

$$Y_{t+1} = \frac{\beta}{\beta - \alpha} Y_t = \left(1 + \frac{\alpha}{\beta - \alpha}\right) Y_t.$$

- Using our general expression for first-order difference equations yields the solution

$$Y_t = \left(1 + \frac{\alpha}{\beta - \alpha}\right)^t Y_0.$$

- As a simulation exercise at home, you can plug in different values for α, β and Y_0 to see how Y_t evolves over time. You can do this in any standard spreadsheet programme such as Microsoft Excel.