Math notes

Pietro Battiston

November 15, 2011

Those are the notes I took during the DEFAP [1] course of Mathematics tought by prof. Weinrich in November-December 2010.

They're certainly full of mistakes. I guess the reader will be mature enough to not attribute them prof. Weinrich, and kind enough to point me[2] at them, so that I can fix them.

---

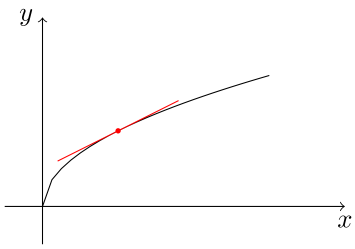
[1] http://scuoledidottorato.unicatt.it/defap
[2] me@pietrobattiston.it

02/11/2010

1 Analysis in the Euclidean space

1.1 Derivatives

Given a function f from $U \subset \mathbb{R}$ (with U, the *domain*, *open*) to $\mathbb{R}$, we can draw its graph and imagine the tangent in some point.



How to calculate how much *"steep"* is f in some point? It's the limit

$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = f'(x) = \frac{df}{dx}(x)$$

(if this limit exists). This is the *derivative*, the slope of the tangent line at the point $(x, f(x))$.

The derivative may exist only for some x, or for all x in the domain: if the latter holds, we say f is *differentiable*: $\forall x \in \operatorname{dom} f\ \exists f'(x)$.

Consequences: f is *increasing* (resp. *decreasing*) if $f'(x) > 0$ (resp. < 0).

It may be that f is neither increasing nor decreasing in some point x: this means $f'(x) = 0$. This is intuitive if we look at the graph.

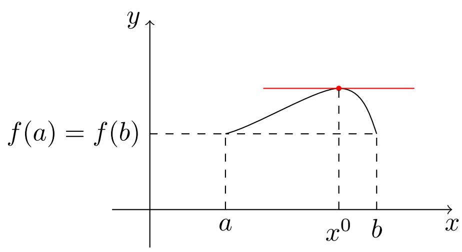
We say $f(x^*)$ is a *relative maximum* if and only if $\exists \varepsilon > 0$ such that

$$f(x^*) \geq f(x)\ \forall x \in (x^* - \varepsilon, x^* + \varepsilon) =: \mathcal{I}_{x_0,\varepsilon}.$$

$\mathcal{I}_{x_0,\varepsilon}$ is called a *neighborhood* of x_0.

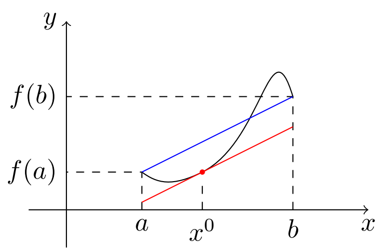
If f is differentiable, this means $f'(x^*) = 0$.

Theorem 1 (Rolle's Theorem)**.** *Left $f : [a,b] \to \mathbb{R}$ continuous on $[a,b]$ and differentiable on (a,b). If $f(a) = f(b)$, then there is a point $x^0 \in (a,b)$ such that $f'\left(x^0\right) = 0$.*



Theorem 2 (Mean-value Theorem)**.** *Suppose that $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ and differentiable on (a,b). Then, there exists $x^0 \in (a,b)$ such that*

$$f(b) - f(a) = f'\left(x^0\right)(b-a).$$



Now consider $f : U \in \mathbb{R}$, $U \subset \mathbb{R}^n$ where

$$\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}.$$

Again, $U = \mathrm{dom}(f)$.
Let $e^i := (\underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{n-i})$. This is called an *elementary vector*.
Given any $x^0 \in \mathrm{dom}\, f \subset \mathbb{R}^n$, consider

$$\frac{\partial f}{\partial x_i}\left(x^0\right) := \lim_{h \to 0} \frac{f(x^0 + he^i) - f\left(x^0\right)}{h};$$

this (if it exists) is called the *partial derivative* of f with respect to x_i.

Remark 3. $\frac{\partial f}{\partial x_i} : \mathrm{dom}\, f \to \mathbb{R}$ *since it associates to any x the value* $\frac{\partial f}{\partial x_i}(x)$.

Since we can do that for any i (for which partial derivatives exist), we obtain n new functions.

Example 4. 1. *Take $f(x_1, x_2) = Ax_1^\alpha x_2^\beta$ with $A, \alpha, \beta > 0$ (a* Cobb-Douglas *function).*

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \alpha A x_1^{\alpha-1} x_2^\beta$$

which (for $x_1, x_2 \neq 0$)[3] *is equal to*

$$\alpha \frac{Ax_1^\alpha x_2^\beta}{x_1} = \alpha \frac{f(x_1, x_2)}{x_1}.$$

Symmetrically,

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

2. *$f(x_1, x_2) = \min\{ax_1, bx_2\}$ with $a, b > 0, x_1, x_2 > 0$ (a* Leontief *function).*
We have to distinguish some cases:

- *$ax_1 < bx_2$:*

$$a(x_1 + h) < bx_2 \text{ for } h \text{ small enough.}$$

So

$$\frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h} = \frac{a(x_1 + h) - ax_1}{h} = a.$$

[3] Which is usually the case in common economic interpretations

- $ax_1 > bx_2$:
  similarly,

$$a(x_1 + h) > bx_2 \text{ for } |h| \text{ small enough.}$$

  So

$$\frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h} = \frac{bx_2 - bx_2}{h} = 0.$$

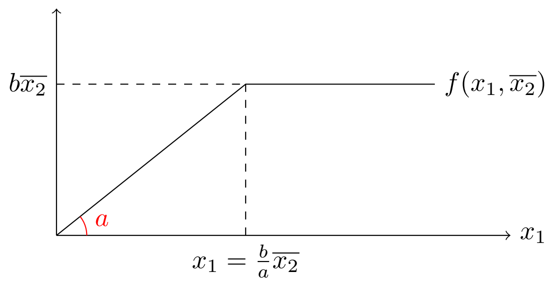
- $ax_1 = bx_2$:

$$\frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h} = \begin{cases} \frac{bx_2 - bx_2}{h} = 0 & \text{if } h > 0 \\ \frac{a(x_1+h) - ax_1}{h} = a & \text{if } h < 0. \end{cases}$$

  As a consequence, if we look at

$$\lim_{h \to 0} \frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h}$$

  the result depends on the sign of h... in other words, this limit does not *exist. So $\frac{\partial f}{\partial x_1}(ax_1, bx_2)$ does not exist in the point $\left(x_1, \frac{a}{b}x_1\right)$.*



1.2 Differential

Take $f : U \to \mathbb{R}$, $U \subset \mathbb{R}$, differentiable.

Example 5. *$f(x) = \frac{1}{2}\sqrt{x} = \frac{1}{2}x^{\frac{1}{2}}$ taken as a production function (it has all the typical properties).*

$f(100) = \frac{1}{2}10 = 5$. What if the inputs are increased by 1?
$f(101) = 5.02494\ldots$, so the increase is $0.02595\ldots$
In general, $f'(x) = \frac{1}{4}x^{-\frac{1}{2}}$. So $f'(100) = \frac{1}{4} \cdot \frac{1}{10} = 0.025$.
Not the same, but very similar. That's why economists talk about marginal productivity *of an input, thought as increase of output corresponding to an increase of input by 1 unit.*

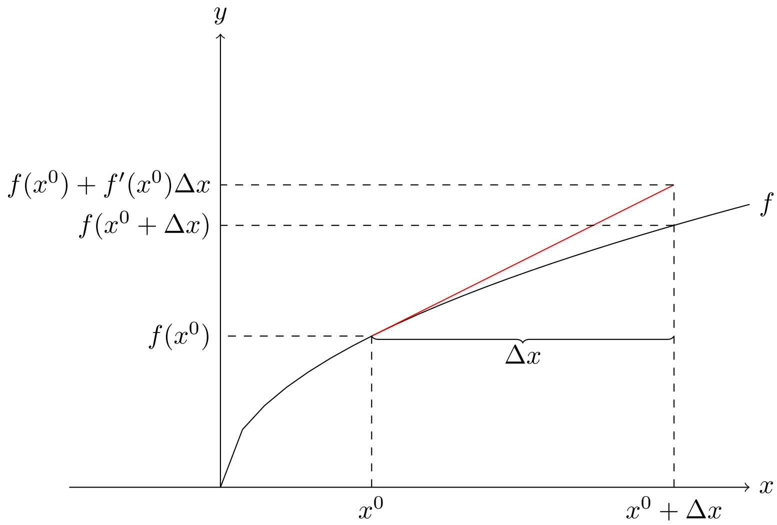
In general, we may consider $h = \Delta x \Rightarrow \Delta y := f(x^0 + \Delta x) - f\left(x^0\right)$: for small values of Δx,

$$\frac{f(x^0 + \Delta x) - f\left(x^0\right)}{\Delta x} \approx f'\left(x^0\right)$$

or

$$f(x^0 + \Delta x) \approx f\left(x^0\right) + f'\left(x^0\right)\Delta x$$

(if we know the "old" value and the derivative in it, we can "forecast" the "new" one).



So I can build the function:

$$x \mapsto f\left(x^0\right) + f'\left(x^0\right)\left(x - x^0\right)$$

which is the equation of the tangent line T at the graph of f in $\left(x^0, f\left(x^0\right)\right)$.

03/11/2010

Let's introduce a new notation:

$$dy = \text{ change of } y \text{ along the tangent line } T.$$

And let's write $dx = \Delta x$. Then,

$$dy = f'\left(x^0\right)dx,$$

and this is a linear function with variables dx, dy ($=$ *differentials*) with origin $\left(x^0, f\left(x^0\right)\right)$.

The smaller the Δx, the bigger the precision of this linear approximation.

So far, we studied the situation for functions in 1 variable. Let's generalize this to higher dimensions.

Given $f : U \to \mathbb{R}$, with $U \subset \mathbb{R}^2$, for $x^0 = (x_1^0, x_2^0) \in U$, we may consider

$$f\left(x_1^0 + \Delta x_1, x_2^0 + \Delta x_2\right)$$
$$\approx f\left(x_1^0, x_2^0\right) + \frac{\partial f}{\partial x_1}\left(x_1^0, x_2^0\right)\Delta x_1 + \frac{\partial f}{\partial x_2}\left(x_1^0, x_2^0\right)\Delta x_2$$

and

$$(x_1, x_2) \overset{T}{\mapsto} f(x_1^0 + x_2^0) + \frac{\partial f}{\partial x_1}(x_1^0, x_2^0)(x_1 - x_1^0) + \frac{\partial f}{\partial x_2}(x_1^0, x_2^0)(x_2 - x_2^0)$$

is the equation of the function describing the two-dimensional plane T tangent to the graph of f at $(x_1^0, x_2^0, f(x_1^0, x_2^0))$, where

$$\mathrm{graph}(f) = \{(x_1, x_2, f(x_1, x_2)) | (x_1, x_2) \in U\} \subset \mathbb{R}^3.$$

This is the generalization of the tangent line to a function in 2 variables.

For $\Delta y = f(x_1, x_2) - f(x_1^0, x_2^0)$, the expression

$$T(x_1, x_2) - f(x_1^0, x_2^0) = \frac{\partial f}{\partial x_1}(x_1^0, x_2^0)(x_1 - x_1^0) + \frac{\partial f}{\partial x_2}(x_1^0, x_2^0)(x_2 - x_2^0)$$

is a linear approximation around x^0. In this case too we can introduce differentials:

$$df = dy = \frac{\partial f}{\partial x_1}(x_1^0, x_2^0)dx_1 + \frac{\partial f}{\partial x_2}(x_1^0, x_2^0)dx_2$$

is the same linear approximation of Δy around x^0.

In general, we consider a function

$$f(x_1^0 + \Delta x_1, \dots, x_n^0 + \Delta x_n) \approx f(x_1^0, \dots, x_n^0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_1^0, \dots, x_n^0)\Delta x_i$$

and the function

$$(x_1, \dots, x_n) \overset{T}{\mapsto} f(x_1^0, \dots, x_n^0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_1^0, \dots, x_n^0)(x_i - x_i^0)$$

describes the n-dimensional *tangent hyperplane* to

$$\mathrm{graph}(f) = \{(x_1, \dots, x_n, f(x_1, \dots, x_n)) | (x_1, \dots, x_n) \in U\} \subset \mathbb{R}^{n+1}$$

at the point $(x^0, f(x_0))$.

We can then look at

$$\begin{aligned} df &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0)dx_i \\ &= \left(\frac{\partial f}{\partial x_1}\left(x^0\right), \dots, \frac{\partial f}{\partial x^n}\left(x^0\right) \right) \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} \\ &= Df\left(x^0\right) dx, \end{aligned}$$

which is called the *total differential* of f.

$$Df\left(x^0\right) = \left(\frac{\partial f}{\partial x_1}\left(x^0\right), \dots, \frac{\partial f}{\partial x_n}\left(x^0\right) \right)$$

is the *Jacobian derivative* of f at x^0 and df, dx_1, $\dots$, dx_n are differentials with

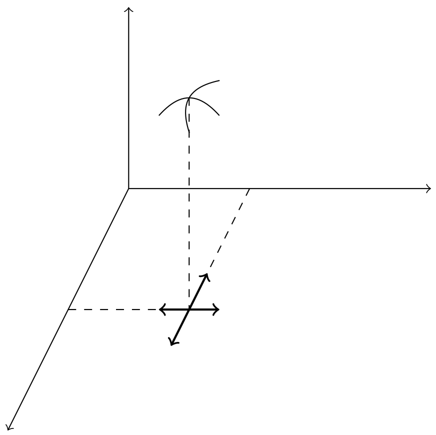
$$dx = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}$$

$df = Df\left(x^0\right) dx$ is a linear approximation of
$\Delta y = f(x) - f\left(x^0\right)$ around x^0.
We can now generalize in another direction...

1.3 Directional derivatives and gradients

Let's recall that

$$\frac{\partial f}{\partial x_1}\left(x_1^0, x_2^0\right) = \lim_{h \to 0} \frac{f\left(x_1^0 + h, x_2^0\right) - f\left(x_1^0, x_2^0\right)}{h} :$$



Definition 6. *A* curve *in $\mathbb{R}^n$ is an n-uple of continuous functions*

$$x(t) = (x_1(t), \dots, x_n(t))$$

with $x_i : I \to \mathbb{R}$ for all $i = 1, \dots, n$ and $I \subset \mathbb{R}$ an interval.

$x_i(t)$ are the coordinate functions and t the parameter *describing the curve. If t is* time, *then $x(t) = (x_1(t), \dots, x_n(t))$ are the coordinates of the point at time t.*

Example 7. $x_i(t) = t$, $0 \leq t \leq 1$, $i = 1, 2 \Rightarrow \{x(t) | t \in [0,1]\} = \{(t,t) | 0 \leq t \leq 1\}$

Let $x(t)$ be a curve in $\mathbb{R}^n$. Consider a sequence $\{h_j\}_{j=1}^{\infty}$ in $\mathbb{R}$ such that $h_j \to 0$ for $j \to \infty$.

Then, given $t_0 \in I$, this induces another sequence: $\{x(t_0 + h_j)\}_{j=1}^{\infty}$ in $\mathbb{R}^n$.

Now, consider

$$\left(\lim_{h_j \to 0} \frac{x_1(t_0 + h_j) - x_1(t_0)}{h_j}, \dots, \lim_{h_j \to 0} \frac{x_n(t_0 + h_j) - x_n(t_0)}{h_j}\right);$$

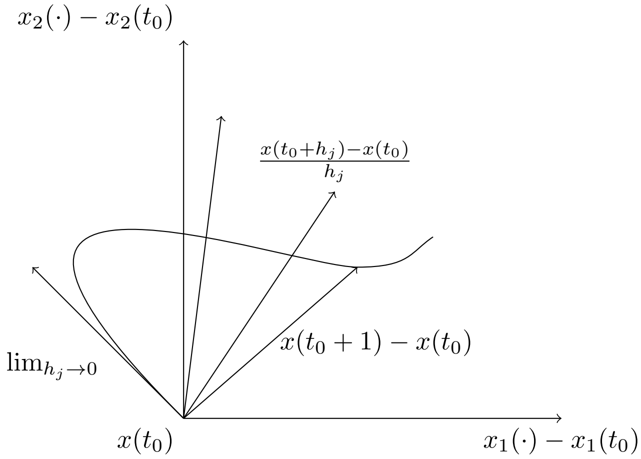
those limits are simply the derivatives:

$$(x_1'(t_0), \dots, x_n'(t_0));$$

this object is called the *velocity* vector of the curve at t_0, and $x_i'(t_0)$ is the *instantaneous velocity* of the i-th coordinate along the curve at t_0.

How can we draw a velocity vector?

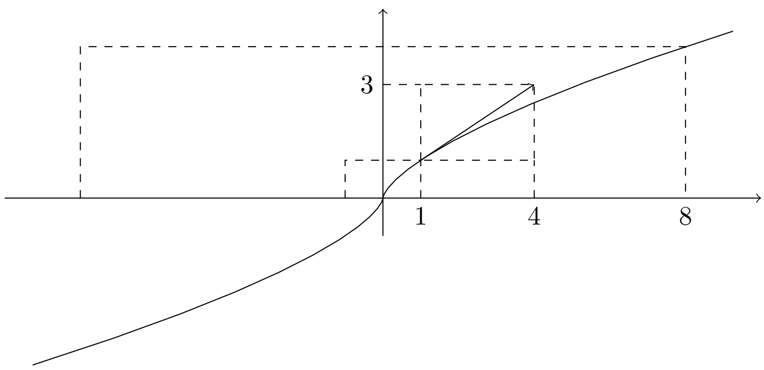
$\frac{x(t_0+h_j)-x(t_0)}{h_j} =$ lenghtening of $x(t_0 + h_j) - x(t_0)$ whenever $h_j < 1$.



we can see that $x'(t_0)$ is a *tangent vector* (=velocity vector) to the curve at $t = t_0$.

Example 8. $x(t) = (t^3, t^2)$, $t_0 = 0$. *Then:*

- *x(0) = (0,0)*
- *x(1) = (1,1)*
- *x(2) = (8,4)*
- *x(-1) = (-1,1) . . .*



$x'(t) = (3t^2, 2t) \Rightarrow x'(1) = (3, 2)$.
More in general, the slope of the tangent vector is: $\frac{2t}{3t^2} = \frac{2}{3t}$.
We can verify it goes to ∞ *for* $t \to 0$.
In fact, $x'(0) = (0, 0)$*, which is not a "tangent vector"... we have a* cusp.

Definition 9. *A curve* $(x_1(t), \ldots, x_n(t))$ *is* regular *if each* $x_i'(t)$ *is continuous in* t *and*

$$(x_1'(t), \ldots, x_n'(t)) \neq (0, \ldots, 0)$$

Remark 10. *Cusps (null vectors) are not necessarily associated to infinite slope.*

04/11/2010

We can now come back to the previous question: how does a function $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$, behave along a curve $x(t)$, with $t \in I$?

What we want to study is the composition

$$\begin{array}{ccccc} \mathbb{R} & & \mathbb{R}^n & & \\ \cup & & \cup & & \\ I & \longrightarrow & U & \longrightarrow & \mathbb{R} \\ \Downarrow & & \Downarrow & & \Downarrow \\ t & \longmapsto & x(t) & \longmapsto & f(x(t)) \end{array}$$

We hence can create $g(t) = f(x(t))$, $t \in I$.

What is $g'(t)$?

In the case $n = 1$, it's easy:

$$\begin{aligned} g'(t) =& \frac{df(x(t))}{dt} \\ =& f'(x(t))x'(t) \end{aligned}$$

(the ordinary formula for functions composition - a.k.a. the *"chain rule"*).

If instead $n > 1$,

$$g'(t) = \frac{\partial f}{\partial x_1}(x(t))x_1'(t) + \cdots + \frac{\partial f}{dx_n}(x(t))x_n'(t).$$

This expression seems more cumbersome than the one-dimensional case. So we'll rewrite it as follows:

$$\begin{aligned} g'(t) =& \left(\frac{\partial f}{\partial x_1}(x(t)), \ldots, \frac{\partial f}{\partial x_n}(x(t))\right) \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix} \\ =& Df(x(t))x'(t); \end{aligned}$$

this is the "new" chain rule: *"chain rule number I"*.[4]

Example 11. $f(x, y) = x^2 + y^2$[5]

$x(t) = y(t) = t$ *(the straight* $45°$ *line):*

$$\{(x(t), y(t)) | t \in \mathbb{R}\} = \{(t, t) | t \in \mathbb{R}\}.$$

Obviously, $x'(t) = 1, y'(t) = 1 \Rightarrow g(t) = f(x(t), y(t)) = t^2 + t^2 = 2t^2 \Rightarrow g'(t) = 4t.$

Let's verify the chain rule yields the same results:

[4] We'll later have number II, III, and IV also.

[5] In 2 dimensions, using x and y is simpler than using subscripts.

$$\begin{aligned} g'(t) =& \frac{\partial f}{x}(x(t), y(t)) \cdot 1 + \frac{\partial f}{\partial y}(x(t), y(t)) \cdot 1 \\ =& 2x(t) + 2y(t) \\ =& 2t + 2t = 4t \end{aligned}$$

Let's now follow a generalization.
If

$$\begin{array}{ccccc} x & : & \mathbb{R}^s & \longrightarrow & \mathbb{R}^n \\ & & \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\ & & t = (t_1, \ldots, t_s) & \longmapsto & (x_1(t), \ldots, x_n(t)) \end{array}$$

We can hence define, for a given function $f : \mathbb{R}^n \to \mathbb{R}$,

$$g(t_1, \ldots, t_s) = f(x_1(t_1, \ldots, t_s), \ldots, x_n(t_1, \ldots, t_s))$$

$$\begin{tikzcd} \mathbb{R}^s \arrow[r, "x(\cdot)"] \arrow[rr, bend right, "g"'] & \mathbb{R}^n \arrow[r, "f"] & \mathbb{R} \end{tikzcd}$$

Now, given any $i = 1, \ldots, s$,

$$\frac{\partial g}{\partial t_i}(t) = \frac{\partial f}{\partial x_1}(x(t))\frac{\partial x_1}{\partial t_i}(t) + \cdots + \frac{\partial f}{\partial x_n}(x(t))\frac{\partial x_n}{\partial t_i}(t);$$

since t_i is the only thing I'm varying, I get the result as a function of just t_i.
Then, I get

$$Dg(t) = \left(\frac{\partial g}{\partial t_1}(t), \ldots, \frac{\partial g}{\partial t_s}(t)\right);$$

this is the *"chain rule number II"*.

Example 12. *Let Q be the capital, $Q = 4K^{\frac{3}{4}}L^{\frac{1}{4}}$ the production function.*
Let K and L vary in time t and in values of the interest rate r according to:

$$K(t, r) = \frac{10t^2}{r}, \qquad L(t, r) = 6t^2 + \underbrace{250r}_{\textit{substitution effect, i.e.}}$$

$$\Rightarrow Q(t, r) = 4\left(\frac{10t^2}{r}\right)^{\frac{3}{4}}\left(6t^2 + 250r\right)^{\frac{1}{4}}$$

We want to calculate the rate of change of Q with respect to t when $t = 10$ and $r = 0.1$.

According to the chain rule number II, since we have 2 variables (K and L), we have two components

$$\frac{\partial Q}{\partial t} = \frac{\partial Q}{\partial K} \cdot \frac{\partial K}{\partial t} + \frac{\partial Q}{\partial L} \cdot \frac{\partial L}{\partial t}.$$

Notice that, differently from above, t is a scalar, not a vector. Economists are often sloppy and use the same letters with different uses (e.g. Q is both the dependent variable and *the function expressing it), and we must capable to give the right interpretation to each object.*[6]

Now we can write

$$K(10, 0.1) = \frac{10 \cdot 100}{0.1} = 10000$$

and

$$L(10, 0.1) = 6 \cdot 100 + 25 = 625.$$

Let's calculate partial derivatives:

$$\frac{\partial Q}{\partial K} = 3K^{-\frac{1}{4}}L^{\frac{1}{4}} = 3\left(\frac{L}{K}\right)^{\frac{1}{4}}.$$

We can now insert those values into

$$\frac{\partial Q}{\partial K} = 3\left(\frac{625}{10000}\right)^{\frac{1}{4}} = 3 \cdot \frac{5}{10} = 1.5,$$

$$\frac{\partial Q}{\partial K} = \frac{1}{4} \cdot 4K^{\frac{3}{4}}L^{-\frac{3}{4}} = \left(\frac{K}{L}\right)^{\frac{3}{4}} = \left(\frac{10}{5}\right)^{3} = 8.$$

All is left to calculate is

$$\frac{\partial K}{\partial t} = \frac{20t}{r} = \frac{200}{0.1} = 2000,$$

$$\frac{\partial L}{\partial t} = 12t = 120.$$

Finally, putting everything together:

$$\frac{\partial Q}{\partial t} = 1.5 \cdot 2000 + 8 \cdot 120 = 3000 + 960 = 3960.$$

What is $Q(10000, 625)$, the amount that the firm can produce? It's $4 \cdot 1000 \cdot 5 = 20000$.

If we measure t in years, for instance, we can say that output incresases of 3960 *units in an year (starting from the given values).*

Since in this case we have constant returns to scale[7]*, this is not even an approximation: this is the real value.*

We could *also do the approximation as yesterday: we would take*

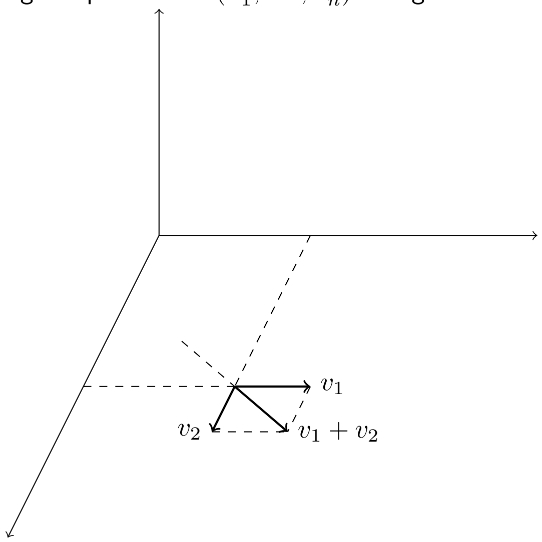
$$\begin{aligned} dQ =& \frac{\partial Q}{\partial t} dt \\ =& \left(\frac{\partial Q}{\partial K}\frac{\partial K}{\partial t} + \frac{\partial Q}{\partial L}\frac{\partial L}{\partial t}\right) dt \end{aligned}$$

and with $dt = 1$ we would get exactly the same result.

[6] Unless what is written is wrong - and that happens.

[7] Q is linear in t.

A step back: we want to calculate the rate of change of a function $f(x_1, \dots, x_n)$ at a given point $x^0 = (x_1^0, \dots, x_n^0)$ in a given direction $v = (v_1, \dots, v_n)$.



We can then define the curve $x(t) = x^0 + tv$, $t \in R$, and look at

$$g(t) = f(x(t)) = f(x^0 + tv) = f(\underbrace{x_1^0 + tv_1}_{x_1(t)}, \dots, \underbrace{x_n^0 + tv_n}_{x_n(t)})$$

$$\Rightarrow g'(t) = \frac{\partial f}{\partial x_1}(x^0 + tv)v_1 + \dots + \frac{\partial f}{\partial x_n}(x^0 + tv)v_n:$$

in particular, if I calculate in 0:

$$\begin{aligned} g'(0) =& \frac{\partial f}{\partial x_1}\left(x^0\right) v_1 + \dots + \frac{\partial f}{\partial x_n}\left(x^0\right) v_n \\ =& \left(\frac{\partial f}{\partial x_1}\left(x^0\right), \dots, \frac{\partial f}{\partial x_n}(x_0)\right) \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \end{aligned}$$

or more coincisely:

$$g'(0) = Df\left(x^0\right) v =: Df_{x_0}(v).$$

This is precisely what we were looking for: the *directional derivative of f in direction v*.

So if $v = e^i = (\underbrace{0, \dots, 0}_{i-1}, 1, 0, \dots, 0)$, we get precisely the ith partial derivative:

$$Df_{x^0}(e^i) = \frac{\partial f}{\partial x_i}\left(x^0\right);$$

this shows that the concept of *directional derivative* is a generalization of the partial ones.

Let's get back to the previous:

Example 13. $Q = 4K^{\frac{3}{4}}L^{\frac{1}{4}} = Q(K, L)$

$x^0 = (K_0, L_0) = (10000, 625)$ *(which was corresponding to* $t = 10$ *- but we don't care about* t *now).*

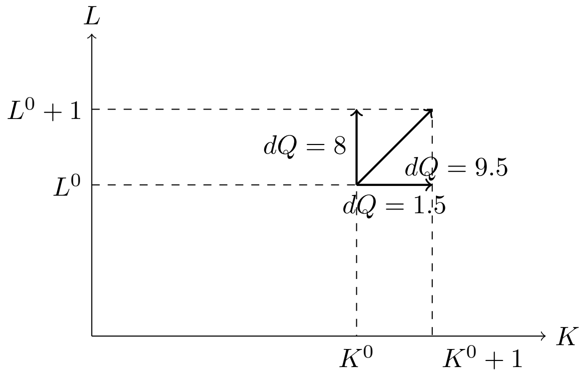
In the compact notation, we have, for instance

$$DQ_{(K^0,L^0)}(1,1)$$

(we move diagonally through a $45°$ *line), and that gives*

$$\begin{aligned} DQ_{(K^0,L^0)}(1,1) =& \frac{\partial Q}{\partial K}(K^0, L^0) \cdot 1 + \frac{\partial Q}{\partial L}(K^0, L^0) \cdot 1 \\ =& 1.5 + 8 = 9.5. \end{aligned}$$

The directional derivative is a simple concept: just take the partial ones and multiply by the vector coordinates!



We can also write

$$dQ = \frac{\partial Q}{\partial k} \cdot dK + \frac{\partial Q}{\partial L} \cdot dL = 9.5$$

which is the way we calculated approximations (then, in this case the result is exact, but we won't bother).

"In this direction", *we get an increase of* 9.5*. Along the* K *axis, it would have been* 1.5*, along the* L *axis* 8*.*

The picture seems to suggest that the oblique direction is "better" than the orthogonal ones... but that vector is also longer!

We want to make the same calculation for a vector of lenght 1: $\|v\| = 1$*, where*

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

and that means

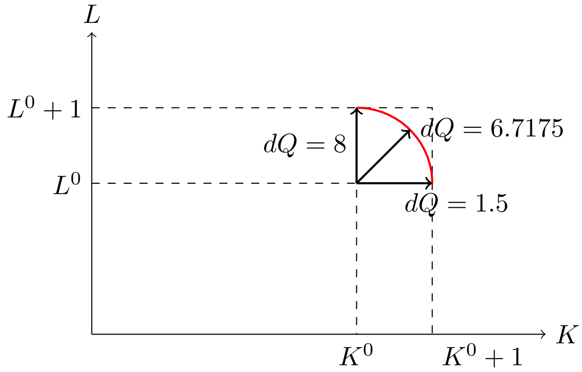
$$\|(a, a)\| = 1 \iff \sqrt{a^2 + a^2} = 1 \iff a = \frac{1}{\sqrt{2}} \Rightarrow v = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Finally,

$$DQ_{K^0,L^0}\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right) = 1.5\frac{1}{\sqrt{2}} + 8\frac{1}{\sqrt{2}} = 9.5\frac{1}{\sqrt{2}} \approx 6.7175\dots$$

So the claim that the 45° line is the best direction was wrong! Just increasing L is already better!

This raises the question: what is the best *direction?*



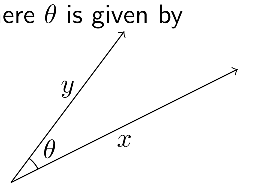
We want to solve the problem

$$max_{(v_1,v_2)} DQ_{(K^0,L^0)}(v_1,v_2) \qquad s.t. \|(v_1,v_2)\| = 1.$$

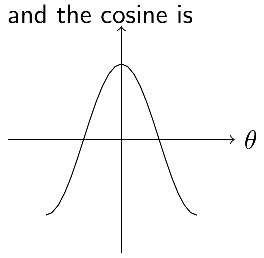
Notice that

$$x,y \in \mathbb{R}^n \Rightarrow x\cdot y = \sum_{i=1}^{n} x_i y_i = \|x\|\|y\|\cos\theta$$

where θ is given by



and the cosine is



(proof on S.B: Theorem 10.3, pp. 215-217).

Now,

$$Df_{x^0}(v) = Df\left(x^0\right)\cdot v = \left\|Df\left(x^0\right)\right\| \underbrace{\|v\|}_{1} \cos\theta$$

and hence the only thing that can change is θ. More precisely, we just want to maximize $\cos\theta$... and that happens for $\theta = 0 \Rightarrow \cos\theta = 1$.

Then $\theta_{\text{max}} = 0$.

This is a general result:

Theorem 14. *Let $f : U \to \mathbb{R}$ be differentiable with continuous partial derivatives, $U \subset \mathbb{R}^n$.*

At any point $x_0 \in U$ with $Df\left(x^0\right) \neq 0$, the vector $Df\left(x^0\right)$ at x^0 points into the direction in which f increases most rapidly.

Exercise 15. *Homework: find the best direction in the example given, and calculate the highest dQ.*

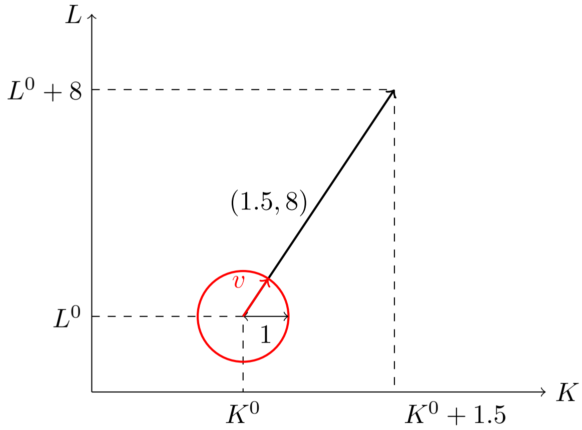
09/11/2010

We have to write the Jacobian:

$$DQ(K^0, L^0) = \left(\frac{\partial Q}{\partial K}(10000, 625), \frac{\partial Q}{\partial L}(10000, 625)\right) = (1.5, 8)$$

and this directly gives us the direction of the optimal v.

We must however normalize it:



We want to have

$$v = aDQ(K^0, L^0), a > 0$$

such that $\|v\| = 1$.

$$\Rightarrow \|aDQ(K^0, L^0)\| = 1$$

We'll just take

$$a = \frac{1}{\|DQ(K^0, L^0)\|} \Rightarrow v = \frac{DQ(K^0, L^0)}{\|DQ(K^0, L^0)\|}$$

since in general

$$\left\|\frac{v}{\|v\|}\right\| = 1.$$

Finally, we can write

$$dQ = DQ(K^0, L^0)\frac{DQ(K^0, L^0)}{\|DQ(K^0, L^0)\|}.$$

Now, it "just happens" that the two components have the same numerator... se we're taking the dot product of a vector by itself, also known as the norm:

$$dQ = \frac{\|DQ(K^0, L^0)\|^2}{\|DQ(K^0, L^0)\|} = \|DQ(K^0, L^0)\| = \|(1.5, 8)\| =$$
$$= \sqrt{1.5^2 + 8^2} = \sqrt{2.25 + 64} = \sqrt{66.25} \approx 8.139\ldots$$

As we expected, this is higher than 8.

In general, we consider, for directional derivatives, expressions of the form

$$Df\left(x^0\right)v = Df_{x^0}(v)$$

only where $\|v\| = 1$.

Sometimes, one may consider the vector $Df\left(x^0\right)$ as a *column* vector:

$$Df\left(x^0\right)^T = \begin{bmatrix} \frac{\partial f}{\partial x_1}\left(x^0\right) \\ \vdots \\ \frac{\partial f}{\partial x_n}\left(x^0\right) \end{bmatrix} = \nabla f\left(x^0\right)$$

which is called *gradient (vector)* of f at x^0.

As we've seen, the gradient always points towards the direction of steepest ascent of the function.

1.4 Jacobians, higher order derivatives, and Hessians

Consider a function

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

$$\begin{array}{cccccc} f & : & \mathbb{R}^n & \longrightarrow & \mathbb{R}^m \\ & & \cup & & \cup \\ & & x & \longrightarrow & f(x) \\ & & \| & & \| \\ & & (x_1, \ldots, x_n) & \longrightarrow & (f_1(x), \ldots, f_m(x)) \end{array}$$

We may consider this function f at a specific point $x^0 \in \mathbb{R}^n$ and take

$$\Delta x = (\Delta x_1, \ldots, \Delta x_n).$$

Then:

$$f_1(x^0 + \Delta x) - f_1\left(x^0\right) \approx \sum_{i=1}^{n} \frac{\partial f_1}{\partial x_i}\left(x^0\right)\Delta x_i$$

but if we can do that for f_1, we can do it for all components:

$$\vdots$$

$$f_m(x^0 + \Delta x) - f_m\left(x^0\right) \approx \sum_{i=1}^{n} \frac{\partial f_m}{\partial x_i}\left(x^0\right)\Delta x_i$$

We can rewrite the right hand side merging all equations in a vectorial form:

$$f(x^0 + \Delta x) - f\left(x^0\right) \approx \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1}\left(x^0\right) & \cdots & \frac{\partial f_1}{\partial x_n}\left(x^0\right) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}\left(x^0\right) & \cdots & \frac{\partial f_m}{\partial x_n}\left(x^0\right) \end{bmatrix}}_{m \times n} \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix}.$$

This will be written

$$Df\left(x^0\right) \cdot \Delta x$$

where now $Df\left(x^0\right)$ is the *Jacobian (matrix)* of f at x^0, and

$$\Delta x = \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix}.$$

The expression above is a linear approximation of

$$\Delta y = f(x^0 + \Delta x) - f\left(x^0\right) \in \mathbb{R}^n.$$

We can also express this in terms of differentials:

$$\begin{bmatrix} dy_1 \\ \vdots \\ dy_m \end{bmatrix} = \begin{bmatrix} df_1\left(x^0\right) \\ \vdots \\ df_m\left(x^0\right) \end{bmatrix} = Df\left(x^0\right) \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix},$$

or, in a more coincise way:

$$df\left(x^0\right) = Df\left(x^0\right) dx,$$

which resembles very much the one-dimensional case in which we had first met the differential.

Example 16. *Let's consider* 2 *commodities with demand functions*

$$q_1 = 6p_1^{-2}p_2^{\frac{3}{2}}y$$
$$q_2 = 4p_1p_2^{-1}y^2$$

(which economically speaking are not that absurd - demand for each good decreases with its price and increases in the price of the other one).

We can write

$$\begin{aligned} q &= (q_1(p_1, p_2, y), q_2(p_1, p_2, y)) \\ &= q(p_1, p_2, y) \\ \Rightarrow &\mathbb{R}^3 \xrightarrow{q} \mathbb{R}^2. \end{aligned}$$

$$\underbrace{D_q}_{2\times 3} = \begin{bmatrix} -12p_1^{-3}p_2^{\frac{3}{2}}y & 9p_1^{-2}p_2^{\frac{1}{2}}y & 6p_1^{-2}p_2^{\frac{3}{2}} \\ 4p_2^{-1}y^2 & -4p_1p_2^{-2}y^2 & 8p_1p_2^{-1}y \end{bmatrix}.$$

Let's assume $p_1^0 = 6$, $p_2^0 = 9$, $y_0 = 2$. *We get:*

$$D_q(6,9,2) = \begin{bmatrix} -3 & 2^{-1}3 & 2^{-1}9 \\ 2^4 3^{-2} & -2^5 3^{-3} & 2^5 3^{-1} \end{bmatrix}.$$

If, for example, $dp_1 = 0.1$, $dp_2 = 0.1$ *and* $dy = -0.1$, *what do we get? We are searching for the linear approximation of the effect of a simultaneous change in all components - and since the different changes have different effects, the result is not obvious. Let's calculate:*

$$\begin{bmatrix} dq_1 \\ dq_2 \end{bmatrix} = \begin{bmatrix} -3 & \frac{3}{2} & \frac{9}{2} \\ \frac{16}{9} & -\frac{32}{27} & \frac{32}{3} \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.1 \\ -0.1 \end{bmatrix} = \begin{bmatrix} \frac{-6+3-9}{20} \\ \frac{48-32-288}{270} \end{bmatrix} = \begin{bmatrix} -\frac{12}{20} \\ -\frac{272}{270} \end{bmatrix} = \begin{bmatrix} -\frac{3}{5} \\ -\frac{136}{135} \end{bmatrix}.$$

So we found the change in the quantities demanded when the given variable changes happen.

We have considered function from $\mathbb{R}^n$ to $\mathbb{R}^m$. We want to extend the considerations above to the behaviour of a function on a curve.

Consider

$$\begin{array}{ccccc} & & \overset{g}{\frown} & & \\ \mathbb{R} & \xrightarrow{x(\cdot)} & \mathbb{R}^n & \xrightarrow{f} & \mathbb{R}^m \\ \Cup & & \Cup & & \Cup \\ t & \longmapsto & x(t) & \longmapsto & f(x(t)) \end{array}$$

$$g(t) = \begin{bmatrix} g_1(t) \\ \vdots \\ g_m(t) \end{bmatrix} = \begin{bmatrix} f_1(x(t) \\ \vdots \\ f_m(x(t)) \end{bmatrix} \Rightarrow g_i(t) = f_i(x(t))$$

$\forall i = 1 \dots m.$

$$\begin{aligned} \Rightarrow g_i'(t) &= \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(x(t))x_j'(t) \forall i = 1 \dots m \\ &= Df_i(x(t))x'(t) \forall i = 1 \dots n \end{aligned}$$

$$\Rightarrow g'(t) = \begin{bmatrix} g_1'(t) \\ \vdots \\ g_m'(t) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x(t)) & \dots & \frac{\partial f_1}{\partial x_n}(x(t)) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x-1}(x(t)) & \dots & \frac{\partial f_m}{\partial x_n}(x(t)) \end{bmatrix} \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix} = Df(x(t))x'(t):$$

this is our *chain rule III*. We can express it also as follows:

$$g'(t) = D(f \circ x)(t)$$

where, in general,

$$g(t) = f(x(t))\ \forall t \Rightarrow g = f \circ x.$$

Example 17. *We'll extend the previous example: we had*

$$\begin{aligned} q_1 =& q_1(p_1, p_2, y) \\ q_2 =& q_2(p_1, p_2, y); \end{aligned}$$

we now assume a functional form for the independent variables too:

$$\begin{aligned} p_1(t) &= \sqrt{12t} \\ p_2(t) &= t^2 \\ y(t) &= t - 1 \end{aligned}$$

(we have added some form of inflation...)

We want to consider

$$\begin{array}{ccccc} \mathbb{R} & \xrightarrow{(p_1,p_2,y)} & \mathbb{R}^3 & \xrightarrow{(q_1,q_2)} & \mathbb{R}^2 \\ \cup & & \cup & & \cup \\ t & \longmapsto & (p_1(t), p_2(t), y(t)) & \longmapsto & (q_1, q_2). \end{array}$$

How is demand changing over time, that is, with respect to t, at $t = 3$?

Remark 18.

$$(p_1(r), p_2(r), y(3)) = (6, 9, 2) = (p_1^0, p_2^0, y^0)$$

which by "chance" are the same numbers as in the former example.

$$g(t) = \begin{bmatrix} q_1(p_1(t), p_2(t), y(t)) \\ q_2(p_1(t), p_2(t), y(t)) \end{bmatrix}$$

We are looking for the variations in demand, which we will calculate as in 1.4:

$$\begin{bmatrix} g_1'(t) \\ g_2'(t) \end{bmatrix} = \begin{bmatrix} \frac{dq_1}{dt} \\ \frac{dq_2}{dt} \end{bmatrix} = \begin{bmatrix} \frac{\partial q_1}{\partial p_1} & \frac{\partial q_1}{\partial p_2} & \frac{\partial q_1}{\partial y} \\ \frac{\partial q_2}{\partial p_1} & \frac{\partial q_2}{\partial p_2} & \frac{\partial q_2}{\partial y} \end{bmatrix} \begin{bmatrix} p_1'(t) \\ p_2'(t) \\ y'(t) \end{bmatrix} = \begin{bmatrix} -3 & \frac{3}{2} & \frac{9}{2} \\ \frac{16}{9} & -\frac{32}{27} & \frac{32}{3} \end{bmatrix} \begin{bmatrix} p_1'(t) \\ p_2'(t) \\ y'(t) \end{bmatrix}$$

Now,

$$\begin{aligned} p_1'(t) =& \sqrt{12}\frac{1}{2}t^{-\frac{1}{2}} \overset{t=3}{=} 1 \\ p_2'(t) =& 2t = 6 \\ y'(t) =& 1, \end{aligned}$$

so

$$\begin{bmatrix} g_1'(t) \\ g_2'(t) \end{bmatrix} = \begin{bmatrix} -3 & \frac{3}{2} & \frac{9}{2} \\ \frac{16}{9} & -\frac{32}{27} & \frac{32}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 6 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{21}{2} \\ \frac{48+32\cdot 3}{27} \end{bmatrix} = \begin{bmatrix} \frac{21}{2} \\ \frac{48}{9} \end{bmatrix}.$$

The message is: although prices are increasing over time, income is increasing too and its effect is, at least at time $t = 3$, dominating the others, so demand is increasing anyway.

There is still one further generalization we can make. Consider finally

$$\mathbb{R}^s \xrightarrow{x(\cdot)} \mathbb{R}^n \xrightarrow{f} \mathbb{R}^m \qquad (g : \mathbb{R}^s \to \mathbb{R}^m)$$

$$\Rightarrow g(t) = \begin{bmatrix} g_1(t_1,\dots,t_s) \\ \vdots \\ g_m(t_1,\dots,t_s) \end{bmatrix} = \begin{bmatrix} f_1(x(t(t_1,\dots,t_s)) \\ \vdots \\ f_m(x(t_1,\dots,t_s)) \end{bmatrix}$$

where

$$x(t) = (x_1(t_1,\dots,t_s),\dots,x_n(t_1,\dots,t_s))$$

so that we can consider, for any $i = 1,\dots,m$ and $h = 1,\dots,s$,

$$\frac{\partial g_i}{\partial t_n}(t).$$

The only difference from 1.4 is that now $x_j'(t)$ becomes a matrix:

$$\begin{aligned}\frac{\partial g_i}{\partial t_n}(t) =& \sum_{j?1}^{m} \frac{\partial f_i}{\partial x_j}(x(t))\frac{\partial x_j}{\partial t_n}(t) \\ =& Df_i(x(t)) \begin{bmatrix} \frac{\partial x_1}{\partial t_n}(t) \\ \vdots \\ \frac{\partial x_n}{\partial t_n}(t) \end{bmatrix} \forall i = 1,\dots,m, \qquad h = 1,\dots,s\end{aligned}$$

which in matrix form is

$$\begin{aligned}Dg(t) =& \left[\frac{\partial g_i}{\partial t_n}\right]_{i,h} \\ =& \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x(t)) & \cdots & \frac{\partial f_1}{\partial x_n}(x(t)) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x(t)) & \cdots & \frac{\partial f_m}{\partial x_n}(x(t)) \end{bmatrix} \begin{bmatrix} \frac{\partial x_1}{\partial t_1}(t) & \cdots & \frac{\partial x_1}{\partial t_s}(t) \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial t_1}(t) & \cdots & \frac{\partial x_n}{\partial t_s}(t) \end{bmatrix}.\end{aligned}$$

So finally to calculate

$$\frac{\partial g_i}{\partial t_n}(t),$$

we will have to calculate the i-th row of the first matrix and the h-th column of the second, and multiply them.

Again, we can rewrite

$$Dg(t) = \underbrace{\underbrace{Df(x(t))}_{m\times n}\underbrace{Dx(t)}_{n\times s}}_{m\times s}$$

which we will call *chain rule IV*.

We can use this result to rewrite in general notation the following:

$$\mathbb{R}^s \xrightarrow{f} \mathbb{R}^n \xrightarrow{g} \mathbb{R}^m \quad (h: \mathbb{R}^s \to \mathbb{R}^m) \Rightarrow h = g \circ f$$

$$\begin{aligned} \Rightarrow Dh(x) =& Dg(f(x))Df(x) \\ =& D(g \circ f)(x); \end{aligned}$$

this is a way to decompose a function in its components in order to derive it: this is nothing more than the well known formula for deriving composite functions, in multidimensional case.

10/11/2010

1.4.1 Higher order derivatives

Given

$$f : \mathbb{R}^n \to \mathbb{R},$$

we have seen that we can consider the partial derivatives

$$\begin{array}{ccccc} \frac{\partial f}{\partial x_i} & : & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & & \Psi & & \Psi \\ & & x & \longmapsto & \frac{\partial f}{\partial x_i}(x) \end{array} \qquad \forall i = 1, \dots, n.$$

Consider

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}(x_1, \dots, x_n)\right);$$

there is a shorter way to write the same thing:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \dots, x_n),$$

and there is a further notation:

$$\frac{\partial^2 f}{\partial x_i \partial x_i} \overset{\text{def}}{=} \frac{\partial^2 f}{\partial x_i^2}.$$

Now: how many derivatives of this form do we have? We have n indexes i and, for each, n indexes j. There is a way to write them together, which is the *Hessian (matrix)*:

$$D^2 f(x) \overset{\text{def}}{=} \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \dots & \frac{\partial f}{\partial x_n^2}(x) \end{bmatrix}.$$

Example 19. *Take*

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

a production function. Then,

$$\begin{aligned}\frac{\partial Q}{\partial x} =& \alpha A x^{\alpha-1}y^{\beta}\\ \frac{\partial Q}{\partial y} =& \beta A x^{\alpha}y^{\beta-1}\\ \Rightarrow D^2Q =& \begin{bmatrix}\alpha(\alpha-1)Ax^{\alpha-2}y^{\beta} & \alpha\beta Ax^{\alpha-1}y^{\beta-1}\\ \alpha\beta Ax^{\alpha-1}y^{\beta-1} & \beta(\beta-1)Ax^{\alpha}y^{\beta-2}\end{bmatrix}\end{aligned}$$

We can observe that the two terms in position $(1, 2)$ and $(2, 1)$ are the same. It may be by chance... but it is not.

Theorem 20 (Yanng's Theorem)**.** *Suppose $f : U \to \mathbb{R}$ is such that all partial derivatives until order 2 exist and are continuous functions, and U is open. Then, for all $x \in U$ and each pair of indices i and j, we have*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x).$$

In other terms, the Hessian matrix $D^2 f(x)$ is symmetric.

Therefore, let's take such a partial derivative of order 2:

$$\begin{array}{cccc} \frac{\partial^2}{\partial x_i \partial x_j} : & \mathbb{R}^n \longrightarrow \mathbb{R} & & \\ & \cup \qquad\qquad \cup & & \\ & x \longmapsto \frac{\partial^2 f}{\partial x_i \partial x_j}(x) & & \forall i, j = 1, \ldots, n. \end{array}$$

But then, we may consider

$$\frac{\partial}{\partial x_l}\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right) = \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_l}(x) \forall i, j, l = 1, \ldots, n.$$

Certain functions may be differentiable an infinity of times.

Definition 21. *A $\mathcal{C}^k$ function f is a function such that all partial derivatives until order k exist and are continuous.*

So for instance in the Yanng's theorem, we could have said simply "suppose that f is $\mathcal{C}^2$"...

This will be used in the next topic...

1.5 Taylor expansion

Let us consider

$$f : \mathbb{R} \to \mathbb{R}$$

$\mathcal{C}^1$; then, as we know,

$$f(x^0 + h) \approx f\left(x^0\right) + f'\left(x^0\right)h. \tag{1}$$

Then, define

$$R(h,x^0) \stackrel{\text{def}}{=} f(x^0+h) - f\left(x^0\right) - f'\left(x^0\right)h,$$

the "approximation error".

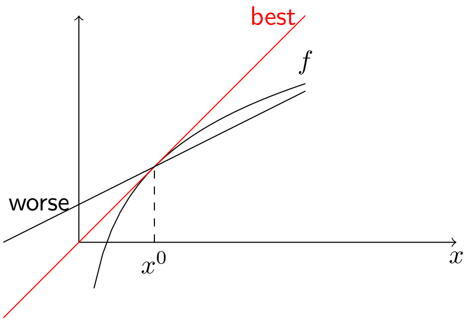
Then,

$$\frac{R(h,x^0)}{h} \stackrel{h\to 0}{\to} 0,$$

as can be deduced by the definition of $R(h,x)$. But obviously the denominator of this definition goes to 0 as $h \to 0$, so the numerator does to... and *faster*!

(1) is the *best* linear approximation of f at x^0.

In what sense?



But this is not necessarily the best approximation *in general*!

We can derive a *quadratic approximation*, assuming f is $\mathcal{C}^2$:

$$f(x^0+h) \approx f\left(x^0\right) + f'\left(x^0\right)h + \frac{1}{2}f''\left(x^0\right)h^2. \tag{2}$$

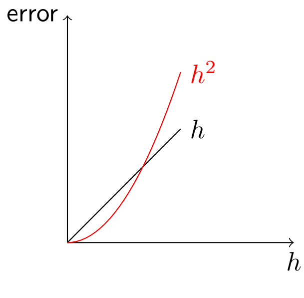
For its error,

$$R_2(h,x^0) \stackrel{\text{def}}{=} f(x^0+h)\ldots,$$

we get

$$\frac{R_2(h,x^0)}{h^2} \stackrel{h\to 0}{\to} 0,$$

and it tends *faster* than $R(h,x^0)$!



Theorem 22 (Taylor)**.** *Let $f: U \to \mathbb{R}$ be a $\mathcal{C}^{k+1}$ function, with $U \subset \mathbb{R}$ an interval.*

Then, for any numbers x^0 and x^0+h in U, there exists a number c between x^0 and x^0+h such that

$$f(x^0+h) = \sum_{n=0}^{h} \frac{1}{n!} f^{(n)}\left(x^0\right)h^n + \frac{1}{(k+1)!}f^{(k+1)}(c)h^{k+1}.$$

where

$$f^{(n)}(x) := \left(f^{(n-1)}\right)'(x) \qquad \forall n \in \mathbb{N}$$

and

$$f^{(0)} := f(x) \qquad \forall x \in U.$$

Remark 23. *If we take $k = 2$, we get exactly the expression* (2).

The formula of the Taylor expansion is a polynomial in the variable h, and we can approximate arbitrary functions (like exponential, logarythm, trigonometric functions…) with it. The error is given by

$$\frac{1}{(k+1)!}f^{(k+1)}(c)h^{k+1},$$

for some c, which however is not “too far away” from x^0, since it is between x^0 and $x^0 + h$.

Remark 24. *This is the k-th* order Taylor expansion, *and*

$$R_k(h, x^0) := \frac{1}{(k+1)!}f^{k+1}(c)h^{k+1}$$

has the following property (which comes immediately from the definition):

$$\frac{R_k(h, x^0)}{h^k} = \frac{1}{(k+1)!}f^{(k+1)}(c)h.$$

Now, what does

$$\frac{1}{(k+1)!}f^{(k+1)}(c)$$

do for $h \to 0$? It goes to

$$\frac{1}{(k+1)!}f^{(k+1)}\left(x^0\right),$$

since $c \to x^0$. And that expression is a real number (since by hypothesis the $k+1$st derivatives exist and are finite). When it is multiplied by h, which tends to 0, we get something that, again, tends to 0.

Proof. Fix x^0 and h such that $x^0 \in U$ and $x^0 + h \in U$.

Define

$$g(t) \stackrel{\text{def}}{=} f(t) - f\left(x^0\right) - \sum_{n=1}^{k} \frac{1}{n!} f^{(n)}\left(x^0\right)(t - x^0)^n - M(t - x^0)^{k+1},$$

where

$$M := \frac{1}{h^{k+1}}\left[f(x^0 + h) - f\left(x^0\right) - \sum_{n=1}^{k} \frac{1}{n!} f^{(n)}\left(x^0\right)h^n\right];$$

then,

$$g\left(x^0\right) = 0.$$

Now,

$$g(x^0+h) = f(x^0+h) - f\left(x^0\right) - \sum_{n=1}^{k} \frac{1}{n!} f^{(n)}\left(x^0\right) h^n - Mh^{k+1}$$
$$= 0.$$

We know from Rolle (since g is a differentiable function) that $\exists c_1 \in [x^0, x^0, h]$ such that $g'(c_1) = 0$, where

$$g'(t) = f'(t) - 0 - f'\left(x^0\right) - \sum_{n=2}^{k} \frac{1}{(n-1)!} f^{(n)}\left(x^0\right)(t-x_0)^{n-1} - (k+1)M(t-x^0)^k.$$

What is the value in x^0? It is 0. So we have 2 points in which the derivative of g becomes 0. So, there must exist $c_2 \in [x_0, c_1]$ such that $g''(c_2) = 0$, where

$$g''(t) = f''(t) - f''\left(x^0\right) - \sum_{n=3}^{k} \frac{1}{(n-2)!} f^{(n)}\left(x^0\right)(t-x^0)^{n-2} - (k+1)kM(t-x_0)^{k-1}.$$

We can now evaluate g'':

$$g''\left(x^0\right) = g''(c_2) = 0:$$

once again, applying Rolle, we conclude

$$\exists c_3 \in [x^0, c_2] : g'''(c_3) = 0.$$

Of course we could continue, it's always the same. The general formula is:

$$\forall i \leq k-1$$
$$g^{(i)}(t) = f^{(i)}(t) - f^{(i)}\left(x^0\right) - \sum_{n=i+1}^{k} \frac{1}{(n-i)!} f^{(n)}\left(x^0\right)(t-x^0)^{n-i}$$
$$-(k+1)\cdot \dots \cdot (k+1-(i-1))M(t-x_0)^{k+1-i}$$

and $\exists c_{i+1}$ between x^0 and c_i such that $g^{i+1}(c_{i+1}) = 0$.
From this formula, we can in particular get that

$$g^{(k-1)}(t) = f^{(k-1)}(t) - f^{(k-1)}\left(x^0\right) - \frac{1}{1} f^{(k)}\left(x^0\right)(t-x^0)$$
$$-(k+1)\cdot \dots \cdot 3M(t-x^0)^2$$

and $\exists c_k$ between x^0 and c_{k-1} such that $g^{(k)}(c_k) = 0$, where

$$g^{(k)}(t) = f^{(k)}(t) - f^{(k)}\left(x^0\right) - (k+1)!M(t-x^0)$$

$$\Rightarrow g^{(k)}\left(x^0\right) = 0;$$

so once more, we have that $g^{(k)}$ takes the same value in 0 and c_k, so applying a last time Rolle, we get that $\exists c_{k+1}$ between x^0 and c_k such that $g^{(k+1)}(c_{k+1}) = 0$ where

$$g^{(k+1)}(t) = f^{(k+1)}(t) - (k+1)!M.$$

This implies that

$$f^{(k+1)}(c_{k+1}) = (k+1)!M$$

which is equivalent to

$$M = \frac{1}{(k+1)!} f^{(k+1)}(c_k + 1).$$

If we multiply both sides by h^{k+1}, we get:

$$\left[f(x^0 + h) - f\left(x^0\right) - \sum_{n=1}^{k} \frac{1}{n!} f^{(n)}\left(x^0\right) h^n \right] = \frac{1}{(k+1)!} f^{(k+1)}(c_k + 1) h^{k+1}$$

$$\iff f(x^0 + h) = f\left(x^0\right) + \sum_{n=1}^{k} \frac{1}{n!} f^{(n)}\left(x^0\right) h^n + \frac{1}{(k+1)!} f^{(k+1)}(c) h^{k+1},$$

where $c = c_{k+1}$.

□

11/11/10

What we have seen is the Taylor approximation *in 1 variable*.

We can imagine that the proof becomes cumbersome in the general (multivariate) case - but the principle is similar: consider

$$f : U \to \mathbb{R}$$

with $U \subset \mathbb{R}^n$. If f is $\mathcal{C}^1$ and x^0, $x^0 + h$ are elments of U, then

$$f(x^0 + h) = f\left(x^0\right) + \frac{\partial f}{\partial x_1}\left(x^0\right) h_1 + \cdots + \frac{\partial f}{\partial x_n}\left(x^0\right) h_n + R_1(h, x^0);$$

it is similar to the linear approximation that we had seen, except for the error term that completes the equality. The error also has an analogous property:

$$\frac{R_1(h, x^0)}{\|h\|} \overset{h \to 0}{\to} 0$$

or, in more concise notation,

$$f(x^0, h) = f\left(x^0\right) + Df\left(x^0\right) h + R_1(h, x^0).$$

We now want to introduce higher order approximations; for this purpose, consider now the following expression:

$$\sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}\left(x^0\right) h_i, h_j;$$

we can rewrite it in a more concise way (in order to plug it in the approximation): first, we explicit the two sums:

$$\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_j}\left(x^0\right)h_i h_j\right),$$

then we can write the same term in vector form:

$$\begin{aligned}
&\left(\sum_{i=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_1}\left(x^0\right)h_i,\ldots,\sum_{i=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_n}\left(x^0\right)h_i\right)\begin{bmatrix}h_1\\ \vdots\\ h_n\end{bmatrix}\\
=&(h_1,\ldots,h_n)\begin{bmatrix}\frac{\partial^2 f}{\partial x_1^2}\left(x^0\right) & \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n}\left(x^0\right)\\ \vdots & \ddots & \vdots\\ \frac{\partial^2 f}{\partial x_n \partial x_1}\left(x^0\right) & \ldots & \frac{\partial^2 f}{\partial x_n^2}\left(x^0\right)\end{bmatrix}\begin{bmatrix}h_1\\ \vdots\\ h_n\end{bmatrix}\\
=&\underbrace{\underbrace{h^T}_{1\times n}\underbrace{D^2 f\left(x^0\right)}_{n\times n}\underbrace{h}_{n\times 1}}_{1\times 1\in\mathbb{R}}.
\end{aligned}$$

We can therefore state the following

Theorem 25. *Let $f : U \in \mathbb{R}\mathcal{C}^2$ with $U \in \mathbb{R}^n$ open and $x^0 \in U$. Then, there exists a $\mathcal{C}^2$ function*

$$h \mapsto R_2(h, x^0)$$

such that, for any $x^0 + h \in U$ with the property that the line segment from x^0 to $x^0 + h$ lies in U,

$$f(x^0+h) = f\left(x^0\right) + Df\left(x^0\right)h + \frac{1}{2}h^T D^2 f\left(x^0\right)h + R_2(h, x^0)$$

and

$$\frac{R_2(h,x^0)}{\|h\|^2} \overset{h\to 0\in\mathbb{R}^n}{\to} 0.$$

There is obviously the possibility to reach even higher order approximations, but for our (economic) purposes the second order will be enough.

However, we want to elaborate a little bit on it.

Example 26. *$n = 2$ gives us*

$$\begin{aligned}
&f\left(x_1^0+h_1, x_2^0+h_2\right)\\
=&f(x_1^0,x_2^0)+\left(\frac{\partial f}{\partial x_1}\left(x^0\right),\frac{\partial f}{\partial x_2}\left(x^0\right)\right)\begin{bmatrix}h_1\\ h_2\end{bmatrix}+\frac{1}{2}(h_1,h_2)\begin{bmatrix}\frac{\partial^2 f}{\partial x_1^2}\left(x^0\right) & \frac{\partial^2 f}{\partial x_1 \partial x_2}\left(x^0\right)\\ \frac{\partial^2 f}{\partial x_1 \partial x_2}\left(x^0\right) & \frac{\partial^2 f}{\partial x_2^2}\left(x^0\right)\end{bmatrix}\begin{bmatrix}h_1\\ h_2\end{bmatrix};
\end{aligned}$$

more specifically:

$$f(x_1,x_2) = x_1^{\frac{1}{4}}x_2^{\frac{3}{4}}$$

$$(x_1^2, x_2^0) = (1,1) = P$$

$$\Downarrow$$

$$
\begin{aligned}
\frac{\partial f}{\partial x_1} &= \frac{1}{4}x_1^{-\frac{3}{4}}x_2^{\frac{3}{4}} \overset{P}{=} \frac{1}{4} \\
\frac{\partial f}{\partial x_2} &= \frac{3}{4}x_1^{\frac{1}{4}}x_2^{-\frac{3}{4}} \overset{P}{=} \frac{3}{4} \\
\frac{\partial^2 f}{\partial x_1^2} &= -\frac{3}{16}x_1^{-\frac{7}{4}}x_2^{\frac{3}{4}} \overset{P}{=} -\frac{3}{16} \\
\frac{\partial^2 f}{\partial x_2^2} &= -\frac{3}{16}x_1^{\frac{1}{4}}x_2^{-\frac{5}{4}} \overset{P}{=} -\frac{3}{16} \\
\frac{\partial^2 f}{\partial x_1 \partial x_2} &= \frac{3}{16}x_1^{-\frac{3}{4}x_2^{-\frac{1}{4}}=\frac{3}{16}}
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
& f(1,1) + \left(\frac{1}{4}, \frac{3}{4}\right)\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \frac{1}{2}(h_1, h_2)\begin{bmatrix} -\frac{3}{16} & \frac{3}{16} \\ \frac{3}{16} & -\frac{3}{16} \end{bmatrix}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\
=& 1 + \frac{1}{4}h_1 + \frac{3}{4}h_2 + \frac{1}{2}\left(-\frac{3}{16}h_1 + \frac{3}{16}h_2, \frac{3}{16}h_1 - \frac{3}{16}h_2\right)\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\
=& 1 + \frac{1}{4}h_1 + \frac{3}{4}h_2 + \frac{1}{2}\left(-\frac{3}{16}h_1^2 + \frac{3}{16}h_1h_2 + \frac{3}{16}h_1^2 - \frac{3}{16}h_2^2\right) \\
=& 1 + \frac{1}{4}h_1 + \frac{3}{4}h_2 - \frac{3}{32}h_1^2 + \frac{3}{16}h_1h_2 - \frac{3}{32}h_2^2.
\end{aligned}
$$

How good is this Taylor approximation? Let's take $h_1 = 0.1$, $h_2 = -0.1$. *Then, we get, for the linear component:*

$$
\begin{aligned}
f(x^0) + Df(x^0)h =& 1 + \frac{1}{4} \cdot 0.1 + \frac{3}{4} \cdot -0.1 \\
=& 1 + \left(-\frac{1}{2}\right)0.1 = 1 - 0.05 = 0.95,
\end{aligned}
$$

while the quadratic part is

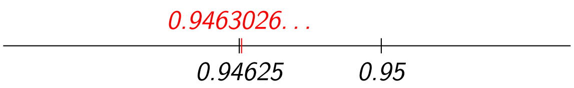
$$
\begin{aligned}
\frac{1}{2}h^T D^2 f(x^0)h =& -\frac{3}{32} \cdot 0.01 + \frac{3}{16}(-0.01) - \frac{3}{32}0.01 \\
=& -\frac{12}{32} \cdot 0.01 = -\frac{3}{8} \cdot 0.01 = -0.00375.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
f(x^0) =& Df(x^0)h + \frac{1}{2}h^T D^2 f(x^0)h \\
=& 0.95 - 0.00375 = 0.94625,
\end{aligned}
$$

where the true value is

$$f(x^0 + h) = f(1.1, 0.9) = 0.9463026\ldots$$

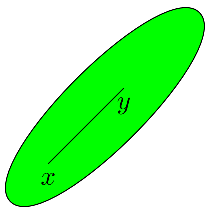


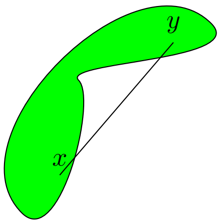
It is interesting to notice that the approximation is not monotonic.

1.6 Convexity, concavity and quasiconcavity

A set $U \subset \mathbb{R}^n$ is *convex* if:

$$x, y \in U \Rightarrow tx + (1-t)y \in U \; \forall t \in (0,1).$$

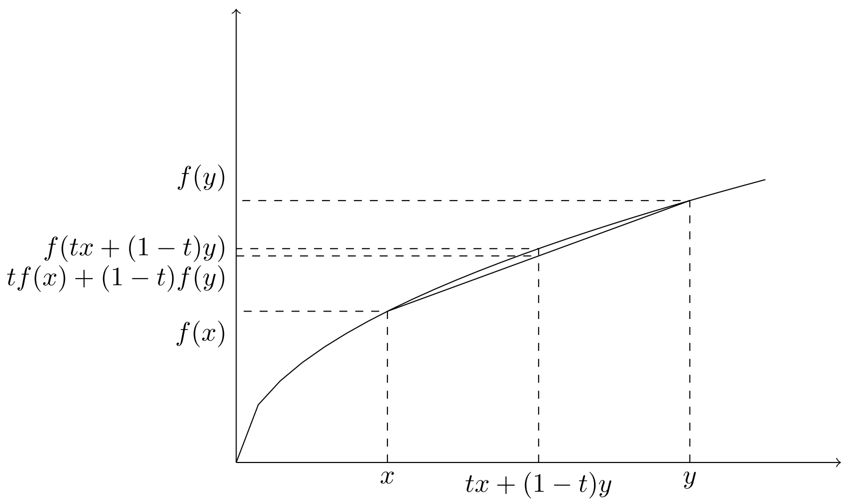




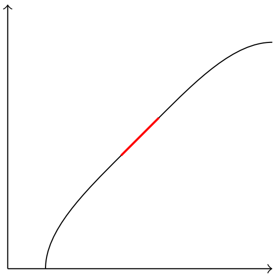
Definition 27. *The function $f : U \in \mathbb{R}, U \subset \mathbb{R}^n$ convex, is* concave *if*

$$f(tx + (1-t)y) \geq tf(x) + (1-t)f(y) \qquad \forall x, y \in U \forall \, t \in (0,1).$$

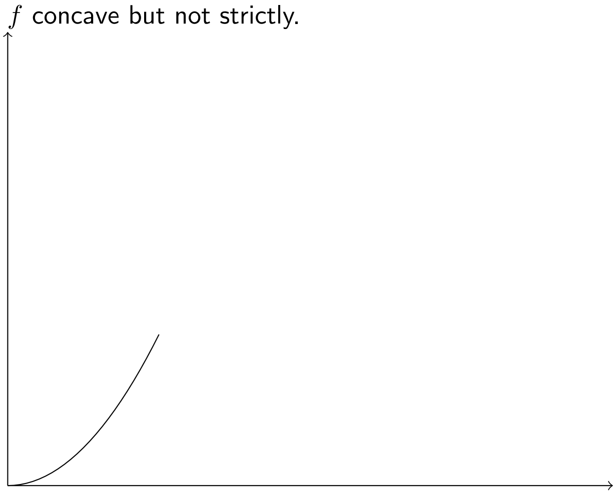
If the inequality is strict $\forall x \neq y$, then f is strictly concave. *A function f is (*strictly*) convex if $\leq$ ($<$) holds in place of $\geq$ ($>$).*



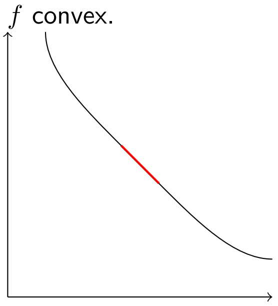
f concave



f concave but not strictly.



f convex.



f convex but not strictly.

Given a function, even relatively simple, it is not easy to ascertain if it is (quasi)concave or (quasi) convex. But if the function has additional properties (i.e. it is differentiable), then it is much easier.

Theorem 28. *The $\mathcal{C}^1$ function $f : U \in \mathbb{R}$, where $U \subset \mathbb{R}^n$ is convex, is concave iff*

$$f(y) \leq f(x) + Df(x)(y - x) \; \forall x, y \in U.$$

f is strictly concave iff the inequality holds strictly $\forall x \in U$ and $y \in U$ with $x \neq y$.

Proof. We show the "$\Rightarrow$" direction: $\forall t \in (0, 1)$:

$$f(ty + (1 - t)x) \geq tf(y) + (1 - t)f(x)$$

$$\Updownarrow$$

$$f(y) \leq \frac{1}{t} \left[f(ty + (1 - t)x) - (1 - t)f(x)\right]$$

$$\Updownarrow$$

$$f(y) \leq f(x) + \frac{f(x+t(y-x)) - f(x)}{t}$$

$$\Downarrow$$

$$f(y) \leq f(x) + \lim_{t\to 0} \frac{f(x+t(y-x)) - f(x)}{t}$$

But what it this limit? Let's introduce the following function:

$$g(t) := f(x + t(y-x)).$$

We can consider this as a curve, and write

$$g'(t) = Df(x + t(y-x))(y-x)$$

$$\Downarrow$$

$$g'(0) = Df(x)(y-x);$$

on the other hand:

$$\begin{aligned} g'(t) &= \lim_{h\to 0} \frac{g(t+h) - g(t)}{h} \\ &= \lim_{h\to 0} \frac{f(x+(t+h)(y-x)) - f(x+t(y-x))}{h} \\ \Rightarrow g'(0) &= \lim_{h\to 0} \frac{f(x+h(y-z)) - f(x)}{h} \\ &= \lim_{t\to 0} \frac{f(x+t(y-z)) - f(x)}{t}, \end{aligned}$$

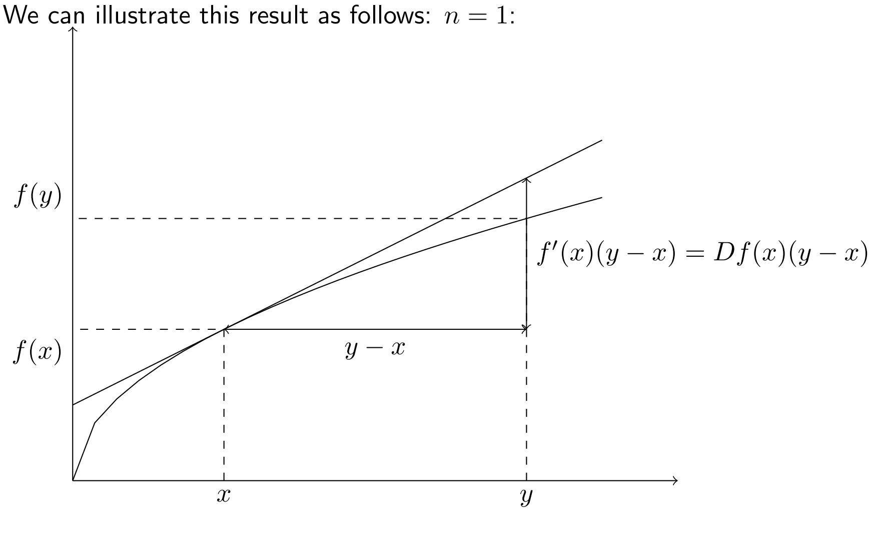
which is precisely the limit seen above. So:

$$\begin{aligned} f(y) \leq& f(x) + g'(0) \\ =& f(x) + Df(x)(y-x), \end{aligned}$$

which is precisely what we had claimed. □

In the proof, we see an interesting thing: another way to get the differential derivatives, by using the formula for directional derivatives but with a generic vector in place of e_i.

We can illustrate this result as follows: $n = 1$:



$f(y) - f(x) \leq Df(x)(y-x) \iff f(y) \leq f(x) + Df(x)(y-x)$.

However, what we saw so far still isn't a feasible approach to study concavity and convexity of functions. Such a method is given by the following:

Theorem 29. *The $\mathcal{C}^2$ function $f : U \in \mathbb{R}$, $U \subset \mathbb{R}^n$ convex, is concave iff $D^2 f(x)$ is negative semidefinite $\forall x \in U$.*

If $D^2 f(x)$ is negative definite $\forall x \in U$, then f is strictly concave.

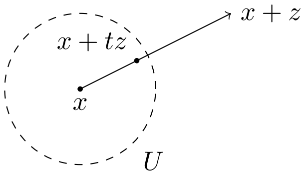
Recall that A is negative semidefinite (definite) if and only if:

$$z^T A z \leq 0 \qquad \forall z \in \mathbb{R}^n$$

$$(z^T A z < 0 \qquad \forall z \in \mathbb{R}^n, z \neq 0)$$

Proof. Again, we show "$\Rightarrow$" only.

Suppose f is concave and $x \in U$, $x \in \mathbb{R}^n$. Then, $\exists t > 0 : x + tz \in U$.



So we can define

$$\begin{aligned} g(t) :=& f(x + tz) \\ \Rightarrow g'(t) =& Df(x+tz)z \\ g''(t) =& z^T D^2 f(x+tz) z \end{aligned}$$

and using the Taylor expansion:

$$g(t) = g(0) + g'(0)t + \frac{1}{2} g''(c) t^2$$

for the "right c", $c \in (-t, t)$.

$$\Rightarrow f(x+tz) = f(x) + Df(x)z \cdot t + \frac{1}{2} z^T D^2 f(x+cz) z \cdot t^2$$

(still for some $c \in (-t, t)$).

The last equality can become

$$\frac{t^2}{2} z^T D^2 f(x+cz) z = f(\underbrace{x+tz}_{y}) - f(x) - Df(x) \underbrace{z \cdot t}_{y-x}.$$

The right hand side is composed by three terms that appeared in theorem 28. Using that theorem, we get that

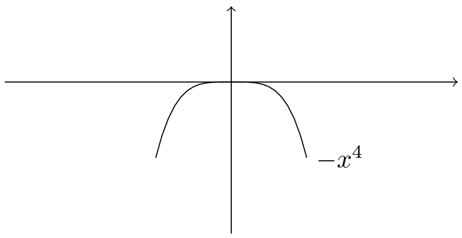
$$\begin{aligned} &\frac{t^2}{2} z^T D^2 f(x+cz) z \leq 0 \\ \Rightarrow & z^T D^2 f(x+cz) z \leq 0 \\ \Rightarrow & c \stackrel{t \to 0}{\to} 0 \\ \Rightarrow & z^T D^2 f(x) z \leq 0 \\ \stackrel{\text{def}}{\Rightarrow} & D^2 f(x) \text{ is negative semidefinite.} \end{aligned}$$

□

Hence, the Hessian matrix tells us if the matrix is negative semidefinite.

Example 30. *Special case $n = 1$: what does the above translate to?*
$D^2 f(x)$ negative semidefinite means $zf''(x)z \leq 0 \forall z \in \mathbb{R}, \forall x \in U$.
But then, z^2 is positive, so the above implies $f''(x) \leq 0 \forall x \in U$.
As we know, this is equivalent to f being a concave function.
If then $f''(x) < 0$, we get that f is strictly concave... but we do not claim the other direction! It is not true! See for instance:

$$f(x) = -x^4 \Rightarrow f'(x) = -4x^3, \; f''(x) = -12x^2 \Rightarrow f''(0) = 0.$$



$D^2 f(x)$ is negative semidefinite but not *negative definite, although f is strictly concave.*

Theorem 31 (Simon and Blume, p.382)**.** *$A \in \mathbb{R}^{n \times n}$ is negative definite iff*

$$|A_{2m}| > 0, |A_{2m+1}| < 0, \; m = 0, 1, \ldots$$

where A_k is the k-th order leading principal submatrix *and $|A_k|$ is the k-th order* leading principal minor.

$$A = \begin{bmatrix} & & \vdots & \vdots \\ & A_k & \vdots & \vdots \\ \cdots & \cdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \ddots \end{bmatrix}$$

$A_n = A$, $A_1 = [a_{11}]$

Example 32. *Let $u(x, y) = x^a y^b, \qquad a, b > 0$.*
Then,

$$D^2 u = \begin{bmatrix} a(a-1)x^{a-2}y^b & abx^{a-1}y^{b-1} \\ abx^{a-1}y^{b-1} & b(b-1)x^a y^{b-2} \end{bmatrix},$$

the Hessian, is negative definite iff:

- $m = 0 \Rightarrow A_{2m+1} = A_1 = \left[a(a-1)x^{a-2}y^b < 0\right]$*, and the determinant is the element itself, which must be < 0.*
- $m = 1 \Rightarrow A_{2m} = A_2 = A$*, and the determinant,*

$$a(a-1)x^{a-2}y^b b(b-1)x^a y^{b-2} - (abx^{a-1}y^{b-1})2,$$

must be positive.

Now, we can assume x and y positive, so the first condition already tells us $a < 1$, while the second can be seen as:

$$\begin{aligned} &abx^{2a-2}y^{2b-2}(ab-a-b+1) - a^2b^2x^{2a-2}y^{2b-2} \\ =&abx^{2a-2}y^{2b-2}(\underbrace{-a-b+1}_{>0}) > 0, \end{aligned}$$

so the second requirement is $a + b < 1$ (which also implies the first one).

Under those condition, $u(x, y) = x^a y^b$ is strictly concave.

16/11/2010

Unfortunately, the characterization of definitness with the principal submatrixes doesn't extend to semidefinitness by simply not requiring that the inequalities are strict: it's instead more complicated: all *principal submatrices* (not just the leading ones[8]) have to satisfy the property that the ones of dimension odd (even) have negative (positive) determinant.[9]

This condition is generally difficult to verify, but we can check it in the simple case seen last time, since the matrix is just 2×2:

$$a + b < 1,\ a, b > 0 \Rightarrow b < 1$$

$$u(x, y) = x^a y^b \text{ is (weakly) concave } \iff a + b \leq 1.$$

More important than concavity for us is another property, that we'll introduce after stating the following:

Theorem 33. *Let $f : U \to \mathbb{R}$ $U \subset \mathbb{R}^n$ convex, $x^0 \in U$ and $\alpha = f(x^0)$; then:*

$$f \text{ concave } \Rightarrow C^+_\alpha := \{x \in U | f(x) \geq \alpha\} \text{ is convex}$$

and

$$f \text{ convex } \Rightarrow C^-_\alpha := \{x \in U | f(x) \leq \alpha\} \text{ is convex .}$$

Proof. Let's take $x, y \in C^+_\alpha$. By definition of C^+_α, $f(x) \geq \alpha$, $f(y) \geq \alpha$. But then,

$$f(tx + (1-t)y) \overset{\text{concavity}}{\geq} tf(x) + (1-t)f(y) \geq t\alpha + (1-t)\alpha = \alpha,$$

which means exactly that

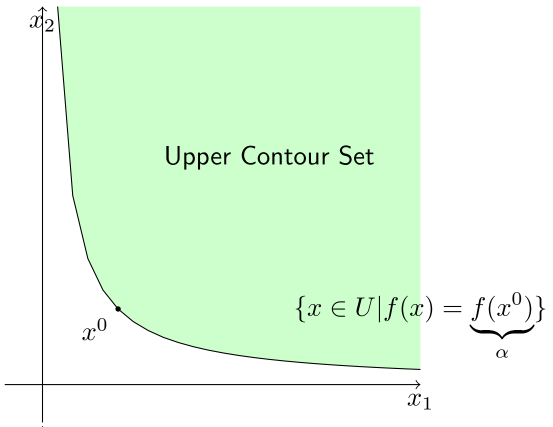
$$tx + (1-t)y \in C^+_\alpha.$$

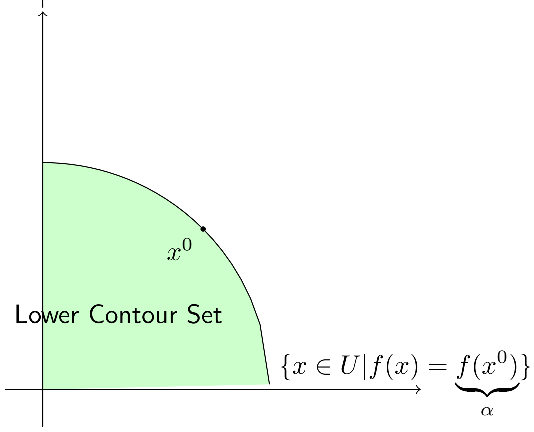
□

Let's illustrate this: take $n = 2$:

[8] If I understand correctly, the point is not "suppress just first row and first column", but more generally "suppress the *same* row and column"

[9] More informations on Simon and Blume





Example 34. $u(x, y) = x^a y^b$. *We know that* $a + b \leq (<) 1$ *if and only if* u *is (strictly) concave.*

On the other hand: consider $T : \mathbb{R} \to \mathbb{R}$ *strictly increasing; then* $f := T \circ u$ *may be non-concave, like in the case* $T(n) = n^c$, *with* $c > 0$, *such that* $(a + b)c > 1$

$$\Rightarrow f(x, y) = (x^a y^b)^c = x^{ac} y^{bc}$$

with $ac + bc = (a + b)c > 1$.

From elementary utility theory, we know that every monotone transformation of a given utility function represents the same preferences order.

From the last inequality, we get f *is not concave. Nevertheless, if we consider the set*

$$\{(x, y) \in U | \underbrace{f(x, y)}_{T(u(x,y))} \geq \alpha\} = \{(x, y) \in U | u(x, y) \geq T^{-1}(\alpha) =: \alpha' \in \mathbb{R}\} = C^+_{\alpha'},$$

which is convex when u *is concave.*

So basically the the upper contour set can be convex even without the function being concave (the implication given goes in one way). So what is the property of the function u such that the C^+_α are indeed convex? It's precisely what we define as *quasi-concavity*.

$$f \text{ concave} \quad \overset{\Rightarrow}{\not\Leftarrow} \quad C_\alpha^+ \text{ convex}$$

$$\Updownarrow$$

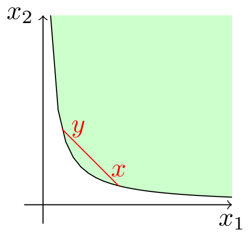
$$f \text{ quasi-concave}$$

Definition 35. *Let $f : U \in \mathbb{R}$, $U \subset \mathbb{R}^n$ convex. Then, f is* quasi-concave *if C_α^+ is convex for any α, that is,*

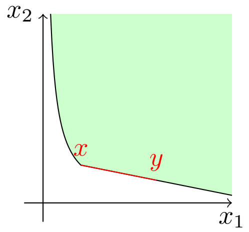
$$f(x) \geq \alpha, f(y) \geq \alpha \Rightarrow f(tx + (1-t)y) \geq \alpha \forall x, y \in U, \alpha \in \mathbb{R} \text{ and } t \in (0,1).$$

If the concluding inequality is strict whenever $x \neq y$, then f is strictly quasi-concave*.*

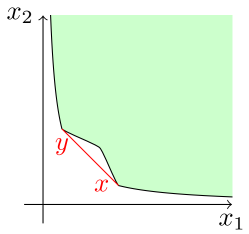
Let's illustrate this:



f strictly quasi-concave



f quasi-concave, but not strictly



f is not quasi-concave

Remark 36.

f quasi-concave

$\Updownarrow$

$$f(tx + (1-t)y) \geq \min\{f(x), f(y)\} \forall x, y, \in U, t \in (0,1) \quad (*)$$

Proof.

- "$\Rightarrow$": we know for sure that

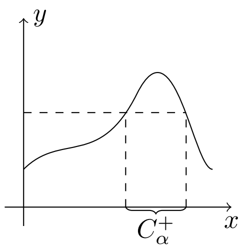
$$\begin{aligned} f(x) &\geq \min\{f(x), f(y)\} =: \alpha \\ f(y) &\geq \alpha. \end{aligned}$$

Now let's assume that f is quasi-concave. This implies that $f(tx+(1-t)y) \geq \alpha$.

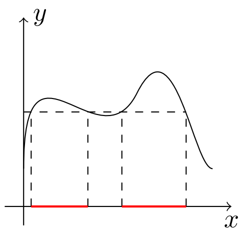
- "$\Leftarrow$": assume we know $f(x) \geq \alpha, f(y) \geq \alpha$. We need to show that the same holds for $f(tx + (1-t)y)$. Now, we know by assumption (see (*)) that this is $\geq \min\{f(x), f(y)\} \geq \alpha$.

□

This concept of quasi-concavity is a little bit intricated: how can we characterize a quasi-concave function's graph? It is easy in the case $n = 1$:



f quasi-concave



f not quasi-concave

Example 37. $u(x, y) = x^a y^b, a, b > 0$

$$u(x, y) = \alpha, \alpha > 0 \Rightarrow x^a y^b = \alpha \iff y = \frac{\alpha^{\frac{1}{b}}}{x^{\frac{a}{b}}} =: f(x)$$

The Cobb-Douglas is always *strictly quasi-concave. Indeed, let φ be the equation of a level curve:*

$$\varphi'(x) = -\frac{a}{b}\alpha^{\frac{1}{b}} x^{-\frac{a}{b}-1} < 0$$

$$\varphi''(x) = \underbrace{\left(-\frac{a}{b} - 1\right)}_{<0} \underbrace{\left(-\frac{a}{b}\right)}_{>0} \underbrace{\alpha^{\frac{1}{b}} x^{-\frac{a}{b}-2}}_{>0} > 0$$

$\Rightarrow \varphi$ *is strictly convex* $\Rightarrow u$ *is strictly quasi-concave.*

Theorem 38. *The $\mathcal{C}^1$ function $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$, is quasi-concave if and only if*

$$f(y) \geq f(x) \Rightarrow Df(x)(y - x) \geq 0 \qquad \forall x, y \in U.$$

Moreover, if

$$f(y) \geq f(x), y \neq x \Rightarrow Df(x)(y - x) > 0 \qquad \forall x, y \in U$$

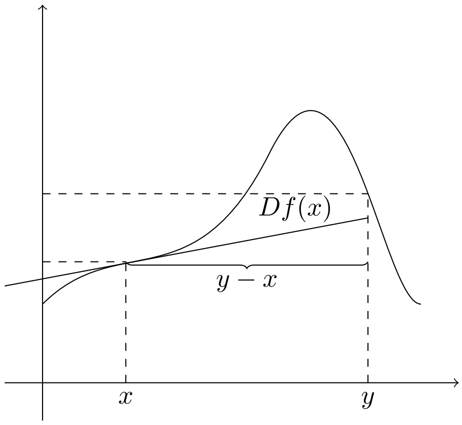
holds, then f is strictly quasi-concave.

Conversely, if f is strictly quasi-concave and $Df(x) \neq 0 \forall x \in U$, then

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

The proof is available in the mathematical appendix of the Mas-Colell WG, p. 934.

We will only draw an example with $n = 1$:



Theorem 39. *The C^2 function $f : U \to \mathbb{R}$ is quasi-concave iff for every $x \in U$ and $z \in \mathbb{R}^n$*

$$Df(x)z = 0 \Rightarrow z^T D^2 f(x)z \leq 0.$$

If $D^2 f(x)$ is negative definite in the subspace

$$\{z \in \mathbb{R}^n | Df(x)z = 0\}$$

for every $x \in U$, then f is strictly quasi-concave.

Again, the proof can be found on the Mas-Colell WG, pag. 935.

1.7 The Implicit Function Theorem

Consider the equations

$$\begin{aligned} f_1(y_1, \ldots, y_m, x_1, \ldots, x_n) =& c_1 \\ & \vdots \\ f_m(y_1, \ldots, y_m, x_1, \ldots, x_n) =& c_m. \end{aligned}$$

Let $y = (y_1, \ldots, y_m), x = (x_1, \ldots, x_n)$ and $c = (c_1, \ldots, c_m)$. Then, we can write:

$$\begin{pmatrix} f_1(y, x) \\ \vdots \\ f_m(y, x) \end{pmatrix} =: f(y, x) = c := \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

where

$$f : U \to \mathbb{R}^n, \qquad U \subset \mathbb{R}^{m+n}.$$

Then, we write

$$D_y f(x,y) := \underbrace{\begin{pmatrix} \frac{\partial f_1}{\partial y_1}(x,y) & \dots & \frac{\partial f_1}{\partial y_m}(x,y) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1}(x,y) & \dots & \frac{\partial f_m}{\partial y_m}(x,y) \end{pmatrix}}_{m\times m}$$

and

$$D_x f(x,y) := \underbrace{\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x,y) & \dots & \frac{\partial f_1}{\partial x_n}(x,y) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x,y) & \dots & \frac{\partial f_m}{\partial x_n}(x,y) \end{pmatrix}}_{m\times n},$$

the two parts of the Jacobian.

Let $(\overline{y}, \overline{x}) \in \mathbb{R}^{m+n}$ satisfy $f(y,x) = c$, that is $f(\overline{x}, \overline{y}) = c$.

We want to consider the following problem: can we find, for small variations of x around $\overline{x}$, values of y such that again $f(x,y) = c$?

In other words, does there exist a function

$$\mathbb{R}^n \supset \mathcal{B}(\overline{x}) \ni x \overset{g}{\mapsto} y \in \mathcal{B}(\overline{y}) \subset \mathbb{R}^n$$

such that

$$f(\underbrace{g(x)}_{\substack{\text{Implicit}\\ \text{function}}}, x) = c \ \forall x \in \mathcal{B}(\overline{x})?$$

Theorem 40 (Implicit Function Theorem)**.** *Let $U \subset \mathbb{R}^{m+n}$ be open and $f : U \to \mathbb{R}^m$ $\mathcal{C}^1$. If for a given $c \in \mathbb{R}^n$ the pair $(\overline{y}, \overline{x}) \in U$ is such that $f(\overline{y}, \overline{x}) = c$ and $\det(D_y f)(\overline{y}, \overline{x}) \neq 0$, then there exist open balls $\mathcal{B}(\overline{x}) \subset \mathbb{R}^n$ and $\mathcal{B}(\overline{y}) \subset \mathbb{R}^m$ and an unique function*

$$g = \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix} : \mathcal{B}(\overline{x}) \to \mathcal{B}(\overline{y})$$

such that $f(g(x), x) = c \ \forall x \in \mathcal{B}(\overline{x})$.

Moreover, g is $\mathcal{C}^1$ and

$$\underbrace{D_g(\overline{x})}_{m\times n} = -\underbrace{[D_y f(\overline{y}, \overline{x})]^{-1}}_{m\times m} \underbrace{D_x f(\overline{y}, \overline{x})}_{m\times n}.$$

Remark 41. *We may not be able to write down explicitly g - we just know it exists.*

Proof. We will not show rigorously: assume it indeed exists. Then,

$$\begin{array}{ccccccc} & & \mathbb{R}^n & & \mathbb{R}^{m+n} & & \mathbb{R}^m \\ & & \cup & & \cup & & \cup \\ \mathcal{B}(\overline{x}) & \ni & x & \xmapsto{(g,id)} & (g(x), x) & \xmapsto{f} & f(g(x), x) \end{array}$$
$$x \xrightarrow{\quad h \quad} f(g(x),x)$$

$$\Rightarrow h(x) = f(g(x), x)$$

which by the chain rule gives

$$Dh(\overline{x}) = Df(g(\overline{x}), \overline{x})D(g, id)(\overline{x})$$

where

$$id = \begin{pmatrix} id_1 \\ \vdots \\ id_n \end{pmatrix}, \qquad id_i(x) = x_i$$

$$\Rightarrow Dh(\overline{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \dots & \frac{\partial f_1}{\partial y_m} & \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & & & & \\ \frac{\partial f_m}{\partial y_1} & \dots & \frac{\partial f_m}{\partial y_m} & \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_m} \end{pmatrix} \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \dots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \dots & \dots & \frac{\partial g_m}{\partial x_n} \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

$$\begin{aligned} &= \Big(\underbrace{D_y f(\overline{y}, \overline{x})}_{m \times m} \quad \underbrace{D_y f(\overline{y}, \overline{x})}_{m \times n} \Big) \begin{pmatrix} Dg(\overline{x})\} \, m \times n \\ I\} \, n \times n \end{pmatrix} \\ &= D_y f(\overline{y}, \overline{x}) Dg(\overline{x}) + D_x f(\overline{y}, \overline{x}). \end{aligned}$$

On the other hand:

$$h(x) = c \qquad \forall x \in \mathcal{B}(\overline{x}) \Rightarrow Dh(\overline{x}) = \underbrace{0}_{m \times n},$$

so putting the two things together we get

$$\begin{aligned} &D_y f(\overline{y}, \overline{x}) Dg(\overline{x}) + D_x f(\overline{y}, \overline{x}) = 0 \\ \Rightarrow &D_y f(\overline{y}, \overline{x}) Dg(\overline{x}) + D_x f(\overline{y}, \overline{x}) = 0 \\ \Rightarrow &Dg(\overline{x}) = -[D_y f(\overline{y}, \overline{x})^{-1} D_x f(\overline{y}, \overline{x})]. \end{aligned}$$

□

17/11/10

Example 42 (The IS-LM model)**.** *This is the most well-known simple macroeconomic model. It is given by two equations:*

$$Y = C(Y - T) + I(r) + G \tag{IS}$$
$$M^S = PL(Y, r) \tag{LM}$$

where C is "consumption", Y is "incoming", T is "taxes", G is "governmental expenditure", r is "interest rate", M^S is "money supply", L is a liquidity demand function, P is "price level". Also,

$$\frac{\partial I}{\partial r} < 0 \qquad \frac{\partial L}{\partial r} < 0$$

so overall the effects take those directions:

$$\begin{aligned} Y =& C(\underbrace{Y - T}_{+}) + I(\underbrace{r}_{-}) + G \\ M^S =& PL(\underbrace{Y}_{+}, \underbrace{r}_{-}). \end{aligned}$$

Y *and* r *are endogenous variables; all the others are exogenous. Hence, with the notation of the Implicit Function Theorem,*

$$\begin{aligned} y =& (Y, r) \\ x =& (T, G, M^S, P) \end{aligned}$$

and

$$f(y, x) = \begin{pmatrix} f_1(Y, r, T, G, M^s, P) \\ f_2(Y, r, T, G, M^s, P) \end{pmatrix} = \begin{pmatrix} Y - C(Y - T) - I(r) - G \\ M^S - PL(Y, r) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Hence, we can define

$$\begin{aligned} Y =& g_1(T, G, M^S, P) \\ r =& g_2(T, G, M^S, P) \end{aligned}$$

and establish

$$f(g_1(\underbrace{T, G, M^S, P}_{x}), g_2(\underbrace{T, G, M^s, P}_{x}), \underbrace{T, G, M^S, P}_{x} = 0)$$

with

$$g : \mathbb{R}^4 \overset{\begin{pmatrix} g_1 \\ g_2 \end{pmatrix}}{\to} \mathbb{R}^2$$

Assuming that we have are in a given situation - a given combination of variables values which satisfies both equations - we may want to study comparative statics *- what is the reaction to (small) shocks:*

$$\begin{aligned} & \begin{pmatrix} \frac{\partial Y}{\partial T} & \frac{\partial Y}{\partial G} & \frac{\partial Y}{\partial M^S} & \frac{\partial Y}{\partial P} \\ \frac{\partial r}{\partial T} & \frac{\partial r}{\partial G} & \frac{\partial r}{\partial M^S} & \frac{\partial r}{\partial P} \end{pmatrix} \\ =& Dg(T, G, M^S, P) \\ =& -D_{(y,r)} f(Y, r, T, G, M^S, P)^{-1} \cdot D_{(T,G,M^S,P)} f(Y, r, T, G, M^S, P) \\ =& -\begin{pmatrix} 1 - C'(Y - T) & -I'(r) \\ -P\frac{\partial L}{\partial Y} & -P\frac{\partial L}{\partial R} \end{pmatrix}^{-1} \cdot \begin{pmatrix} C'(Y - T) & -1 & 0 & 0 \\ 0 & 0 & 1 & -L(Y, r) \end{pmatrix} \\ \overset{10}{=} & \end{aligned}$$

This is precisely where the condition on the determinant of the square matrix becomes important:

$$\det D_y f(y,x) = (\underbrace{1 - C'(Y-T)}_{+})\left(\underbrace{-P\frac{\partial L}{\partial r}}_{+}\right) - \underbrace{I'(r)}_{-}\,P\,\underbrace{\frac{\partial L}{\partial Y}}_{+} > 0,$$

so

$$\begin{aligned} D_g(T,G,M^S,P) &= -\frac{1}{\det}\begin{pmatrix} -P\frac{\partial L}{\partial r} & I'(r) \\ P\frac{\partial L}{\partial Y} & 1-C'(Y-t)\end{pmatrix}\cdot\begin{pmatrix} C'(Y-T) & -1 & 0 & 0 \\ 0 & 0 & 1 & -L(Y,r)\end{pmatrix} \\ \Rightarrow \frac{\partial Y}{\partial T} &= -\frac{1}{\det}(-P\frac{\partial L}{\partial r}C'(Y-T) + I'(r)\cdot 0) \\ &= \underbrace{\frac{1}{\det}}_{+}\cdot\underbrace{P\frac{\partial L}{\partial r}}_{-}\cdot\underbrace{C'(Y-T)}_{+} \le 0; \end{aligned}$$

this already explains that an increase in taxes causes a decrease in income.

A similar analysis can be applied to any other combination of variables.

But we still know nothing about the magnitude *of such effects. We can try to* estimate *the economic functions. More specifically, assume*

$$\begin{aligned} C(Y-T) =& 100 + 0.8(Y-T) \\ I(r) =& 500 - 50r \\ L(Y,r) =& 500 + 0.2Y - 25r; \end{aligned}$$

then

$$\det D_y f(y,x) = ((1-0.8)(-P(-25)) - (-50)P\cdot 0.2)) = 15P.$$

Hence,

$$\begin{aligned} Dg &= -\frac{1}{15P}\begin{pmatrix} 25P & -50 \\ 0.2P & 0.2\end{pmatrix}\cdot\begin{pmatrix} 0.8 & -1 & 0 & 0 \\ 0 & 0 & 1 & -L(Y,r)\end{pmatrix} \\ &= \begin{pmatrix} -\frac{4}{3} & \frac{5}{3} & \frac{10}{3P} & -\frac{10L(Y,r)}{3P} \\ -\frac{4}{375} & \frac{1}{75} & -\frac{1}{75P} & \frac{L(Y,r)}{75P}\end{pmatrix} \end{aligned}$$

Now, we can for instance assume

$$\begin{aligned} (\overline{Y},\overline{r}) =& (1200, 8.8) \\ (\overline{T},\overline{M}^S,\overline{P}) =& (400, 400, 520, 1), \end{aligned}$$

verify that $f(\overline{Y},\overline{r},\overline{T},\overline{G},\overline{M}^S,\overline{P}) = 0$ *and calculate*

$$D_g = \begin{pmatrix} -\frac{4}{3} & \frac{5}{3} & \frac{10}{3} & -\frac{5200}{3} \\ -\frac{4}{375} & \frac{1}{75} & -\frac{1}{75} & \frac{104}{15}\end{pmatrix}.$$

We now want to see a *geometric interpretation* of the Implicit Function Theorem.
Let $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^2$ open, with $(\overline{x}, \overline{y}) \in U$.
Consider the *level set* (or *level curve*)

$$L_f(f(\overline{x}, \overline{y})) := \{(x, y) \in U | f(x, y) = f(\overline{x}, \overline{y})\}.$$

Let $\frac{\partial f}{\partial y}(\overline{x}, \overline{y}) \neq 0$. Then, there exists an open interval $\mathcal{B}(\overline{x}) \subset \mathbb{R}$, an open interval $\mathcal{B}(\overline{y}) \subset \mathbb{R}$ and a function

$$g : \mathcal{B}(\overline{x}) \to \mathcal{B}(\overline{y})$$

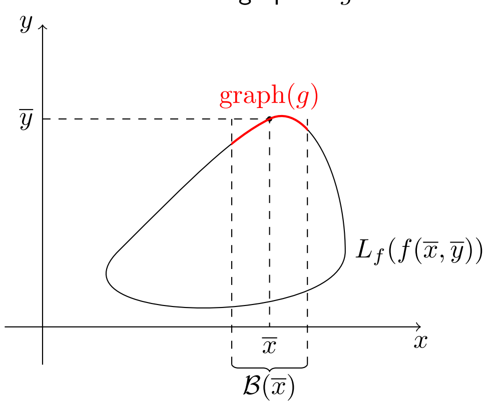
such that

$$f(x, g(x)) = f(\overline{x}, \overline{y}) \; \forall x \in \mathcal{B}(\overline{x}).$$

It is evident that the assumptions are the ones of the Implicit Function Theorem (only with x and y reversed to match their usual function in two dimensions).
We now want to study the graph of this function g:

$$\begin{aligned}
\mathrm{graph}(g) =& \{(x, g(x)) | x \in \mathcal{B}(\overline{x})\} \\
=& \{(x, y) | x \in \mathcal{B}(\overline{x}), y = g(x)\} \\
=& \{(x, y) | x \in \mathcal{B}(\overline{x}), y \in \mathcal{B}(\overline{y}) \text{ and } f(x, y) = f(\overline{x}, \overline{y})\} \\
\subset& \{(x, y) | (x, y) \in U, f(x, y) = f(\overline{x}, \overline{y})\} \\
=& L_f(f(\overline{x}, \overline{y})).
\end{aligned}$$

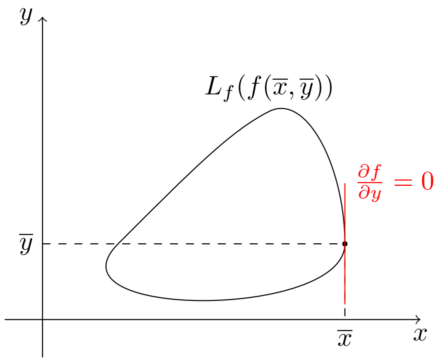
This means that the graph of g is a subset of the level curve for f:



What is the slope (inclination) of $L_f(f(\overline{x}, \overline{y}))$ at $(\overline{x}, \overline{y})$?
It is

$$\begin{aligned}
g'(\overline{x}) &= -\frac{\frac{\partial f}{\partial x}(\overline{x}, \overline{y})}{\frac{\partial f}{\partial y}(\overline{x}, \overline{y})} \\
&= -\left(D_y f(\overline{x}, \overline{y})\right)^{-1} D_x f(\overline{x}, \overline{y}).
\end{aligned}$$

Can we apply this formula to all points? Not really:



in this point, we have that the slope is infinite - it cannot be the graph of a function from x to y, and indeed the condition on the invertibility (which in this case is simply "difference from 0") is not respected.

Still in the case of two variables, consider a function

$$h(x) = f(x, g(x)) = f(\overline{x}, \overline{y}) \Rightarrow h'(x) = 0 \overset{(*)}{=} \frac{\partial f}{\partial x}(\overline{x}, g(\overline{x})) + \frac{\partial f}{\partial y}(\overline{x}, g(\overline{x}))g'(\overline{x}),$$

where $(*)$ is a simple application of the chain rule.

The same can be rewritten as:

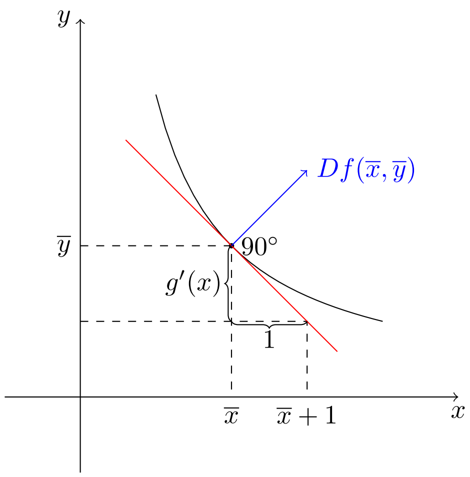
$$h'(\overline{x}) = \begin{pmatrix} \frac{\partial f}{\partial x}(\overline{x}, g(\overline{x})) & \frac{\partial f}{\partial y}(\overline{x}, g(\overline{x})) \end{pmatrix} \begin{pmatrix} 1 \\ g'(\overline{x}) \end{pmatrix} = Df(\overline{x}, \overline{y}) \begin{pmatrix} 1 \\ g'(\overline{x}) \end{pmatrix} = 0$$

Now, we recall that

$$x \cdot y = \|x\|\|y\| \cos\theta \Rightarrow \cos\theta = 0$$

and that happens for $\theta = \pm\frac{\pi}{2} = 90°$: x and y are orthogonal.

Coming back to our example, $Df(\overline{x}, \overline{y})$ and $(1, g'(\overline{x}))$ are orthogonal:



So the gradient vector $Df(\overline{x}, \overline{y})$ (or $\nabla f(\overline{x}, \overline{y})$) is orthogonal to the level curve passing through the point $(\overline{x}, \overline{y})$.

This is a general result.

18/11/10

Alternative method:

$$\begin{aligned} f_1(y_1,\ldots,y_m,x_1,\ldots,x_n) =& c_1 \\ &\vdots \\ f_m(y_1,\ldots,y_m,x_1,\ldots,x_n) =& c_m \end{aligned}$$

can be linearly approximated as:

$$\begin{aligned} \frac{\partial f_1}{\partial y_1}dy_1 + \cdots + \frac{\partial f_1}{\partial y_m}dy_m + \frac{\partial f_1}{\partial x_1} + \cdots + \frac{\partial f_1}{\partial x_n}dx_n =& 0 \\ &\vdots \\ \frac{\partial f_m}{\partial y_1}dy_1 + \cdots + \frac{\partial f_m}{\partial y_m}dy_m + \frac{\partial f_m}{\partial x_1} + \cdots + \frac{\partial f_m}{\partial x_n}dx_n =& 0. \end{aligned}$$

This last expression can be rewritten in a more concise way:

$$\underbrace{\underbrace{D_y f(\overline{y},\overline{x})}_{m\times m} \underbrace{\begin{pmatrix} dy_1 \\ \vdots \\ dy_n \end{pmatrix}}_{m\times 1}}_{m\times 1} + \underbrace{\underbrace{D_x f(\overline{y},\overline{x})}_{m\times n} \underbrace{\begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix}}_{n\times 1}}_{m\times 1} = \underbrace{0}_{m\times 1}$$

If $D_y f(\overline{y},\overline{x})$ is invertible, then

$$\begin{pmatrix} dy_1 \\ \vdots \\ dy_m \end{pmatrix} = -\left(D_y f(\overline{y},\overline{x})\right)^{-1} D_x f(\overline{y},\overline{x}) \begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix}.$$

In particular, if $dx_i = 0 \quad \forall i \neq k$, I can obtain some direct result without doing all the calculations:

$$\begin{pmatrix} dy_1 \\ \vdots \\ dy_n \end{pmatrix} = -\left(D_y f(\overline{y},\overline{x})\right)^{-1} \begin{pmatrix} \frac{\partial f_1}{\partial x_k}(\overline{y},\overline{x})dx_k \\ \vdots \\ \frac{\partial f_m}{\partial x_k}(\overline{y},\overline{x})dx_k \end{pmatrix}.$$

Example 43. *Again the IS-LM model:*

$$\begin{aligned} \begin{pmatrix} dY \\ dr \end{pmatrix} =& -\begin{pmatrix} \frac{\partial f_1}{\partial Y} & \frac{\partial f_1}{\partial r} \\ \frac{\partial f_2}{\partial Y} & \frac{\partial f_2}{\partial r} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial f_1}{\partial T}dT \\ \frac{\partial f_2}{\partial T}dT \end{pmatrix} \\ =& -\frac{1}{\det} \begin{pmatrix} -p\frac{\partial L}{\partial r} & I'(r) \\ p\frac{\partial L}{\partial Y} & 1 - C'(Y-t) \end{pmatrix} \begin{pmatrix} C'(Y-T)dT \\ 0 \cdot dT \end{pmatrix} \\ \Rightarrow dY =& \frac{1}{\det} p \frac{\partial L}{\partial r} C'(Y-T)dT \\ =& -\frac{4}{3}dT. \end{aligned}$$

Indeed, we can verify that $\frac{\partial Y}{\partial T} = -\frac{4}{3}$, as we had already seen.

Some authors will always use the Implicit Function Theorem, other will proceed as we just did.

Even if we are interested in the effect of only one variable, we still have to calculate *all* the matrix to invert: this is an intrinsec need, because of *feedback*: variables have an effect one on the other.

2 Static optimization

2.1 Unconstrained optimization

Let

$$f : U \in \mathbb{R}, \qquad U \subset \mathbb{R}^n.$$

Then, x^* is a

- *local maximizer (minimizer)* if there is an open ball $\mathcal{B}(x*) \subset U$ such that

$$f(x^*) \overset{(\leq)}{\geq} f(x) \; \forall x \in \mathcal{B}(x^*);$$

- *global maximizer (minimizer)* if

$$f(x^*) \overset{(\leq)}{\geq} f(x) \; \forall x \in U.$$

Theorem 44. *Suppose f is $\mathcal{C}^1$ and $x^* \in \mathbb{R}^n$ is a local maximizer or minimizer of f. Then,*

$$\frac{\partial f}{\partial x_i}(x^*) = 0 \qquad \forall i, \ldots, n$$

or, in more concise notation,

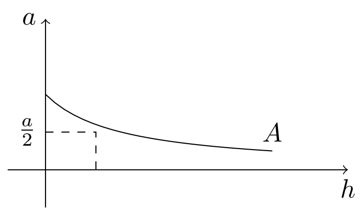
$$Df(x^*) = 0 (\in \mathbb{R}^n).$$

Proof. Suppose x^* is a local maximizer but contrary to our claim

$$\frac{\partial f}{\partial x_i}(x^*) = a > 0$$

for some i. Then, for h sufficiently small we get

$$A(h) := \frac{f(x^* + he^i) - f(x^*)}{h} > \frac{a}{2}.$$

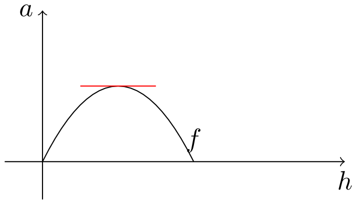


So,

$$f(x^* + he^i) > f(x^*) + \frac{a}{2}h > f(x^*).$$

This implies that x^* is not a local maximizer: we have a contradiction. $\square$

In the one-dimensional case, it is quite intuitive:



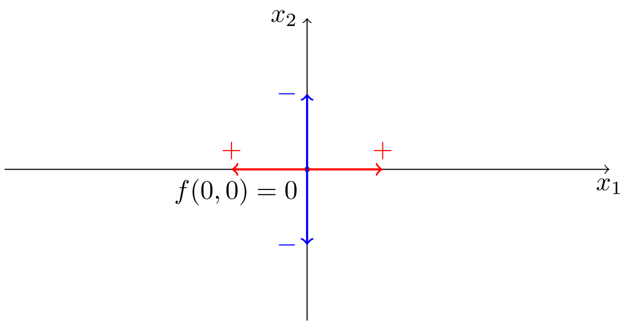
Definition 45.

$$Df(x) = 0 \iff x \textit{ is a } \text{critical point}.$$

By the way, x^* maximizer $\Rightarrow$ x^* is a critical point, but the opposite implication is not true.

Example 46.

$$\begin{aligned} & f(x_1, x_2) = x_1^2 - x_2^2 \\ \Rightarrow & Df(x_1, x_2) = (2x_1, -2x_2) \\ \Rightarrow & Df(0,0) = (0,0) \\ \Rightarrow & (0,0) \textit{ is a critical point for } f. \end{aligned}$$



Since in every neighborhood of $(0,0)$ *we can easily find point on which* f *is smaller or larger, that is neither a local maximum nor a minimum.*

Theorem 47. *Suppose that* f *is* $\mathcal{C}^2$ *and* $Df(x^*) = 0$.

(i) If x^* *is a local maximizer, then* $D^2 f(x^*)$ *is negative semidefinite.*

(ii) If $D^2 f(x^*)$ *is negative definite, then* x^* *is a local maximizer.*

Proof. Let $x \in \mathbb{R}^n$ and

$$\varphi(\varepsilon) := f(x^* + \varepsilon z), \qquad \varepsilon \in \mathbb{R} \text{ small}$$

$$\Rightarrow \begin{cases} \varphi'(\varepsilon) = & Df(x^* + \varepsilon z)z \\ \varphi''(\varepsilon) = & z^T D^2 f(x^* + \varepsilon z) \end{cases}$$

We can hence rewrite φ using its Taylor expansion:

$$\varphi(\varepsilon) = \varphi(0) + \varphi'(0)\varepsilon + \frac{1}{2}\varphi''(0)\varepsilon^2 + R(\varepsilon),$$

and we know that

$$\frac{R(\varepsilon)}{\varepsilon^2} \overset{\varepsilon \to 0}{\to} 0.$$

So,

$$f(x^* + \varepsilon z) = f(x^*) + Df(x^*)z\varepsilon + \tfrac{1}{2} z^T D^f(x^*) z\varepsilon^2 + R(\varepsilon)$$

$$\Updownarrow$$

$$f(x^* + \varepsilon z) - f(x^*) \overset{(*)}{=} \tfrac{1}{2} z^T D^2 f(x^*) z\varepsilon^2 + R(\varepsilon)$$

Moreover, we know that the left hand side is ≤ 0 if x^* is a maximizer. So:

$$\frac{1}{2} z^T D^2 f(x^*) z\varepsilon^2 + R(\varepsilon) \leq 0,$$

which in turn we can rewrite as follows:

$$z^T D^2 f(x^*) z + \frac{2}{\varepsilon^2} R(\varepsilon) \leq 0.$$

Now: what happens if we let ε tend to 0? We know that the second term tends to 0 too. So:

$$\begin{aligned} & z^T D^2 f(x^*) z \leq 0 \\ \Rightarrow & D^2 f(x^*) \text{ is negative semidefinite.} \end{aligned}$$

We have shown part 47 of the theorem. If we now start from the hypothesis that the Hessian is negative semidefinite, vice versa:

$$\begin{aligned} & z^T D^2 f(x^*) z < 0 \qquad \forall z \neq 0 \\ \Rightarrow & \exists \varepsilon > 0 : z^T D^2 f(x^*) z + \frac{2}{\varepsilon^2} R(\varepsilon) < 0 \\ \overset{(*)}{\Rightarrow} & f(x^* + \varepsilon z) < f(x^*) \\ \Rightarrow & x^* \text{ is a local maximizer.} \end{aligned}$$

This proves 47.

□

So:

$$\begin{gathered} D^2 f(x^*) \text{ negative definite} \quad . \\ \Big\Downarrow (ii) \\ x^* \text{ local maximizer} \\ \Big\Downarrow (i) \\ D^2 f(x^*) \text{ negative semidefinite} \end{gathered}$$

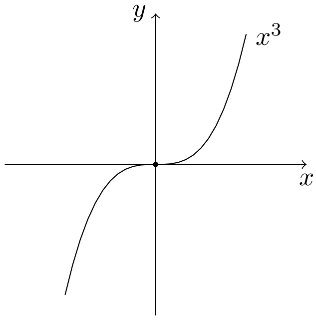
We know the first implication cannot be reversed: neither it is possible with the second one.

Example 48.

$$f(x) = x^3$$

$$\Rightarrow \begin{cases} Df(x) = & 3x^2 \\ D^2 f(x) = & 6x \Rightarrow D^2 f(0) = 0 \end{cases} \Rightarrow D^2 f(0) \textit{ is negative semidefinite because } zD^2 f(0) z = 0 \leq 0.$$

But: $x = 0$ *is neither a maximum nor a minimum:*



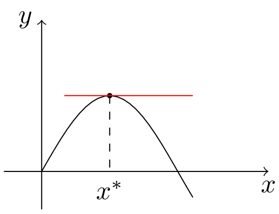
Theorem 49. *Any critical point* x^* *of a concave function is a global maximizer.*

Proof. A previous theorem on concave functions told us that:

$$f(x) \leq f(x^*) + \cancel{Df(x^*)}(x - x^*) \qquad \forall x \in \operatorname{dom} f$$

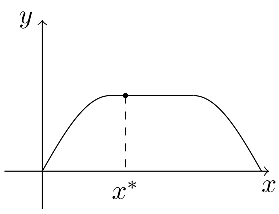
$$\Updownarrow$$

$$f(x) \leq f(x^*) \qquad \forall x \in \operatorname{dom} f$$



$\square$

Moreover, f strictly concave $\Rightarrow$ such an x^* is unique; we don't have something like



2.2 Constrained optimization

Start with equality constraint:

$$\max f(x)$$

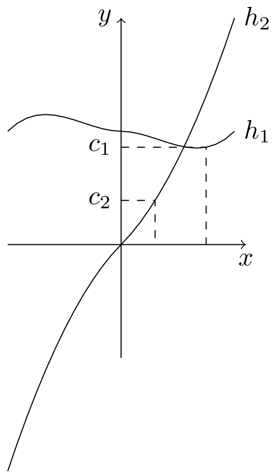
$$\text{subject to} \quad \left.\begin{array}{c} h_1(x) = c_1 \\ \vdots \\ h_m(x) = c_m \end{array}\right\} h(x) = c \iff c \in C$$

where $f, h_1, \ldots, h_m : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$, $n \geq m$ and

$$C := \{x \in U | h_j(x) = c_j \; \forall j = 1, \ldots, m\}$$

is the *constraint set*.

For instance:



is not a valid example, since C is empty.

Definition 50. $x^* \in C$ *is a* local constrained maximizer *if there is an open ball* $\mathcal{B}(x^*) \subset U$ *such that*

$$f(x^*) \geq f(x) \; \forall x \in \mathcal{B}(x^*) \cap C.$$

Definition 51. $x^* \in C$ *is a* global constrained maximizer *if*

$$f(x^*) \geq f(x) \; \forall x \in U \cap C.$$

Theorem 52. *Suppose $f, h_1, \ldots, h_m$ are $\mathcal{C}^1$, x^* is a local constrained maximizer and $Dh_1(x^*), \ldots, Dh_m(x^*)$ are linearly independent vectors (* "constraints qualification" *condition, or CQ).*

Then, there exists $\mu_1, \ldots, \mu_n \in \mathbb{R}$ ("Lagrange multipliers") such that

$$\frac{\partial f}{\partial x_i}(x^*) = \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i}(x^*) \qquad \forall i = 1, \ldots, n$$

or, in more concise notation,

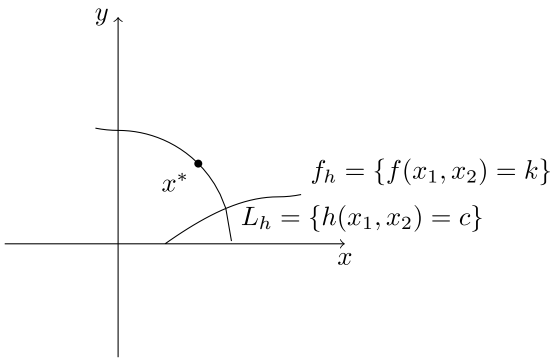
$$Df(x^*) = \sum_{j=1}^{m} \mu_j Dh_j(x^*).$$

Let's try to illustrate why these numbers should exist.

Proof. The complete proof can be found on Mas-Colell Winston Green, pp. 956-957.

Here, we only consider the special case $m = 1$, $n = 2$. We have one constraint only, on two variables:

$$\max_{h(x_1,x_2)=c} f(x_1, x_2).$$



x^* is the maximizer.

The slope of h_g at x^* is equal to the slope of L_h:

$$-\frac{\frac{\partial f}{\partial x_1}(x^*)}{\frac{\partial f}{\partial x_2}(x^*)} = -\frac{\frac{\partial h}{\partial x_1}(x^*)}{\frac{\partial h}{\partial x_2}(x^*)}$$

$$\Rightarrow \exists \mu \in \mathbb{R} : \frac{\partial f}{\partial x_1}(x^*))\mu \frac{\partial h}{\partial x_1}(x^*)$$

and for the same μ:

$$\frac{\partial f}{\partial x_2}(x^*) = \mu \frac{\partial h}{\partial x_2}(x^*).$$

So

$$Df(x^*) = \mu Dh(x^*);$$

this is exactly what we wanted to show.

□

Back to the general case, or $x = (x_1, \ldots, x_n) \in \operatorname{dom} f$ and $\mu = (\mu_1, \ldots, \mu_m) \in \mathbb{R}^m$: define the *Lagrange function*

$$L(x, \mu) := f(x) - \sum_{j=1}^{m} \mu_j [h_j(x) - c_j].$$

Then, we can observe the following thing:

$$Df(x^*) = \textstyle\sum_{j=1}^{m} \mu_j Dh_j(x^*)$$
$$\Updownarrow$$
$$D_x L(x^*, \mu) = 0.$$

Moreover, we can also notice that

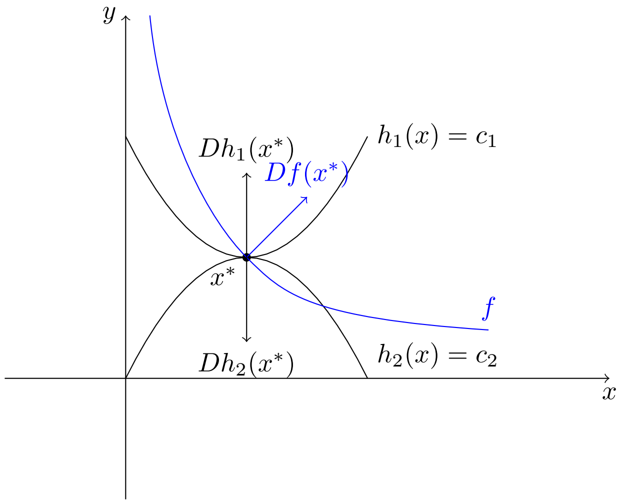
$$h(x^*) = c \iff D_\mu L(x^*, \mu) = 0.$$

Finally, if we want to form the Lagrange conditions, we need just to write the Lagrange function and put all its partial derivatives equal to 0.

Let's now discuss the meaning of the Constraint Qualification.

We require that all the gradients are linearly independent.

Example 53. *We take $n = m = 2$.*



In this case, $C = \{x^*\}$ and hence x^* is necessarily a solution to

$$\max_{\substack{h_1(x)=c_1,\\ h_2(x)=c_2}} f(x),$$

but $\nexists \mu_1, \mu_2 \in \mathbb{R}$ such that

$$Df(x^*) = \mu_1 Dh_1(x^*) + \mu_2 Dh_2(x^*)$$

because $Dh_1(x^*)$ and $Dh_2(x^*)$ are not linearly independent: CQ does not hold.

For the records, this is very rare in economic applications.

So far, we considered constrained maximizations with only *equality* constraints: we now face the problem of also *inequalities*:

$$\max f(x) \tag{P}$$

subject to

$$\left.\begin{array}{c} \left.\begin{array}{c} g_1(x) \le b_1 \\ \vdots \\ g_k(x) \le b_k \end{array}\right\} g(x) \le b \\ \left.\begin{array}{c} h_1(x) = c_1 \\ \vdots \\ h_m(x) = c_m \end{array}\right\} h(x) = c \end{array}\right\} \stackrel{\text{def}}{\iff} x \in C$$

where $f, g_1, \ldots, g_k, h_1, \ldots, h_m : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$, $n \ge k + m$, $k, m \ge 0$.

Now the CQ takes the following form:

those constraints that hold at x^* with equality give rise to linearly independent gradients,

that is, the vectors

$$\{Dg_j(x^*)|g_j(x^*) = b_j\} \cup \{Dh_j(x^*)|j = 1, \ldots, m\}$$

are linearly independent.

That said, we can state the following:

Theorem 54 (Kuhn-Tucker)**.** *Suppose that x^* is a solution to (P) and CQ is satisfied at x^*. Then there are multipliers $\lambda \geq 0$, one for each inequality constraint, and $\mu_j \in \mathbb{R}$, one for each equality constraint, such that:*

1. *For every $i = 1, \ldots, n$:*

$$\frac{\partial f}{\partial x_i}(x^*) = \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i}(x^*) + \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i}(x^*)$$

or, in more concise notation,

$$Df(x^*) = \sum_{j-1}^{k} \lambda_j Dg_j(x^*) + \sum_{j=1}^{m} \mu_j Dh_j(x^*),$$

2. *For every $j = 1, \ldots, k$*

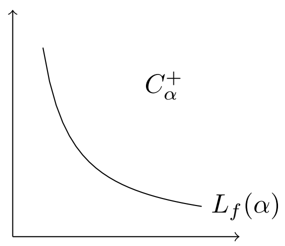
$$\underbrace{\lambda_j}_{\geq 0} \underbrace{[g_j(x^*) - b_j]}_{\leq 0} = 0$$

(this second condition is called "complementary slackness").

It's easy to see that for $k = 0$ we fall back to the formulae for the case with no inequalities.

22/11/10

Let's take a function f of 2 variables. We know that f is quasi-concave $\iff$ C_α^+ is convex. But this is equivalent to $L_f(\alpha)$ being convex:



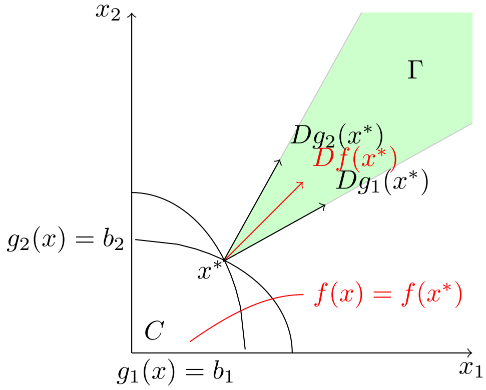
Indeed,

$$\begin{aligned} &f(x, y) = \alpha \\ \Rightarrow &y = \varphi(x, \alpha) \\ &y'' > 0 \Rightarrow \varphi \text{ convex} \iff L_f(\alpha) \text{ convex} \end{aligned}$$

We had already seen that in the case of the Cobb Douglas. This method is a very special case (the function must be in 2 variables *and* the curve level must be representable as a function from one to the other), but it is a special case that will be very useful in our economic activity - it typically applies to utility functions for 2 goods.

Regarding the equalities and disequalities problem formulated last time, it is evident that the choice of $\leq$ instead than $\geq$ doesn't imply any loss of generality - it is sufficient to multiply by -1.

We want to understand *why* the conditions given are reasonable and what they mean. We provide a sketch of the proof, illustrating the case $n = 2$, $k = 2$, $m = 0$.



In this situation:

- x^* is the maximum
- $Dg_1(x^*)$ and $Dg_2(x^*)$ span the *cone*

$$\Gamma := \{x \in \mathbb{R}^2 | \exists \lambda_1, \lambda_2 \geq 0 \text{ with } x = \lambda_1 Sg_1(x^*) + \lambda_2 Dg_2(x^*)$$

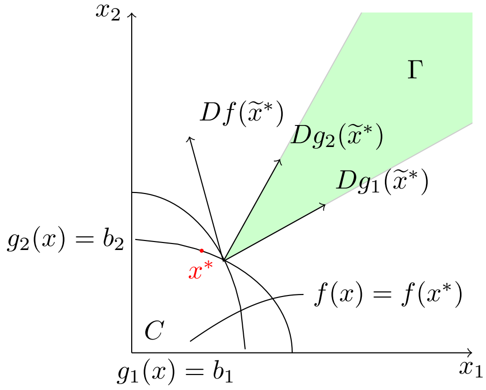
- $Df(x^*) \in \Gamma$

Hence, $Df(x^*) = \lambda_1 Dg_1(x^*) + \lambda_2 Dg_2(x^*)$ for some $\lambda_1, \lambda_2 \geq 0$. This is precisely what we claim with the Kuhn-Tucker theorem.

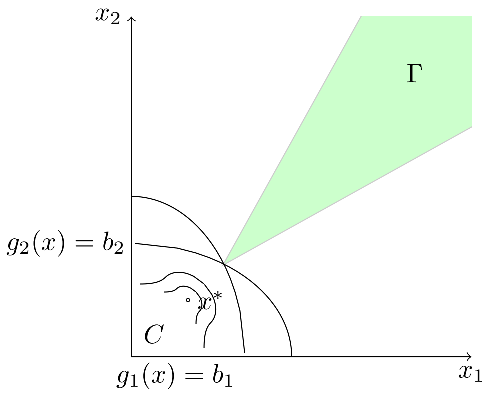
In another very special case, consider

$$\max f(x) \qquad \text{s.t. } g(x) = b;$$

if $D_f(\tilde{x}^*) \notin \Gamma$, then we can "move" a bit in its direction, remaining on the level curve of $g(x)$:



$Df(\widetilde{x}^*) \notin \Gamma$
$\Rightarrow$ it is not a maximizer. x^* is, and hence by Kuhn-Taker we have the multipliers.
There is a third possibility we must consider:



x^* is the maximizer.

$$Df(x^*) = \underbrace{\lambda_1}_{=0} Dg_1(x^*) + \underbrace{\lambda_2}_{=0} Dg_2(x^*) = 0$$

This is an *unconstrained* maximization problem.
Kuhn-Tucker just joins all the special cases in a single theorem.
Let's get back to the general case, with the Lagrange function:

$$L(x, \lambda_1, \ldots, \lambda_K, \mu_1, \ldots, \mu_m) := f(x) - \sum_{j=1}^{k} \lambda_j [g_j(\lambda) - b_j] - \sum_{j=1}^{m} \mu_j [h_j(x) - c_j]$$

$$\Rightarrow \frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{k} \lambda_j \frac{\partial g_i}{\partial x_i} - \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i} = 0$$

and

$$\lambda_j[g_j(x) - b_j] = 0,\ \lambda_j \geq 0,\ g_j(x) - b_j \leq 0\ \forall j = 1, \ldots, k$$
$$h_j(x) = c_j \quad \forall j = 1, \ldots, m.$$

All these together provide the set of Kuhn-Tucker (necessary) conditions.

There is an interesting special case of inequality constraints: *non-negativity constraints*, which is very typical in economics (for instance we don't want prices to be negative).

$$x_i \geq 0 \text{ for some (/for all) } i \iff -x_i \leq 0$$

(where the second line fits in the requirements of the theorem, by just putting, for some j,

$$g_j(x) = -x_i$$

with $b_j = 0$). However, it is interesting to transform this:

$$\begin{aligned}\Rightarrow &- \lambda_j[g_j(x) - b_j] \\ &- \lambda_j(-x_i) = \lambda_j x_i.\end{aligned}$$

Moreover, if we rename $\lambda_j =: \nu_i \Rightarrow \nu_i x_i$,

$$\begin{aligned}\Rightarrow &L(x, \lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_m, \nu_1, \ldots, \nu_n) \\ &= f(x) - \sum_{j=1}^{k} \lambda_j[g_j(x) - b_j] - \sum_{j=1}^{m} \mu_j[h_j(x) - c_j] + \sum_{i=1}^{n} \nu_i x_i\end{aligned}$$

we can express the Kuhn-Tucker conditions as:

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i} - \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i} + \nu_i = 0 \qquad \forall i = 1, \ldots, n \tag{1}$$

(under the assumption that *all* of the variables are subject to negativity constraints) and, for what concerns the complementary slackness condition,

$$\nu_i x_i = 0,\ \nu_i \geq 0,\ x_i \geq 0 \qquad \forall i = 1, \ldots, n \tag{2}$$

(recall $g_j = 0$).
Moreover, (1) and (2) are equivalent to

$$\begin{aligned}\frac{\partial d}{\partial x_1} \leq \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i} + \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i}&, \\ x_i \geq 0&\end{aligned} \tag{3}$$

with "$\leq$" being "$=$" when $x_i > 0$.

Both ways of formulating those conditions are used by different authors. And this is not the end of the story, since there is a further way of expressing this: (3) and

$$\left(\frac{\partial f}{\partial x_i} - \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i} - \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i} \right) x_i = 0 \qquad \forall i = 1, \ldots, n.$$

Example 55.

$$\max f(x_1, x_2) = x_1 + x_2^{\alpha}, \alpha > 0$$

$$\begin{aligned}\textit{s.t. } p_1 x_1 + p_2 x_2 &\leq I \\ x_1 \geq 0,\ x_2 &\geq 0\end{aligned}$$

We explicitly write

$$L(x_1, x_2, \lambda, \nu_1, \nu_2) = x_1 + x_2^{\alpha} - \lambda(p_1 x_1 + p_2 x_2 - I) + \nu_1 x_1 + \nu_2 x_2$$

and from this function we get the conditions

$$\Rightarrow \frac{\partial L}{\partial f_1} = 1 - \lambda p_1 + \nu_1 = 0 \quad (1)$$
$$\frac{\partial L}{\partial x_2} = \alpha x_2^{\alpha-1} - \lambda p_2 + \nu_2 = 0 \quad (2)$$
$$\nu_1 x_2 = 0 \quad (3)$$
$$\nu_2 x_2 = 0 \quad (4)$$
$$\lambda(p_1 x_1 + p_2 x_2 - I) = 0$$
$$\lambda, \nu_1, \nu_2 \geq 0,\ x_1, x_2 \geq 0$$
$$p_1 x_1 + p_2 x_2 - I \leq 0.$$

Solving this system is not immediately trivial. Can we simplify something?

$$(1) \Rightarrow \lambda = \frac{1+\nu_1}{p_1} > 0$$
$$\Rightarrow p_1 x_1 + p_2 x_2 = I. \quad (5)$$

We can hence consider various cases.

case	ν_1	ν_2	λ	(x_1, x_2)	$f(x_1, x_2)$
1	0	0	$\frac{1}{8}$	$\left(\frac{11}{8}, \frac{1}{4}\right)$	$\frac{23}{16}$
2	$+$	0	$\frac{3}{2}$	$(0, 3)$	9
3	0	$+$	$\frac{1}{8}$	$\left(\frac{3}{2}, 0\right)$	$\frac{3}{2}$
4	$+$	$+$	/	/	/

Assume: $\alpha = 2$, $p_1 = 8$, $p_2 = 4$, $I = 12$.

- Case 1*:* $\nu_1 = \nu_2 = 0$

$$(1) \Rightarrow \lambda = \frac{1}{p_1} = \frac{1}{8} \overset{(2)}{\Rightarrow} 2x_2 = \frac{1}{8} \cdot 4 = \frac{1}{2} \Rightarrow x_2 = \frac{1}{4}$$
$$\overset{(5)}{\Rightarrow} 8x_1 + 4\frac{1}{4} = 12 \Rightarrow x_1 = \frac{11}{8}$$
$$\Rightarrow f(x,y) = \frac{11}{8} + \left(\frac{1}{4}\right)^2 = \frac{22+1}{16} = \frac{23}{16}$$

This is a possible candidate. *We'll now examine if there are others.*

- Case 2*:* $\nu_1 > 0$, $\nu_2 = 0$.

$$(3) \Rightarrow x_1 = 0 \overset{(5)}{\Rightarrow} 4x_2 = 12 \Rightarrow x_2 = 3$$
$$\overset{(2)}{\Rightarrow} 2 \cdot 3 = 4\lambda \Rightarrow \lambda = \frac{3}{2} > 0$$
$$\overset{(1)}{\Rightarrow} \nu_1 = \frac{3}{2} \cdot 8 - 1 = 12 - 1 = 11$$
$$\Rightarrow f(x,y) = 0 + 3^2 = 9$$

$9 > \frac{23}{16} \Rightarrow$ *this candidate is already better than the first one.*

- Case 3: $\nu_1 = 0$, $\nu_2 > 0$.

$$(4) \Rightarrow x_2 = 0 \overset{(5)}{\Rightarrow} 8x_1 = 12 \Rightarrow x_1 = \frac{12}{8} = \frac{3}{2}$$

$$\overset{(1)}{\Rightarrow} 8\lambda_1 = 1 \Rightarrow \lambda = \frac{1}{8} > 0$$

$$\overset{(2)}{-} \frac{1}{8} \cdot 4 + \nu_2 = 0 \Rightarrow \nu_2 = \frac{1}{2}$$

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

This is uninteresting. But there is still one case to consider.

- Case 4: $\nu_1 > 0$, $\nu_2 > 0$.

$$(3) \Rightarrow x_1 = 0, \qquad (4) \Rightarrow x_2 = 0$$

$$\overset{(5)}{\Rightarrow} 0 + 0 = 12$$

and from this contradiction we know that this case provides no candidates

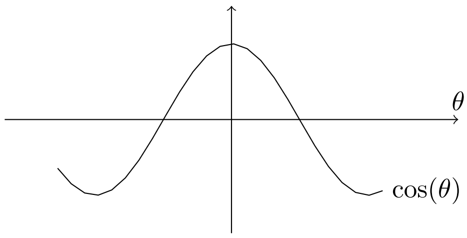
So by looking at the table we filled, we get that the winner is

$$(x_1^*, x_2^*, \lambda^*, \nu_1^*, \nu_2^*) = \left(0, 3, \frac{3}{2}, 11, 0\right).$$

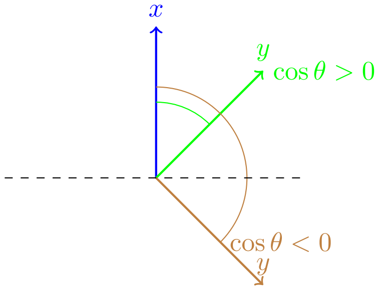
We see that this is a so called corner solution*:* $x_1^* = 0$*. This means that without Kuhn-Tucker (applying the simple Lagrange theorem, or studying the marginal rate of substitution), we would have never found this result.*

We may guess *under what conditions* those necessary conditions become sufficient. To answer, recall

$$x \cdot y = \|x\|\|y\| \cos\theta \Rightarrow \text{sign}(x \cdot y) = \text{sign}(\cos\theta)$$



We see that for θ lower than $90°$, the sign is always positive, and it is negative otherwise.



Theorem 56. *Suppose there are no equality constraints and that every function g_j is quasi-convex.*

Suppose also that the objective function f satisfies, for all $x, y \in \operatorname{dom} f$, $x \neq y$:

$$f(y) > f(x) \Rightarrow Df(x)(y-x) > 0 \tag{*}$$

Then, if $x^* \in C$ satisfies the Kuhn-Tucker conditions, it follows that x^* is a maximizer.

Remark 57. *f is quasi-convex $\iff$ C_α^- convex $\forall \alpha$*

Proof. Suppose, to the contrary, that there exists $y \in C$ such that $f(y) > f(x^*)$.

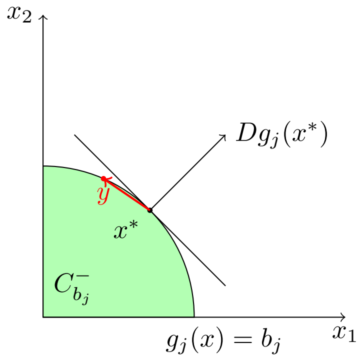
$$y \in C \Rightarrow g_j(y) \leq b_j \qquad \forall j = 1, \ldots, k$$

moreover,

$$Df(x^*)(y - x^*) > 0$$

by $(*)$. Now,

$$\left.\begin{matrix} g_j(y) \leq b_j \\ g_j \text{ quasi-convex} \end{matrix}\right\} \Rightarrow Dg_j(x^*)(y-x^*) \leq 0 \; \forall j = 1, \ldots, k$$



So

$$Df(x^*)(y-x^*) \overset{\text{Kuhn-Tucker}}{\Rightarrow} \sum_{j=1}^{h} \lambda_j \underbrace{Dg_j(x^*)(y-x^*)}_{\leq 0} \leq 0$$

which is a contradiction to $(*)$.

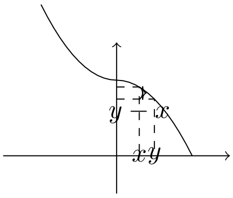
□

23/11/2010

To this point, we have expressed the Kuhn-Tucker theorem in the most general form.

Remark 58. 1. *Condition $(*)$ is satisfied when f is concave or when f is quasi-concave and $Df(x) \neq 0 \; \forall x \in \operatorname{dom} f$.*

2. *Condition $(*)$ cannot be dispensed with.*
Let's consider point (1), with f concave:



If instead f *is quasi-concave,*

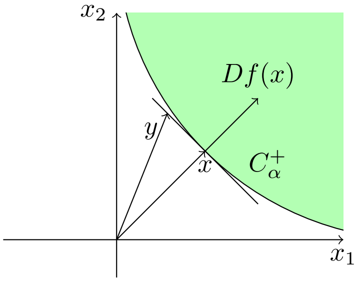
$$Df(x) \neq 0 \ \forall x \in \operatorname{dom} f$$

$$f(y) > f(x) \overset{\text{quasi-concavity}}{\Rightarrow} Df(x)(y-x) \geq 0$$

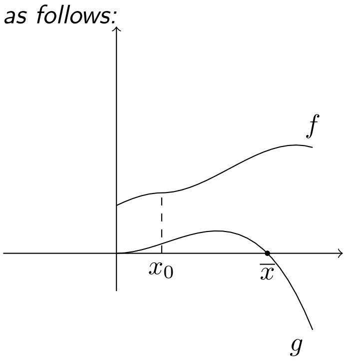
If $Df(x)(y-x) = 0$*, then*

$$\begin{aligned} &Df(x) \text{ and } (y-x) \text{ are orthogonal} \\ \Rightarrow &f(y) \leq f(x), \end{aligned}$$

as we can verify with a picture:



For what concerns (2)*, given* $f : \mathbb{R} \to \mathbb{R}$ *and* $g : \mathbb{R} \to \mathbb{R}$ *as follows:*



$\max f(x)$ *subject to* $g(x) \leq 0 \iff x \leq \overline{x} \Rightarrow x^* = \overline{x}$.

$$\text{Kuhn-Tucker: } \begin{cases} L(x,\lambda) = f(x) - \lambda g(x) \Rightarrow \frac{\partial L}{\partial x} = f'(x) - \lambda g'(x) = 0 \\ \lambda g(x) = 0 \end{cases}$$

x_0 *and* $\lambda = 0$ *satisfy Kuhn-Tucker... but* x_0 *is not a maximizer.*
This is possible because $(*)$ *is violated:*

$$f(\overline{x}) > f(x_0) \not\Rightarrow \underbrace{Df(x_0)}_{f'(x_0)=0}(\overline{x} - x_0) > 0$$

This is enough for the *sufficiency* of Kuhn-Tucker conditions. What about *unicity*? We can show the following:

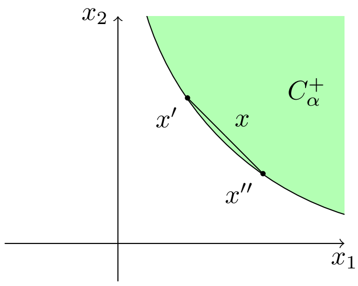
Theorem 59. *Suppose that in the general format (p) of the problem, the constraint set C is convex and the objective function f is strictly quasi-concave.*

Then, the global constrained maximizer is unique.

Proof. Let $x' \neq x''$ be both maximizers.

$$\begin{aligned} \Rightarrow x := tx' + (1-t)x'',\ t \in (0,1) \\ f(x') = f(x''),\ x' \neq x'' \\ \Rightarrow f(x) > f(x') \end{aligned}$$

because of strict quasi-concavity of f.

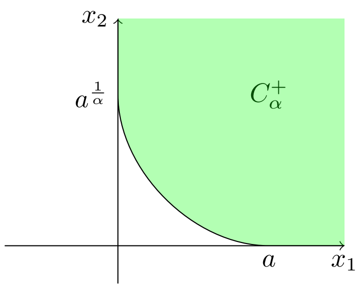


□

Example 60. $f(x_1, x_2) = x_1 + x_2^\alpha$ *for* $\alpha > 0$, $x_1, x_2 \geq 0$.

Let

$$\begin{aligned} &f(x_1, x_2) = a > 0 \\ \Rightarrow &x_1 + x_2^\alpha = a \\ \iff &x_2 = (a - x_1)^{\frac{1}{\alpha}} = \varphi(x_1) \\ \Rightarrow &\begin{cases} \varphi'(x_1) = & \frac{1}{\alpha}(a - x_1)^{\frac{1}{\alpha}-1}(-1) < 0 \\ \varphi''(x_1) = & \left(\frac{1}{\alpha} - 1\right)\frac{1}{\alpha}\underbrace{(a - x_1))}_{>0}^{\frac{1}{\alpha}-2}(-1)^2 \\ \end{cases} \\ &\frac{1}{\alpha} - 1 > 0 \iff \alpha < 1 \iff \varphi \text{ strictly convex} \iff f \text{ strictly quasi-concave.} \end{aligned}$$



$Df(x_1, x_2) = \left(1, \alpha x_2^{\alpha-1}\right) \neq (0,0)$, *so it is clear that this function never has a gradient equal to* 0*, so it satisfies our conditions for sufficiency* and *for uniqueness (for* $\alpha < 1$*): condition* $(*)$ *in the corresponding theorem is satisfied if* $\alpha \leq 1$ *and the solution is unique if* $\alpha < 1$.

$$\max x_1 + x_2^\alpha$$
$$\textit{s.t. } p_1 x_1 + p_2 x_2 \leq I$$
$$x_1, x_2 \geq 0$$

and let $\alpha = \frac{1}{2}$, $p_1 = 8$, $p_2 = 4$, $I = 12$.
We could now setup a table with 4 *cases. But that may be not worthwhile.*

- *Case 1*: $\nu_1 = \nu_2 = 0$.

$$\begin{aligned}
(1) \Rightarrow & \lambda = \frac{1}{8} \overset{(2)}{\Rightarrow} \frac{1}{2} x_2^{-\frac{1}{2}} - \frac{1}{8} \cdot 4 = 0 \\
\iff & x_2^{-\frac{1}{2}} = 1 \iff \frac{1}{\sqrt{x_2}} = 1 \iff x_2 = 1 \\
\overset{(5)}{\Rightarrow} & 8x_1 + 4 = 12 \iff x_1 = 1
\end{aligned}$$

This is *the only* solution: $(x_1^*, x_2^*, \lambda \cdot \nu_1^*, \nu_2^*) = \left(1, 1, \frac{1}{8}, 0, 0\right)$.

Because of the stated unicity conditions, we know all other cases should bring to contradictions.

We were lucky in find the solution at the first case... but on average, the new theoretic results we know save us 50% of time!

We want to elaborate a bit further on this case: can we always be sure we get an interior solution just because $\alpha = \frac{1}{2}$?

Let us assume

$$\alpha = \frac{1}{2}, \ p_1 = 15, \ p_2 = 3, \ I = 12;$$

is the optimum still interior?

- *Case 1*: $\nu_1 = \nu_2 = 0$. As before,

$$\begin{aligned}
(1) \Rightarrow & \lambda = \frac{1}{15} \overset{(2)}{\Rightarrow} \frac{1}{2} x_2^{-\frac{1}{2}} - \frac{1}{15} \cdot 3 = 0 \\
\iff & \frac{1}{\sqrt{x_2}} = \frac{2}{5} \Rightarrow x_2 = \frac{25}{4} \\
\overset{(5)}{\Rightarrow} & 15x_1 + 3 \cdot \frac{25}{4} = 12 \\
\iff & x_1 = \frac{12 - \frac{75}{4}}{15} = \frac{48 - 75}{15 \cdot 4} < 0
\end{aligned}$$

Since this is not the right case, the solution will not be interior.

- *Case 2*: $\nu_1 > 0, \nu_2 = 0$

$$\begin{aligned}(3) \Rightarrow & x_1 = 0 \overset{(5)}{\Rightarrow} 3x_2 = 12 \\ \iff & x_2 = 4 \overset{(2)}{\Rightarrow} \frac{1}{2} \cdot 4^{-\frac{1}{2}} = 3\lambda \\ \Rightarrow & \lambda = \frac{1}{12} > 0 \overset{(1)}{\Rightarrow} \nu_1 = 15\frac{1}{12} - 1 = \frac{1}{4} > 0;\end{aligned}$$

because of the unicity, *the* solution is

$$(x_1^*, x_2^*, \lambda^*, \nu_1^*, \nu_2^*) = (0, 4, \frac{1}{12}, \frac{1}{4}, 0).$$

Graphically,

$$\begin{aligned}f(1,1) &= 1 + 1^{\frac{1}{2}} = 2 \\ f(0,4) &= 0 + 4^{\frac{1}{2}} = 2\end{aligned}$$

which means that the change in prices didn't imply a change in utility for the consumer: he was able to keep the same by changing the consumption bundle.

$$f(x_1, x_2) = x_1 + x_2^{\frac{1}{2}} = 2 \Rightarrow x_2 = (2 - x_1)^2 = \begin{cases} 4 & \text{if } x_1 = 0 \\ 1 & \text{if } x_1 = 1 \\ 0 & \text{if } x_1 = 2 \end{cases}$$

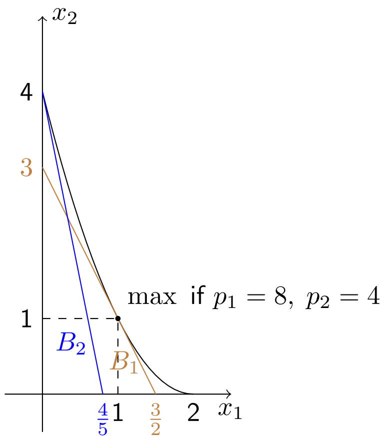
So if we call this function $\varphi(x_1)$, we get

$$\varphi'(x_1) = 2(2 - x_1)(-1) < 0 = \begin{cases} -4 & \text{if } x_1 = 0 \\ -2 & \text{if } x_1 = 1 \\ 0 & \text{if } x_1 = 2 \end{cases}$$

For what concerns the budget lines,

$$\begin{aligned}B_1 :& 8x_1 + 4x_2 = 12 \\ B_2 :& 15x_1 + 3x_2 = 12.\end{aligned}$$

The slopes of the two budget constraints are respectively -2 and -5. While in the first case the solution will be where φ is tangent to the budget constraint - that is, $\varphi'(x_1) = -2$, there is no point such that $\varphi'(x_1) = -5$, so we have necessarily a corner solution.



This is not an interior solution - if the consumer could buy a negative quantity of good 1, he would. This is not an easy predictable result - even in such trivial setups - so the Kuhn-Tucker theorem provides a crucial tool.

There is still another situation, namely

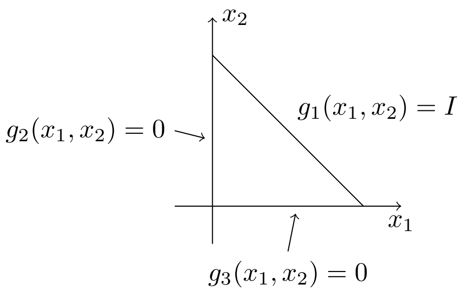
$$\alpha = 2,\ p_1 = 8,\ p_2 = 4,\ I = 14 \Rightarrow x_1^* = 0, x_2^* = 3;$$

illustrate this as an exercise.

It is necessary to find the level curve and understand which are the three candidates (the function is not quasi-concave), what they mean...

Constraint qualification: in practice, this is a topic of almost no relevance, but once in life we will still do this: analyze the constraint functions and see if they are linearly independent.

$$\begin{aligned} p_1x_1 + p_2x_2 \leq & I, \qquad x_1 \geq 0,\ x_2 \geq 0 \\ g_1(x_1, x_2) = & p_1x_2 + p_2x_2 \\ g_2(x_1, x_2) = & -x_1 \\ g_3(x_1, x_2) = & -x_2 \\ & \Downarrow \\ Dg_1(x_1, x_2) = & (p_1, p_2) \\ Dg_2(x_1, x_2) = & (-1, 0) \\ Dg_3(x_1, x_2) = & (0, -1). \end{aligned}$$



It is evident that we can never satisfy all 3 constraints: at most 2 of them can be binding simultaneously $\Rightarrow$ I always need to check at most 2 of the gradients together. Now: the first two (for $p_2 \neq 0$) obviously are, and similarly it is easy to verify the other two pairs, for $p_1, p_2 > 0$. And this is all we have to verify: CQ is satisfied.

This is all we had to say about optimization in the narrow sense, but this has a very important implication in economics. . .

2.3 The Envelope Theorem

We already saw that by changing the conditions, the solutions of optimization problems can change in an apparently unforecastable way: we want to now study how the maximizer and the maximum value change as one or several of the parameters (such as b_j or c_j) change.

Formally, set

$$\begin{matrix} f, & & \\ g_1, \dots, g_k, & : & U \times A \longrightarrow \mathbb{R} \\ h_1, \dots, h_n & & \uplus \\ & & (x, a) \end{matrix}$$

where $U \subset \mathbb{R}^n$, $A \subset \mathbb{R}^s$.

The problem, for a given $a \in A$, is:

$$\max_x f(x,a) \text{s.t.} \qquad \begin{aligned} g_1(x,1) &\leq 0 \\ &\vdots \\ g_k(x,a) &\leq 0 \\ h_1(x,a) &= 0 \\ &\vdots \\ h_m(x,a) &= 0. \end{aligned}$$

This form of the constraints is no restriction of generality as any equation

$$\widetilde{h}(x) = c$$

can be rewritten as

$$h(x,c) = \widetilde{h}(x) - c = 0.$$

The conditions on $g_1, \dots, g_k, h_1, \dots, h_m$ can be summarized as

$$x \in C(\alpha).$$

Let's further define

$$v(a) := \max\{f(x,a) | x \in C(a)\} = f(x^*(a), a)$$

(where x^* indicates the maximizer or maximizers), the *value function*.

Our goal is to understand how $v(a)$ changes as a changes.

Theorem 61 (Envelope theorem)**.** *Assume the value function $v(a)$ is differentiable at $\overline{a} \in A$, and that $\lambda_1, \ldots, \lambda_k$, $\mu_1, \ldots, \mu_m$ are the values of the Lagrange multipliers associated with the maximizer x^*. Then, we have the following result:*

$$\frac{\partial v}{\partial a_q}(\overline{a}) = \frac{\partial f}{\partial a_q}(x^*(\overline{a}), \overline{a}) - \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial a_q}(x^*(\overline{a}), \overline{a}) - \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial a_q}(x^*(\overline{a}), \overline{a}),$$

or, in more concise notation,

$$Dv(\overline{a}) = D_a f(x^*(\overline{a}), \overline{a}) - \sum_{j=1}^{k} \lambda_j D_a g_j(x^*(\overline{a}), \overline{a}) - \sum_{j=1}^{m} \mu_j D_a h_j(x^*(\overline{a})),$$

where obviously D_a is the derivative of D only with respect to a (not x).

Proof. Recall that we have

$$\begin{array}{ccccc} a & \longmapsto & (x^*(a), a) & \longmapsto & f(x^*(a), a) = v(a) \\ \Cup & & \Cup & & \Cup \\ \mathbb{R}^s & & \mathbb{R}^{n+s} & & \mathbb{R} \end{array}$$

$$\begin{aligned}
\frac{\partial v}{\partial a_q}(\overline{a}) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^*(\overline{a}), \overline{a}) \frac{\partial x_i^*}{\partial a_q}(\overline{a}) + \frac{\partial f}{\partial a_q}(x^*(\overline{a}), \overline{a}) \\
&\overset{\text{Kuhn-Tucker}}{=} \frac{\partial f}{\partial a_q}(x^*(\overline{a}), \overline{a}) + \sum_{i=1}^{n} \left\{ \left[\sum_{j=1}^{k} \lambda_j (x^*(\overline{a}), \overline{a}) + \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i}(x^*(\overline{a}), \overline{a}) \right] \frac{\partial x_i^*}{\partial a_q}(\overline{a}) \right\} \\
&= \frac{\partial f}{\partial a_q}(x^*(\overline{a}), \overline{a}) + \sum_{j=1}^{k} \lambda_j \sum_{i=1}^{n} \frac{\partial g_j}{\partial x_i}(x^*(\overline{a}), \overline{a}) \frac{\partial x_i^*}{\partial a_q}(\overline{a}) + \sum_{j=1}^{m} \mu_j \sum_{i=1}^{n} \frac{\partial h_j}{\partial x_i}(x^*(\overline{a}), \overline{a}) \frac{\partial x_i^*}{\partial a_q}(\overline{a}).
\end{aligned}$$

We know, from the formulation of the problem, that

$$h_j(x^*(\overline{a}), \overline{a}) = 0 \qquad \forall \overline{a} \in A,\ \forall j = 1, \ldots, n,$$

so $z_j(\overline{a}) := h_j(x^*(\overline{a}), \overline{a})$ is constant in $\overline{a}$. So:

$$\begin{aligned}
0 = \frac{\partial z_j}{\partial a_q}(\overline{a}) &= \sum_{i=1}^{n} \left[\frac{\partial h_j}{\partial x_i}(x^*(\overline{a}), \overline{a}) \frac{\partial x_i^*}{\partial a_q}(\overline{a}) \right] + \frac{\partial h_j}{\partial a_q}(x^*(\overline{a}), \overline{a}) \\
\iff \sum_{i=1}^{n} \left[\frac{\partial h_j}{\partial x_i}(x^*(\overline{a}), \overline{a}) \frac{\partial x_i^*}{\partial a_q}(\overline{a}) \right] &= -\frac{\partial h_j}{\partial a_q}(x^*(\overline{a}), \overline{a}) \qquad \forall j = 1, \ldots, m.
\end{aligned}$$

What we did for equality constraints, we can obviously do it with *in*equality constraints:

$$\sum_{i=1}^{n} \frac{\partial g_j}{\partial x_i}(x^*(\overline{a}), \overline{a}) \frac{\partial x_i^*}{\partial a_q}(\overline{a}) = -\frac{\partial g}{\partial a_q}(x^*(\overline{a}), \overline{a})$$

$\forall \overline{a} \in A, \forall q = 1, \ldots, s,\ \forall j = 1, \ldots, k$ s.t. $g_j(x^*(\overline{a}), \overline{a}) = 0$;

hence, we can now put things together as follows:

$$\frac{\partial v}{\partial a_q}(\overline{a}) = \frac{\partial f}{\partial a_q}(x^*(\overline{a}), \overline{a}) - \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial a_q}(x^*(\overline{a}), \overline{a}) - \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial a_q}(x^*(\overline{a}), \overline{a})$$

But what happens if we do *not* have the equality? The corresponding λ_j is 0, so the second term is not relevant.

This is precisely what we wanted to show. □

To get a better understanding of the significance of this, we'll look at examples. But first, let's make a

Remark 62.

$$v(a) = f(x^*(a), a)$$

can be rewritten in a more complicated way as

$$f(x^*(a), a) - \sum_{j=1}^{k} \lambda_j^*(a) g_j(x^*(a), a) - \sum_{j=1}^{m} \mu_j^*(a) h_j(x^*(a), a)$$

since the h_j are null, and if the g_j are smaller than 0, the corresponding λ are 0. So we can reformulate as:

$$L(x^*(a), \lambda^*(a), \mu^*(a), a),$$

the Lagrange function.

So the Envelope Theorem says

$$\frac{\partial v}{\partial a_q} = \frac{\partial L}{\partial a_q}(x^*(a), \lambda^*(a), \mu^*(a), a)$$

or in other terms:

$$Dv(a) = D_a L(x^*(a), \lambda^*(a), \mu^*(a), a).$$

Example 63. *Let $n = s = 1$, $k = m = 0$ (no constraints).*

This means that

$$\begin{aligned} v(a) &= \max_x f(x, a) \\ &= f(x^*(a), a) \\ \overset{\text{envelope theorem}}{\Rightarrow} v'(a) &= \frac{\partial f}{\partial a}(x^*(a), a) \end{aligned}$$

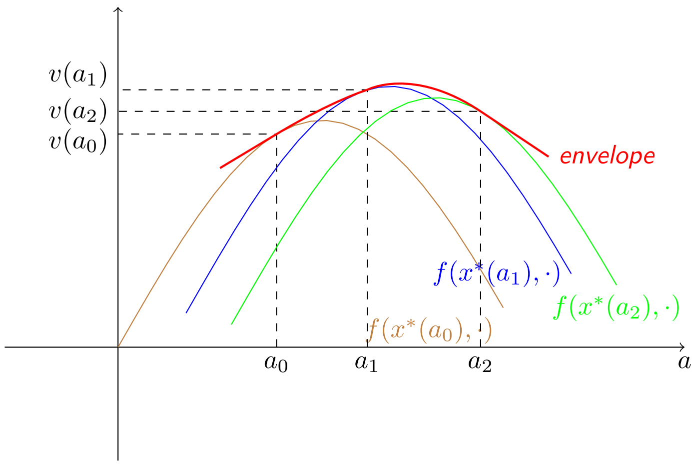
while simply applying the chain rule would have given us:

$$v'(a) = \frac{\partial f}{\partial x}(x^*(a), a) \frac{\partial x^*}{\partial a}(a) + \frac{\partial f}{\partial a}(x^*(a), a);$$

the two expressions must be equal, and this is true indeed, since we know that in an unconstrained maximum the partial derivatives ($\frac{\partial f}{\partial x}(x^*(a), a)$) must be $= 0$.

Of course, given any a,

$$v(a) = f(x^*(a), a) \geq f(x, a)\ \forall x$$



24/11/10

Typical economic interpretations of the envelope theorem are the distinctions between short run and long run curves. For instance, in the above, a_0 would be the stock of capital, which doesn't change in the immediate but will change in time if x (the labour force) changes; then, day by day, the optimum will be reached on the current value on a, and the function of production (v) in time is the envelope.

Example 64. *Let's take $n = s = 1$ as before, but now $k = 0$, $m = 1$.*

$$\max f(x) \qquad \textit{s.t. } h(x,c) = \widetilde{h}(x) - c = 0$$
$$\Rightarrow v'(c) = -\mu \underbrace{\frac{\partial h}{\partial c}(x^*(c), c)}_{=-1} = \mu$$

Example 65. *A similar thing happens if we add an inequality constraint:*

$$n = s = 1, \qquad k = 1,\ m = 0$$

$$\max f(x) \qquad \textit{s.t. } g(x,b) = \widetilde{g}(x) - b \leq 0$$
$$\Rightarrow v'(b) = -\lambda \underbrace{\frac{\partial g}{\partial b}(x^*(b), b)}_{=-1} = \lambda \begin{cases} = 0 & \textit{if } g(x^*(b), b) < 0 \\ \geq 0 & \textit{if } g(x^*(b), b) = 0 \end{cases}.$$

Example 66.

$$\max f(x_1, x_2) = x_1 + x_2^\alpha$$
$$\textit{s.t. } p_1 x_1 + p_2 x_2 \leq I$$
$$x_1 \geq 0,\ x_2 \geq 0$$

where now

$$a = (\alpha, p_1, p_2, I)$$
$$\Rightarrow v(\alpha, p_1, p_2, I)$$
$$\Rightarrow \frac{\partial v}{\partial a_q} \overset{E.T.}{=} \frac{\partial f}{\partial a_q} - \lambda \frac{\partial}{\partial a_q} \underbrace{(p_1 x_1 + p_2 x_2 - I)}_{g}$$
$$\Rightarrow \frac{\partial v}{\partial \alpha}(x_1 + x_2^\alpha) = x_2^\alpha \ln x_2$$
$$\frac{\partial v}{\partial p_1} = -\lambda x_1 \leq 0$$

(if prices rise, the consumer gets hurt, and he gets hurt even more if he used to consume much of that good). Analogously,

$$\frac{\partial v}{\partial p_2} = -\lambda x_2 \leq 0$$
$$\frac{\partial v}{\partial I} = -\lambda(-1) = \lambda > 0;$$

this is the marginal utility of money, *which is clearly positive.*[11]

[11] It's clear that here we are giving a *cardinal*, not just ordinal, meaning to the utility.

Remark 67. *We always put $x_i \geq 0$ just because it's what is most interesting for us, thinking to economic applications.*

Example 68. *Let's consider a profit function $\Pi(k,l) = 16k^{\frac{1}{2}}l^{\frac{1}{2}} - 2k - 4l$.*

We want to

$$\max \Pi(k,l) \qquad \textit{s.t. } k \leq \overline{k},$$

where the latter is a capacity constraint. *We set up the Lagrange function:*

$$\begin{aligned} L &= 16k^{\frac{1}{2}}l^{\frac{1}{2}} - 2k - 4l - \lambda(k - \overline{k}) \\ \frac{\partial L}{\partial K} &= 8k^{-\frac{1}{2}}l^{\frac{1}{2}} - 2 - \lambda \overset{(1)}{=} 0 \\ \frac{\partial L}{\partial l} &= 8k^{\frac{1}{2}}l^{-\frac{1}{2}} - 4 \overset{(2)}{=} 0 \\ \underbrace{\lambda}_{\geq 0}\underbrace{(k - \overline{k})}_{\leq 0} &\overset{(3)}{=} 0. \end{aligned}$$

We can write condition (2) as follows:

$$\begin{aligned} \left(\frac{k}{l}\right)^{\frac{1}{2}} &= \frac{4}{8} = \frac{1}{2} \\ \iff \left(\frac{l}{k}\right)^{\frac{1}{2}} &\overset{(4)}{=} 2 \\ (1) \Rightarrow \left(\frac{l}{k}\right)^{\frac{1}{2}} &= \frac{2+\lambda}{8} \\ \Rightarrow \lambda = 14 > 0 &\overset{(3)}{\Rightarrow} k^* = \overline{k}. \end{aligned}$$

Now, this allows us also to determine l^*, because

$$\begin{aligned} (4) \Rightarrow & l^* = 4\overline{k} \\ \Rightarrow & \Pi^* = \Pi(k^*, l^*) = 16\overline{k}^{\frac{1}{2}}(4\overline{k})^{\frac{1}{2}} - 2\overline{k} - 4 \cdot 4\overline{k} \\ & = 32\overline{k} - 18\overline{k} = 14\overline{k} = v(\overline{k}) \end{aligned}$$

which is the value function, *which shows how $\overline{k}$ influences the profit.*

We can hence calculate the derivative:

$$v'(\overline{k}) = 14$$

and observe that it is equal to λ. That is of course not by chance.

Economists think of marginal increase as a one unit *increase. So Π^* increases, when we increase k from $\overline{k}$ to $\overline{k}+1$, from $14\overline{k}$ to $14\overline{k}+1$.*

This means that if the firm can buy a new machine to increase k by one, it will buy it only if the machine costs at most 14, which is the utility of one more unit of capital.

14 is called the shadow price. *That's why we will often call λ "shadow price" even in problems in which multipliers have nothing to do with prices.*

After having sufficiently illustrated the Envelope Theorem, we can switch to the new part:

3 Correspondences and related theorems

3.1 Continuity concepts of correspondences

Example 69. *Let's consider our - by now very familiar - example of a consumer, but with a specificity: the exponent is* 1*:*

$$\max x_1 + x_2 \qquad \text{s.t. } p_1x_1 + p_2x_2 \leq I, \qquad x_1, x_2 \geq 0.$$

The solution x_1^* *is* $x_1^*(p_1, p_2, I)$ *(for* $i = 1, 2$*). In particular, if we concentrate on*

$$x_1^*(p_1, \overline{p}_2, \overline{I})$$

where $\overline{p}_2$ *and* $\overline{I}$ *are, for the moment, considered as fixed* $\Rightarrow x_1^* = x_1^*(p_1)$*.*

$$x_1^*(p_1) \begin{cases} = \frac{I}{p_1} & \text{if } p_1 < \overline{p}_2 \\ \in \left[0, \frac{\overline{I}}{p_1}\right] & \text{if } p_1 = \overline{p}_2 \\ = 0 & \text{if } p_1 > \overline{p}_2 \end{cases}$$

but this is not so nice, because that is no more a function, so we can not treat it with the mathematical tools seen so far. We will have to generalize (some of) them.

Definition 70. *Given* $X \subset \mathbb{R}^k$*, a* correspondence $f : X \to Y \subset \mathbb{R}^m$ *is a rule that assigns to every* $x \in X$ *a set* $f(x) \subset Y$*.*

Remark 71. *If* $f(x)$ *is a singleton* $\forall x \in X$*, that is* $f(x) = \{y\}$*, then* f *can be identified with a function in the usual sense, that is*

$$f(x) = \{y\} \iff f(x) = y$$

Example 72. *Let's take again the previous case: we have seen that* $x_1^*(p_1)$ *was not determined for* $p_1 = \overline{p}_2$*; we can rewrite that as follows:*

$$x_1^*(p_1) = \begin{cases} \left\{\frac{\overline{I}}{p_1}\right\} & \text{if } p_1 < \overline{p}_2 \\ \left[0, \frac{\overline{I}}{p_1}\right] & \text{if } p_1 = \overline{p}_2 \\ \{0\} & \text{if } p_1 > \overline{p}_2, \end{cases}$$

keeping in mind that this is a correspondence.

Since continuity is a crucial property when working with functions, we would like to extend the concept to correspondences:

Definition 73. *Given* $X \subset \mathbb{R}^n$ *and the closed set* $Y \subset \mathbb{R}^m$*, the correspondence*

$$f : X \to Y$$

has a closed graph *if, for any two sequences* $\{x^n\}$ *in* X *and* $\{y^n\}$ *in* Y *with* $x^n \to x \in X$*,* $y^n \to y$ *and* $y^n \in f(x^n)$ *for every* n*, we have*

$$y \in f(x).$$

f is upper hemicontinuous *("uhc") if it has a closed graph and the images of compact*[12] *sets are bounded, that is for any compact set $U \subset X$, the set*

$$f(U) := \{y \in Y | y \in f(x) \text{ for some } x \in U\}$$

is bounded.

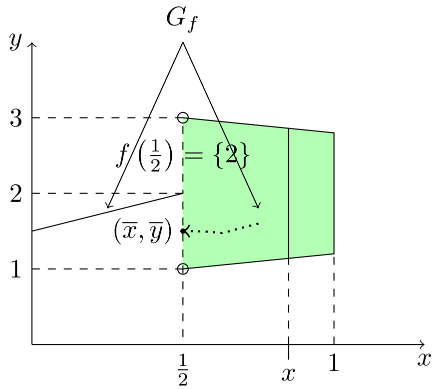
Remark 74. *If Y is compact, then:*

$$f \text{ has a closed graph } \iff f \text{ uhc.}$$

We can now try to illustrate what uhc means:

Example 75.

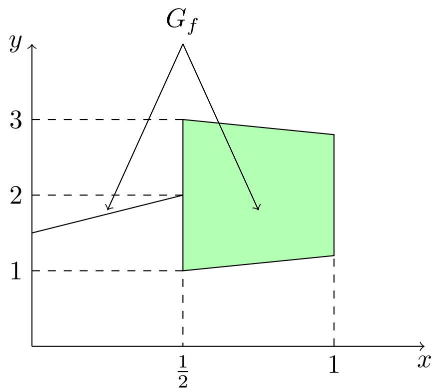
$$X = [0, 1], \qquad Y = [0, 3].$$



(a) *Consider $(\overline{x}, \overline{y})$ s. t. $\overline{x} = \frac{1}{2}$, $\overline{y} \in (1, 2) \Rightarrow (\overline{x}, \overline{y}) \notin G_f$, take $\{x^n\} \to \overline{x}$, $\{y^n\} \to \overline{y}$ s.t. $y^n \in f(x^n) \forall n \Rightarrow \overline{y} \notin f(\overline{x})$.*

So f is not uhc.

(b) *f would be uhc if $f(\frac{1}{2}) = [1, 3]$:*



Remark 76. *If $f(x) = \{y\} = y$ is a function, then*

$$f \text{ uhc} \Rightarrow f \text{ continuous,}$$

[12]A *compact* set is a closed and bounded set.

because

$$\begin{array}{ccc} x^n & \longrightarrow & x \\ \downarrow & & \downarrow \\ f(x^n) & \longrightarrow & f(x) \\ \| & & \| \\ y^n & \longrightarrow & y. \end{array}$$

This helps us see correspondences as a generalization of functions.

However, there is another way to generalize the concept of continuity:

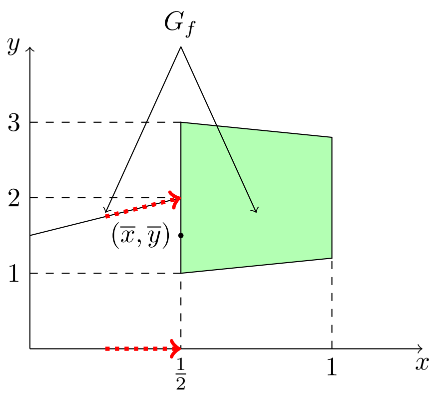
Definition 77. *Given $X \subset \mathbb{R}^k$ and $Y \subset \mathbb{R}^m$ compact, the correspondence $f : X \to Y$ is* lower hemicontinuous *("lhc") if, for every sequence $\{x^n\}$ with $x^n \to x \in X$ and every $y \in f(x)$ we can find a sequence $\{y^n\}$ in Y with $y^n \to y$ and $y^n \in f(x^n)$ for all n.*

Let's try to illustrate this:

$$\begin{array}{ccc} x^n & \longrightarrow & x \\ \downarrow & & \downarrow \\ f(x^n) & & f(x) \\ \Cup & & \Cup \\ \exists y^n & \longrightarrow & y; \end{array}$$

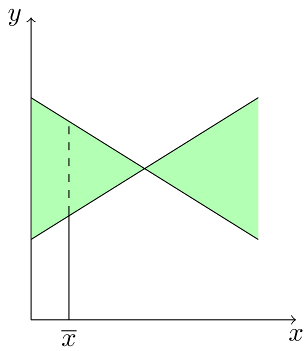
knowing that f (a function) is lhc means:

$$\begin{array}{ccc} x^n & \longrightarrow & x \\ \downarrow & & \downarrow \\ f(x^n) & \longrightarrow & f(x) \\ \| & & \| \\ \exists y^n & \longrightarrow & y. \end{array}$$



Consider $x^n \to \left(\frac{1}{2}\right)^-$.
$\overline{y} \in f(\overline{x}),\ 1 < \overline{y} < 2$
$y^n \in f(x^n) \Rightarrow y^n \to 2 \neq \overline{y}$.
So $y^n \not\to \overline{y} \Rightarrow f$ is not lhc.
We may wonder if the function of the case (a), in which $f\left(\frac{1}{2}\right) = \{2\}$, is lhc... it is.[13]

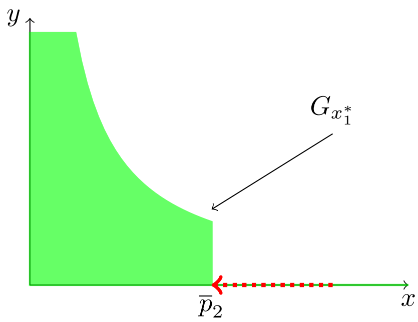
Definition 78. *A correspondence f is* continuous *if f is uhc* and *lhc.*



Example 79.
A continuous correspondence.

Example 80.

$$x_1^*(p_1) = \begin{cases} \left\{\frac{\overline{I}}{p_1}\right\} & \text{if } p_1 < \overline{p}_2 \\ \left[0, \frac{\overline{I}}{p_1}\right] & \text{if } p_1 = \overline{p}_2 \\ \{0\} & \text{if } p_1 > \overline{p}_2, \end{cases}$$



Our function is not lhc - see the red sequence. However, it is uhc - intuitively, it is not possible to approach a point outside the graph from inside the graph.

We will see next if uhc is considered "nice enough" for our economic purposes (if with it it is feasible to find equilibria, applicate fixed point theorems...)

25/11/2010

3.2 Theorem of the maximum

Let's look at the problem

[13]Notice that the circled points $\left(\frac{1}{2}, 3\right)$ and $\left(\frac{1}{2}, 1\right)$ are *excluded* from G_f.

$$\max_x f(x,a) \qquad \text{s.t. } g(x,a) \le 0$$
$$h(x,a) = 0$$

where the functions can obviously be in an arbitrary number of variables.

And let $C(a)$ be the set of points satisfying those conditions.

Let moreover $x^*(a)$ be the solution of the problem, and

$$v(a) = f(x^*(a), a).$$

How fundamental are the assumptions about continuity and differentiability of functions? That is the content of the

Theorem 81 (Theorem of the Maximum)**.** *Suppose that the constraint correspondence*

$$\begin{array}{ccc} A & \xrightarrow{C} & \mathbb{R}^n \\ \Cup & & \Cup \\ a & \longmapsto & C(a) \end{array}$$

is continuous and f is a continuous function. Then, the maximizer correspondence

$$\begin{array}{ccc} A & \xrightarrow{C} & \mathbb{R}^n \\ \Cup & & \cup \\ a & \longmapsto & x^*(a) \end{array}$$

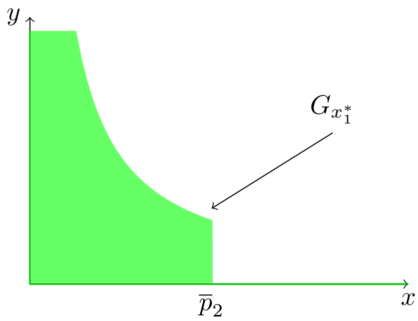
is (not continuous, as we would hope, but...) uhc and the value function

$$\begin{array}{ccc} A & \xrightarrow{v} & \mathbb{R} \\ \Cup & & \Cup \\ a & \longmapsto & v(a) \end{array}$$

is indeed continuous.

Example 82. *It is sufficient to remind yesterday's example:*

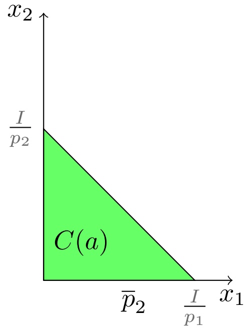
$$\max x_1 + x_2$$
$$\textit{s.t. } p_1x_1 + p_2x_2 \le I \qquad x_1, x_2 \ge 0$$



and remind the function is uhc but not lhc. . . though it is easy to see that

$$a = (p_1, p_2, I) \longmapsto C(a)$$
$$\qquad\qquad\qquad\qquad \|$$
$$\{(x_1, x_2) | p_1 x_2 + p_2 x_2 \leq I,\ x_1 \geq 0, x_2 \geq 0\}$$

is indeed continuous.



Proof. Consider two sequences $a^n \to a \in A$ and $x^n \to x \in \mathbb{R}^n$, with $x^n \in x^*(a^n)$ for each n.

We want to show that the maximizer correspondence is uhc. This means we must show (from the definition of "uhc") that $x \in x^*(a)$.

We know that $x^n \in x^*(a^n)$, and that obviously implies $x^n \in C(a^n)$. But we know that C is continuous (and, in particular, is uhc) and hence $x \in C(a)$. Let $y \in C(a)$; again, since C is (continuous, and hence) lhc, we get

$$\exists y^n \in C(a^n)\ \forall n$$
$$\text{s.t. } y^n \to y.$$

However, since x^n is a maximizer, we deduct that

$$f(y^n, a^n) \leq f(x^n, a^n)\ \forall n$$
$$\Rightarrow f(y^n, a^n) \xrightarrow{f \text{ cont.}} f(y, a)$$
$$\wedge| \qquad\qquad \wedge|$$
$$f(x^n, a^n) \xrightarrow{f \text{ cont.}} f(x, a)$$
$$\Rightarrow f(y, a) \leq f(x, a)$$
$$\Rightarrow x \in x^*(a):$$

this proves the property of the uhc of the correspondence... and we already know this is the strongest continuity result that we can get: we must live witht the fact that the demand correspondence is not necessarily lhc.

But we will now see this is not too bad. . .

3.3 Fixed-point Theorems

Example 83. *Consider a given market price, for some good,*

$$p_{t+1} = \frac{D(p_t)}{S(p_t)} \cdot p_t =: f(p_t)$$

where D and S are respectively demand *and* supply *of the good. Obviously,*

$$p_{t+1} = p_t \iff D(p_t) = S(p_t) \iff p_t = p_{t+1} = f(p_t),$$

which is a market equilibrium*. In that case, $p_t = p^*$ is a* fixed point *of the function f.*

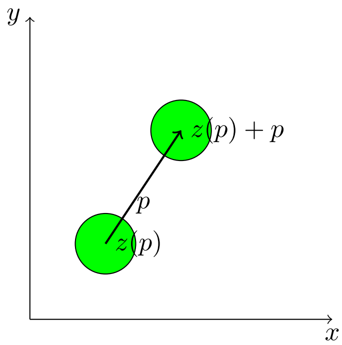
In general, equilibria of economic models can always be seen as fixed points of some functions.

Example 84. *Consider an economy with n markets (commodities). Consider the aggregate excess demand correspondence*

$$\begin{array}{ccc} X & \xrightarrow{z} & \mathbb{R}^n \\ \Cup & & \cup \\ p & \longmapsto & z(p) \quad = D(p) - S(p) \end{array} .$$

p^* is a general equilibrium if $\mathbb{R}^n \ni 0 \in z(p^*)$.

Now, let $f(p) := z(p) + p$ (for each element in the set $z(p)$, we add p.



Finally, p^* is a general equilibrium if and only if

$$\begin{aligned} & 0 \in z(p^*) \\ & \iff p^* \in z(p^*) + p^* \\ & \iff p^* \in f(p^*) \\ & \iff p^* \text{ is a fixed point of } f. \end{aligned}$$

In general,

Definition 85. *Let $f : X \to X \subset \mathbb{R}^n$ be a correspondence: $x \in X$ is a* fixed point *of f if $x \in f(x)$.*

It is evident that this is the analogous of the definition of fixed point of a function - in that case, $x = f(x)$.

□

Remark 86. *Gerard Debreu obtained a Nobel Prize because he was able to find a general equilibrium... in other words, to find a fixed point.*

The term that becomes 0 in the fixed point is the *excess demand and supply*.

This relation between the economic concept of equilibrium and the mathematical concept of fixed points makes it natural for us to investigate *when* it is the case that a given correspondence admits a fixed point: fixed point theorems.

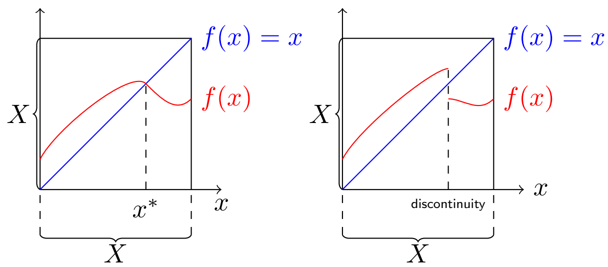
Theorem 87 (Kakutani's (Brouwer's) Fixed Point Theorem)**.** *Suppose that $X \subset \mathbb{R}^n$ is a non-empty, compact, convex set and*

$$f : X \to X$$

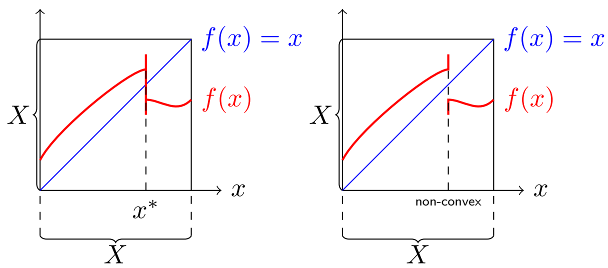
is an uhc correspondence (function), with $f(x) \neq \emptyset$ and $f(x)$ convex for all $x \in X$.[14] *Then, f has a fixed point.*

The proof can be found in hundreds of volumes... we just wan to get a bit of intuition.

- *Case 1*: f is a function



- *Case 2*: f is a correspondence



4 Dynamics

4.1 Difference equations

Example 88. *Consider money in a bank account at given moments in time:*

$$y_{n+1} = y_n + \rho y_n = (1+\rho) y_n$$

[14] Both properties are trivially true if f is a function

where

$$y_n = \text{currency units in the bank account in period } n$$
$$\rho = \text{interest rate per period.}$$

Then, a natural question is: given y_0, ρ, what is y_n?
Of course, that it easy, because

$$\begin{aligned} y_1 =&(1+\rho)y_0 \\ y_2 =&(1+\rho)^2 y_0 \\ &\dots \\ y_n =&(1+\rho)^n y_0; \end{aligned}$$

the above is an example of a linear difference *equation, that we write as*

$$y_{n+1} = a y_n, \ a \in \mathbb{R}$$

and the solution is

$$y_n = a^n y_0.$$

Things become a bit more interesting if we consider a system of *two* linear difference equations, that is:

$$\begin{aligned} x_{n+1} =& a x_n + b y_n \\ y_{n+1} =& c x_n + d y_n. \end{aligned}$$

This is the problem we want to solve. Let's first introduce a notation:

$$z_{n+1} := \begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \underbrace{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}_{A} \begin{pmatrix} x_n \\ y_n \end{pmatrix} = A z_n.$$

In the case $b = c = 0$, there is no interdependence between the two variables, we fallback to the previous case and we say the equations are *uncoupled*.

If instead $b \neq 0$ or $c \neq 0$ (or both), then we must proceed in a different way. We will use the concept of eigenvalues and eigenvectors.

Let r_i and v_i, $i = 1, 2, \dots$ be the eigenvalues and eigenvectors corresponding to the matrix A, that is,

$$A v_i = r_i v_i \qquad \forall i = 1, 2$$

We can rewrite this as follows:

$$\begin{aligned} &\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} v_{i1} \\ v_{i2} \end{pmatrix} = \begin{pmatrix} r_i v_{i1} \\ r_2 vi2 \end{pmatrix} \qquad \forall i = 1, 2 \\ \iff &\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix} = \begin{pmatrix} r_1 v_{11} & r_2 v_{21} \\ r_1 v_{12} & r_2 v_{22} \end{pmatrix} \\ &= \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix} \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} \\ \iff &AP = PD \end{aligned}$$

where

$$P = \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix}, \qquad \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix}.$$

If P^{-1} exists (that is, $\det P \neq 0$), then

$$P^{-1}AP = D;$$

then set

$$Z_n := P^{-1}z_n \qquad \forall n$$

and hence

$$\begin{aligned} Z_{n+1} =& P^{-1}z_{n+1} = P^{-1}Az_n \\ =& P^{-1}APZ_n = DZ_n. \end{aligned}$$

Then, if we set

$$Z_n = \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \qquad \forall n,$$

we obtain

$$\begin{pmatrix} X_{n+1} \\ Y_{n+1} \end{pmatrix} = D \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$$
$$\iff \begin{pmatrix} X_{n+1} \\ Y_{n+1} \end{pmatrix} = \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \end{pmatrix};$$

this is an *uncoupled* system in X_n, Y_n, with solutions

$$\begin{aligned} X_n =& r_1^n X_0 \\ Y_n =& r_2^n Y_0. \end{aligned}$$

This is great... but doesn't give us the values we wanted to know. How can we get back x_n and y_n? Recall

$$\begin{aligned} \begin{pmatrix} x_n \\ y_n \end{pmatrix} =& z_n = PZ_n \\ =& \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix} \begin{pmatrix} r_1^n X_0 \\ r_2^n Y_0 \end{pmatrix} \\ =& r_1^n X_0 \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} + r_2^n Y_0 \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}. \end{aligned}$$

What can we say about X_0 and Y_0? By varying them, we get all possible solutions to the system

$$z_{n+1} = Az_n. \qquad (*)$$

Therefore, setting $c_1 = X_0$ and $c_2 = Y_0$, $v_1 = \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix}$ and $v_2 = \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}$, we get the general solution to (*) as

$$z_n = c_1 r_1^n v_1 + c_2 r_2^n v_2 \qquad \forall n.$$

In fact:

$$\begin{aligned} z_{n+1} =& c_1 r_1^{n+1} v_1 + c_2 r_2^{n+1} v_2 \\ =& c_1 r_1^n \cdot r_1 v_1 + c_2 r_2^n \cdot r_2 v_2 \\ =& c_1 r_1^n A v_1 + c_2 r_2^n A v_2 \\ =& A(c_1 r_1^n v_1 + c_2 r_2^n v_2) \\ =& A z_n; \end{aligned}$$

so whatever are c_1 and c_2, this expression solves indeed our problem.

That said, I will want to have a *specific* solution, with a given value for x_0 and a given value for y_0. How can I get that? Before we find out, let's state what we obtained, in a more general case:

Theorem 89. *Let A be a $k \times k$ matrix with k distinct real eigenvalues $r_1, \ldots, r_k$ and corresponding eigenvectors $v_1, \ldots, v_k$. Then the general solution of the system of difference equations*

$$z_{n+1} = A z_n$$

is

$$z_n = c_1 r_1^n v_1 + \cdots + c_k r_k^n v_k.$$

Now let's come back to the problem of putting desired initial values: if x_0 and y_0 are given, then

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} = Z_0 = P^{-1} z_0 = \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Next time, we shall consider an example.

29/11/2010

We have already done the major theoretical part of the topic of difference equations, and seen more in detail the case of a system of two equations.

We will now see an

Example 90. *Let*

$$\left.\begin{aligned} x_{n+1} = & \quad -x_n + 3y_n \\ y_{n+1} = & \qquad 2x_n \end{aligned}\right\} \Rightarrow A = \begin{pmatrix} -1 & 3 \\ 2 & 0 \end{pmatrix}$$
$$\Rightarrow Av_i = r_i v_i,\ i = 1,2$$
$$\Rightarrow (A - r_i I)v_i = 0,\ i = 1,2$$

Now,

$$\begin{aligned} \det(A - rI) = & \det \begin{pmatrix} -1-r & 3 \\ 2 & -r \end{pmatrix} \\ = & (1+r)r - 6 = r^2 + r - 6 \\ = & (r+3)(r-2) = 0 \end{aligned}$$

and from this it is obvious which are the eigenvalues: $r_1 = -3$, $r_2 = 2$.

From them, we can find the eigenvectors from

$$\begin{aligned} & (A - r_i I)v_i = 0,\ i = 1,2 \\ \Rightarrow & (A - r_1 I)v_1 = \begin{pmatrix} -1+3 & 3 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \Rightarrow & 2v_{11} + 3v_{12} = 0 \\ \Rightarrow & \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} = \lambda \begin{pmatrix} 3 \\ -2 \end{pmatrix} \end{aligned}$$

with $\lambda \neq 0$. *Similarly, for the other eigenvalue,*

$$\begin{aligned} & (A - r_2 I)v_2 = \begin{pmatrix} -1-2 & 3 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix} \\ \Rightarrow & \begin{cases} -3v_{21} + 3v_{22} = & 0 \\ 2v_{21} - 2v_{22} = & 0 \end{cases} \\ \Rightarrow & \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix},\ \lambda \neq 0 \\ \Rightarrow & z_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} = c_1(-3)^n \begin{pmatrix} 3 \\ -2 \end{pmatrix} + c_2 \cdot 2^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \end{aligned}$$

and this is the general solution of the problem. But what is the expression for given x_0, y_0*?*

As already seen,

$$
\begin{aligned}
\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} &= \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\
&= \begin{pmatrix} 3 & 1 \\ -2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\
&= \frac{1}{5} \begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\
&= \frac{1}{5} \begin{pmatrix} x_0 - y_0 \\ 2x_0 + 3y_0 \end{pmatrix} \\
\Rightarrow \begin{pmatrix} x_n \\ y_n \end{pmatrix} &= \frac{1}{5}(x_0 - y_0)(-3)^n \begin{pmatrix} 3 \\ -2 \end{pmatrix} + \frac{1}{5}(2x_0 + 3y_0)2^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}.
\end{aligned}
$$

It is quite clear that for $n \to \infty$, the solution does not converge to a limit. This would happen instead if the eigenvalues were smaller than 1 in absolute value.

In general, if there is at least one eigenvalue bigger than 1, there is at least some vector - its eigenvactor - for which the solution tends to "explode".

4.2 Differential Equations

Example 91 (Savings account)**.** *Assume we know*

$$
\begin{aligned}
y(t+1) &= (1+\rho)y(t) \\
\Rightarrow \frac{y(t-1) - y(t)}{y(t)} &= \rho
\end{aligned}
$$

where ρ is the annual interest rate.

If the interest is paid every Δt fraction of the year, then we get

$$
\frac{y(t+\Delta t)}{y(t)} = \rho \Delta t;
$$

Example 92. *If interest is paid every month, then $\Delta t = \frac{1}{12}$.*

Now interest can also be payed continuously (continuous compounding), we have to get Δ_t to 0.

Consider the differential equation:

$$
\lim_{\Delta t \to 0} \frac{y(t+\Delta t) - y(t)}{\Delta t} = \rho y(t)
$$

The solution is

$$
\begin{aligned}
y(t) =& ke^{\rho t}, \ k \in \mathbb{R} \\
&\Rightarrow y'(t) = ke^{\rho t}\rho = \rho y(t)
\end{aligned}
$$

(notice that only the exponential function has a derivative equal to the function itself); if $y(0) = y_0$ is given, then

$$
\begin{aligned}
y(0) &= ke^0 = k \\
&\Rightarrow x = y_0
\end{aligned}
$$

Definition 93. *A* first-order differential equation *is given by*

$$y'(t) = F(y(t), t)$$

If $F(y(t), t)$ features t separately - that is, if the equation can be written as $y'(t) = F(y, d)$ - then it is called autonomous *or* time-indepedent, *otherwise it is* non-autonomous *and* time dependent

Example 94.

$$y'(t) = [y(t)]^2 + t^2$$

is a non-autonomous second order *differential equation. There is* no solution.

Instead,

$$y'(t) = -\frac{1}{t}y(t) + t^3$$

admits a solution[15]

We want to discuss a particular class of differential equations:
consider

$$y'(t) = a(t)y(t) + b(t)$$

where $a(t)$ and $b(t)$ are real functions. Then the solution exists and is given by

$$y(t) = [k + B(t)]e^{A(t)} \tag{3}$$

where $A(t)$ is any function such that $A'(t) = a(t)$, and B is any function such that

$$B'(t) = b(t)e^{A(t)} \ \forall t$$

Proof.

$$\begin{aligned}(3) \Rightarrow y'(t) =& b(t)\cancel{e^{-A(t)}}\cancel{e^{A(t)}} + \underbrace{[k + B(t)]^e e^{A(t)}}_{y(t)} a(t) \\ =& a(t)y(t) + b(t)\end{aligned}$$

□

Remark 95. *If $F(t)$ is given by*

$$F(t) = \int_{t_0}^{t} f(s)ds \qquad \text{for some } t_0 \in R,$$

then $F'(t) = f(t)$ (this is the Fundamental Theorem of Integrated Calculus*).*

[15]See exercise 26 in the problem sets.

Therefore, we can write

$$F = \int^t f(s)ds \Rightarrow y(t) = [k + B(t)]e^{A(t)}$$
$$= \left[k + \int^t b(s)e^{-A(s)}dy\right] e^{A(t)}$$
$$= \left[k + \int^t b(s)e^{-\int^s a(u)du}ds\right] \cdot e^{\int^t a(s)ds}.$$

After finding an antiderivative, the recipe is quite easy to apply.

Example 96. *In chapter 15, we will see Dynamic Optimization. For the moment, we pospone the discussion of that example.*

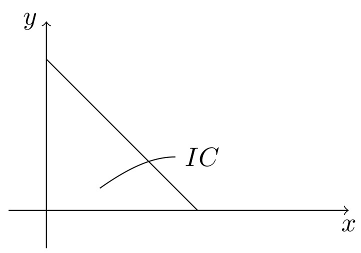
4.3 Dynamic Optimization

For the topic we will now treat, see also Lambert, Ch. 7.

4.3.1 Introduction

Economic problems can be classified in many ways: one way is the cathegorization "static" (without time being an essential aspect of the problem) vs. "dynamic" (with time).

Example 97. *The consumer's problem*



is a typical static *problem, while studying what happens, for instance, when the constraints change in time is a* dynamic *one.*

We can further subdivide dynamic problems between the ones *separable across time* and the ones which are not.

$$\max_{x_t: t=1,\ldots,T} \sum_{t=1}^{T} f(t, x_t)$$
$$\text{subject to } g(t, x_t) \leq b_t, \qquad t = 1, \ldots, T$$

this problem is separable. For instance, consider

$$f(t, x_t) = \frac{1}{(1+r)^t} h(x_t)$$

(where the natural interpretation of $1+r$ is of a *discount factor*). We can rephrase it as

$$\sum_{t=1}^{T} \left\{ \begin{matrix} \max_{x_t} f(t, x_t) \\ \text{subject to } g(t, x_t) \le b_t \end{matrix} \right\} \overset{\text{solution}}{\Rightarrow} \sum_{t=1}^{T} f(t, \hat{x}_t) \Rightarrow \hat{x} = (\hat{x}_1, \dots, \hat{x}_T);$$

the solution of the problem (in *discrete time*) is reached by solving a series of static problems - which we already know how to solve. If all dynamic problems were of this form, life would be easy.

This distinction also holds for problems in *continuous* time, where the sum is typically replaced by an integral:

$$\max_{x(t): 0 \le t \le T} \int_0^T f(t, x(t)) dt$$
$$\text{subject to } g(t, x(t)) \le b(t), \ 0 \le t \le T.$$

Again, it is clear that this maximization can be faced by considering independently each point in time:

$$\max_{x} f(t, x) \text{ subject to } g(t, x) \le b(t) \text{ for } 0 \le t \le T$$

$$\hat{x}(t) \ \forall t \in [0, T]$$
$$\Rightarrow \int_0^T f(t, \hat{x}(t)) dt.$$

We now consider problems which are *not* separable across time.

$$\max_{x_t} \sum_{t=1}^{T} f(t, x_t, x_{t-1})$$
$$\text{subject to } g(t, x_t, x_{t-1}) \le b_t \qquad \forall t = 1, \dots, T$$

Now, the decision taken in period $t-1$ influences what can be done in period t. Similarly, in the continuous case,

$$\max \int_0^T f(t, x(t), x'(t)) dt$$
$$\text{subject to } g(t, x(t), x'(t)) \le b(t) \qquad \forall t \in [0, T]$$

is not separable.

There are three approaches to dynamic optimization:

- *optimal control theory*
- *calculus of variations*
- *dynamic programming*.

Historically, calculus of variations precedes optimal control theory, which however is more general and complete.

On the other hand, both techniques work in continuous times, while dynamic programming works in discrete time. Since we have limited time available, we will concentrate on optimal control theory.

4.3.2 Two Examples of Dynamic Optimization

We consider two problems, and will try to "extract" from them a general formulation.

(a) an individual derives income from the interest paid at rate i on her savings S to be allocated between consumption C and new savings $S'(t) = I$.

$$iS(t) = \text{interest paid at moment } = C(t) + I(t).$$

Moreover, it is generally assumed that $C(t) \geq 0$, while $I(t)$ can have any sign (saving stocks *can* decline).

Moreover, we assume that initially $S(0) = S_0 > 0$.

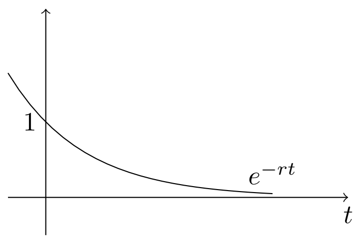
We can think of the following formalization:

$$\max_{C,S} \int_0^T e^{-rt} u(C(t))dt$$

$$\text{subject to } \begin{array}{rl} S'(t) = & iS(t) - C(t) \\ S(0) = & S_0 \\ S(T) \geq & 0 \end{array}$$

(where the final 0 doesn't really play an important role - it could be replaced with some other constant).

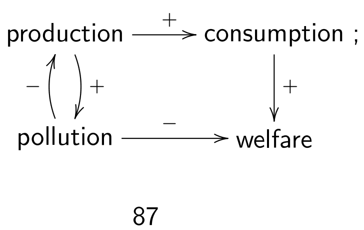
r is the *discount rate*:



$C : [0, T] \to \mathbb{R}$ is a *control* variable,
$S : [0, T] \to \mathbb{R}$ is a *state* variable.

Obviously, by choosing C one indirectly chooses S.

(b) A society unfortunately produces pollution of the air by CO_2, in the following way:



how can this problem be faced? Let

$$\begin{aligned}
M(t) =& \text{concentration of } CO_2 \text{ at time } t,\\
E(t) =& \text{rate of emission of } CO_2,\\
C(t) =& \underbrace{\underbrace{f(E(t))}_{\substack{\text{Production}\\ \text{function}}} - \underbrace{h(M(t))}_{\substack{\text{Pollution}\\ \text{damage}}}}_{\text{Production}}\\
& \text{with } f' > 0, h' > 0, f'' < 0, h'' > 0.
\end{aligned}$$

Of course, the concentration of CO_2 varies over time, as (by assumption):

$$\begin{aligned}
M'(t) =& aE(t) - bM(t), \text{ with}\\
a =& \text{ technological constant}\\
b =& \text{rate of dissipation of } CO_2 \text{ into the outer atmosphere.}
\end{aligned}$$

By the way, this model was *indeed* published 20-30 years ago in the American Revue.

Our decision problem is formalized as:

$$\max \int_0^T a^{-rt} U(f(E(t)) - h(M(t)))dt$$

$$\text{subject to } \begin{aligned} M'(t) =& \quad aE(t) - bM(t),\\ M(0) =& \quad M_0,\\ M(T) \leq& \quad M_t, \end{aligned}$$

where by M_t we indicate some fatal level of pollution. $E(t)$ is the control variable, and $M(t)$ is the state variable.

4.3.3 The Optimal Control Theory Format

Let

$$\begin{aligned}
x(t) =& \text{state variable}\\
u(t) =& \text{control variable ;}
\end{aligned}$$

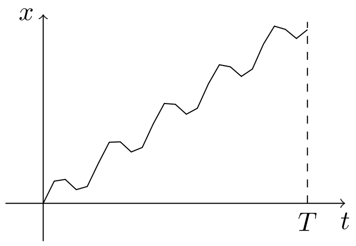
then, we would like to solve

$$\max \int_0^T f(t, x(t), u(t))dt \tag{P}$$

$$\text{subject to} \quad \begin{aligned} x'(t) =& \; g(t, x(t), u(t))\\ x(0) =& \; x_0\\ x(T) :& \text{ some condition to be specified.} \end{aligned}$$

The solution is a couple of functions $\hat{x}(\cdot)$, $\hat{u}(\cdot)$ which give rise to the *optimal trajectories*, or *optimal time paths*

$$\{(t, \hat{x}(t)) | t \in [0, T]\}$$
$$\{(t, \hat{u}(t)) | t \in [0, T]\}.$$



Finally, we can derive the *maximum value*

$$V = \int_0^T f(t, \hat{x}(t), \hat{u}(t)) dt$$

of

$$\begin{cases} \max \int_0^T f(t, x, u) dt \\ \text{s.t. } x' = g(t, x, u) \\ x(0) = x_0 \\ x(T) : \text{ some condition.} \end{cases}$$

4.3.4 Optimal Control Theory: a Lagrangian approach

Consider

$$\max \int_0^T f(t, x, u) dt$$
$$\text{subject to } x' = g(t, x, u) \ \forall t \in [0, T].$$

Then, define

$$L := \int_0^T \{f(t, x, u) - \lambda(t) [x'(t) - g(t, x, u)]\} dt \tag{PL}$$

with $\lambda(t)$ being the *costate variable*.

Remark 98. $L = L(x(\cdot), u(\cdot), \lambda(\cdot), x'(\cdot))$.

- $L(\hat{x}(\cdot), \hat{u}(\cdot), \lambda(\cdot), \hat{x}'(\cdot)) = V \ \forall \lambda(\cdot)$, *since the only difference between the two is the term in square brackets... but that term disappears (by assumption).*

- $\exists \hat{\lambda}(\cdot)$ *such that solving*

$$\max_{x(\cdot), u(\cdot)} L(x(\cdot), u(\cdot), \hat{\lambda}(\cdot), x'(\cdot)) = V$$

This is something we did not discuss when considering static optimization, but we can briefly verify why it is so.

Exercise 99. [16] *Let $f : \mathbb{R} \to \mathbb{R}$ be a concave function, and $g : \mathbb{R} \to \mathbb{R}$ a convex one.*

Given the problem

$$\max f(x) \text{ subject to } g(x) \leq b$$

show that x^* is an constrained maximizer if and only if x^* is an unconstrained maximizer of $L(x, \lambda^*)$, where λ^* is the multiplier at the solution of the constrained problem.

From Kuhn Tucker, we know that if (x^*, λ^*) is a solution, then

$$\frac{\partial L}{\partial x}(x^*, \lambda^*) = f'(x^*) - \lambda^* g'(x^*) = 0 \tag{1}$$
$$\lambda^*[g(x^*) - b] = 0$$

x^{**} is a solution to the unconstrained optimization problem

$$\max_x L(x, \lambda^*) = f(x) - \lambda^*[g(x) - b]$$
$$\Rightarrow \frac{\partial L}{\partial x}(x^{**}, \lambda^*) = f'(x^{**}) - \lambda^* g'(x^{**}) = 0 \tag{2}$$

From equation (1), we get that

$$\frac{f'(x^*)}{g'(x^*)} = \lambda^* \overset{(2)}{=} \frac{f'(x^{**})}{g'(x^{**})}$$

In particular, since we have sufficient conditions (on f and g) for the uniqueness of the solution, we get that indeed $x^* = x^{**}$.

From the static case, this result translates to the dynamic one.

For the moment, we know that to solve our original problem we want to maximize (in an unconstrained way) the Lagrange function. If we also find the right $\hat{\lambda}$, then it will give us the maximizer.

There is one problem that we will have to face: maximizing (PL) is easy if the problem is separable across time... otherwise, it is not, in general (x' and x appear both in the function to be maximized).

What we will do is eliminate *x', and consider:*

$$\int_0^T -\lambda(t) x'(t) dt.$$

[16] Problem 22

From the general formula $\int fg' = fg - \int f'g$ (integration by parts *) we get the above is equal to*

$$\begin{aligned}
&[-\lambda(t)x(t)]_0^T + \int_0^T \lambda'(t)x(t)dt \\
&= \int_0^T \lambda'(t)x(t)dt - \lambda(T)x(T) + \lambda(0)x(0) \\
\Rightarrow L &= \int_0^T \{f(t,x,u) + \lambda(t)g(t,x,u) + \lambda'(t)x(t)\}dt + \lambda(0)x(0) - \lambda(T)x(T);
\end{aligned}$$

we see that we have thrown out x' - we now only have to maximize the term in curly brackets. On the other hand, we added λ' in... but notice we don't maximize *with respect to it!*

30/11/10

We have started dynamic optimization and formulated the problem in a general mathematical formalization.

We have observed that maximizing Lagrange function with the "right" λ function, we get the solution of our maximization problem. However, maximizing the integral which appears in the Lagrange function is difficult if x' appears inside it (the problem is not separable), so we have to apply integration by parts to get rid of the problem.

Definition 100. *$H(\lambda, t, x, u) := f(t,x,u) + \lambda g(t,x,u)$ is called the* Hamiltonian *for the problem.*

We get now, as expression for the function L,

$$L = \int_0^T [H(\lambda, t, x, u) + \lambda' x]dt + \lambda(0)x(0) - \lambda(T)x(T),$$

and we can now introduce the following:

Theorem 101 (Pontryagin's Maximum Principle)**.** *If the solution $(\hat{x}, \hat{u})$ to the problem (P) exists, then there exists a function λ such that $\hat{x}$ and $\hat{u}$ maximize $H + \lambda' x$.*

The complete proof can be found in the original work: *Pontryagin et al. (1962)* or in the book by *Kamien and Schwarz (1981)*.

The theorem only gives a necessary condition, but it is not difficult to transform it into a sufficient one:

Theorem 102 (Mangasarian, 1966)**.** *If $f(t,x,u)$ is concave in x and u, and if $g(t,x,u)$ is linear in x and u, then the necessary condition of Pontryagin's theorem is also sufficient.*

In the two economic examples we introduced last time, in fact, these conditions will be satisfied, so the necessary condition of Pontryagin's theorem will also be sufficient.

So our goal is now to maximize indeed $H + \lambda' x$.

4.3.5 The Hamiltonian conditions

The problem is

$$\max_{x,u} H(\lambda, t, x, u) + \lambda' x.$$

This maximization yields necessary conditions:

$$\frac{\partial H}{\partial u} = 0 \qquad (1)$$

$$\frac{\partial H}{\partial x} + \lambda' = 0 \qquad \Longleftrightarrow \lambda' = -\frac{\partial H}{\partial x} \qquad (2)$$

$$x' = g(t, x, u) \qquad \Rightarrow \frac{\partial H}{\partial \lambda} = x' \qquad (3)$$

transversality condition. (4)

They are called *Hamiltonian conditions*.

4.3.6 Transversality condition

We first list the transversality conditions that we will adopt:

end-point condition	**transversality condition**
(1) $x(T) = x_T$	no condition
(2) $x(T) \geq x_T$	$\lambda(T) \geq 0,\ \lambda(T)[x(T) - x_T] = 0$
(3) $x(T) \leq x_T$	$\lambda(T) \leq 0,\ \lambda(T)[x(T) - x_T] = 0$
(4) x_T free	$\lambda(T) = 0$

and then try to justify them. If we insert the terminal conditions in the Lagrangian, we get

$$L = \int_0^T [H(\lambda, t, x, u) + \lambda' x] dt + \lambda(0)x(0) - \lambda(T)x(T) - \underbrace{\mu[x(T) - x_T]}_{=0 \text{ at solution}}.$$

Now we maximize $\max_{x,u} L$, and hence put

$$\begin{aligned} &\frac{\partial L}{\partial x(T)} = 0 \\ &\Rightarrow -\lambda(T) - \mu = 0 \\ \Longleftrightarrow\ &\lambda(T) = -\mu \\ &\Rightarrow V = \int_0^T [H(\lambda, t, \hat{x}, \hat{u}) + \lambda' \hat{x}] + \lambda(0)\hat{x}(0) - \cancel{\lambda(T)\hat{x}(T)} + \cancel{\lambda(T)\hat{x}(T)} - \lambda(T) x_T. \end{aligned}$$

We can hence now look at

$$\frac{\partial V}{\partial x_T} = -\lambda(T):$$

Case 2 : $x(T) \geq x_T$.

Consider what happens if we decrease x_T: the constraint is relaxed, becomes weaker, so the maximum value V can only increase. So

$$\frac{\partial V}{\partial x_T} \leq 0 \Rightarrow \lambda(T) \geq 0$$

Case 3 : $x(T) \leq x_T$. If now we increase x_T, again the maximum value of V can only increase:

$$\frac{\partial V}{\partial x_T} \geq 0 \Rightarrow \lambda(T) \leq 0$$

Case 1 : $x(T) = x_T$.

The sign of λ depends on the sign of $\frac{\partial V}{\partial x_T}$, which can be positive or negative - we can make no prediction.

Case 4 : $x(T)$ free. This means

$$0 = \frac{\partial L}{\partial x(T)} = -\lambda(T)$$

4.3.7 Interpreting the costate variables

$$V = \int_0^T [H(\lambda, t, \hat{x}, \hat{u}) + \lambda' \hat{x}]\, dt + \lambda(0) x_0 - \lambda(T) x_T$$

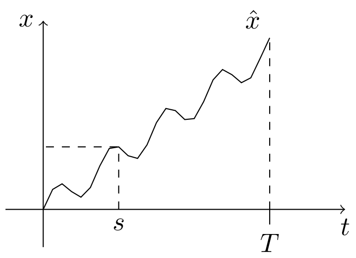
because $\hat{x}(0) = x_0$. Hence, we can now look at what happens when we change the initial value:

$$\frac{\partial V}{\partial x_0} = \lambda(0)$$

$\Rightarrow \lambda(0)$ express the sensitivity of the maximum value to a change in the initial value x_0 of the state variable.

In fact, more generally, at $s \in [0, T]$,

$$\frac{\partial V}{\partial x_s}(\hat{x}(s)) = \lambda(s).$$



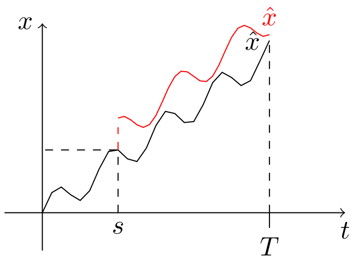
Consider

$$
\begin{aligned}
V = & \max \int_0^T f(t,x,u)dt \quad \text{subject to} \quad & x' = g(t,x,u) \\
& & x(0) = x_0 \\
& & x(T): \text{ some condition} \\
= & \max \int_0^s f(t,x,u)dt \quad \text{subject to} \quad & x' = g(t,x,u) \\
& & x(0) = x_0 \\
& & x(s) = \hat{x}(s) \\
+ & \max \int_s^T f(t,x,u)dt \quad \text{subject to} \quad & x' = g(t,x,u) \\
& & x(s) = \hat{x}(s) =: x_s \\
& & x(T): \text{ some condition}
\end{aligned}
$$

$$\overset{\text{def}}{=} V_1 + V_2$$

$$
\begin{aligned}
&\Rightarrow V_2 \int_s^T [H(\lambda, t, \hat{x}, \hat{u}] + \lambda' \hat{x} dt + \lambda(s) x_s - \lambda(T) x_T \\
&\Rightarrow \frac{\partial V_2}{\partial x_s}(\hat{x}(s)) = \lambda(s)
\end{aligned}
$$

How can we visualize this? The consumer had planned some consumption path, but at some point her savings may unexpectedly increase:



$\lambda(s)$ expresses the sensitivity of the maximum value to an exogenous change in the state variable at time $t = s$.

4.3.8 Using the Hamiltonian condition to solve problems

Recall problem A:

$$
\max \int_0^T e^{-rt} U(C) dt \quad \text{subject to} \quad
\begin{array}{rcl}
S' = & iS - C & \\
S(0) = & S_0 & (x = S) \\
S(T) \geq & 0 & (u = C)
\end{array}
$$

Let's form the Hamiltonian:

$$H = f + \lambda g = e^{-rt}U(C) + \lambda(iS - C)$$

$$\Rightarrow \text{Hamiltonian conditions:} \quad \begin{array}{c} (1)\frac{\partial H}{\partial C} = e^{-rt}U'(C) - \lambda = 0 \\ (2)\lambda' = -\lambda i \\ (3)S' = iS - C \\ (4)\lambda(T) \geq 0,\ \lambda(T)S(T) = 0. \end{array}$$

We now impose those conditions:

$$(2) \Rightarrow \lambda'(t) = -i\lambda(t)$$

Recall that we had seen the general case $y'(t) = \rho y(t)$, with solution $y(t) = y_0 e^{\rho t}$, so in this case we get

$$\lambda(t) = \lambda(0)e^{-it} \qquad \forall t.$$
$$\lambda(T) = \lambda(0)e^{-iT}$$

Now, we know $\lambda(0) = \frac{\partial V}{\partial S_0} > 0$, because if in the beginning there are more savings, it's plausible that the consumer will be better off (consume more) - it is obvious if she changes her consumption path projects, while if she doesn't reprogram then, she will just consume more in the first period (and we assume U is increasing in C). Hence, $\lambda(T) > 0$.

Having this, we can apply (4) and get, from the complementary slackness condition, that

$$S(T) = 0.$$

Now,

$$(1) \Rightarrow e^{-rt}U'(C(t)) - \underbrace{\lambda(0)e^{-it}}_{\lambda(t)} = 0$$
$$\iff U'(C(t)) = \lambda(0)e^{(r-i)t} \tag{5}$$

Therefore, if $(U')^{-1}$ exists, then

$$C(t) = (U')^{-1}\left(\lambda(0)e^{(r-i)t}\right).$$

It is quite typical that the derivative of U does exist - often, U is assumed to be concave.

A further assumption is then often a specification of the utility function. We will now assume:

$$U(C) = \ln C:$$

so that we can really solve the particular problem. We have

$$U'(c) = \frac{1}{C} \overset{(5)}{\Rightarrow} \frac{1}{C(t)} = \lambda(0)e^{(r-i)t}$$
$$\iff C(t) = \frac{1}{\lambda(0)e^{(r-i)t}}. \tag{6}$$

Once we find $\lambda(0)$, we are hence able to determine $C(t)$. Let's come back to S:

$$S'(t) = iS(t) - C(t).$$

Recall that

$$y'(t) = a(t)y(t) + b(t)$$

has as solution

$$y(t) = [k + B(t)]\, e^{A(t)}$$

where $A'(t) = a(t)$ and $B'(t) = b(t)e^{-A(t)}$.

Here: $a(t) = i$, $b(t) = -C(t) \overset{(6)}{=} -\frac{1}{\lambda(0)e^{(r-i)t}}$.

We now face the typical obstacle of this class of problems: find the antiderivative. In this case, however, it is particularly easy:

$$\begin{aligned}
A(t) =& it \\
\Rightarrow B'(t) =& b(t) = -\frac{1}{\lambda(0)e^{(r-i)t}} \cdot e^{-it} \\
=& -\frac{1}{\lambda(0)e^{rt}} = -\frac{1}{\lambda(0)}e^{-rt} \\
\Rightarrow B(t) =& -\frac{1}{r}\left(-\frac{1}{\lambda(0)}e^{-rt}\right) \\
=& \frac{1}{r\lambda(0)}e^{-rt}
\end{aligned}$$

Now we have $A(t)$ and $B(t)$, we can compute

$$\begin{aligned}
S(t) &= \left[k + \frac{1}{r\lambda(0)}e^{-rt}\right] e^{it} \qquad \forall t \\
\Rightarrow S(T) &= \left[k + \frac{1}{r\lambda(0)}e^{-rT}\right] e^{iT}
\end{aligned}$$

but we also know (from condition (4)) that $S(T) = 0$. Since $e^{it} \neq 0$, we have that

$$\begin{aligned}
k =& -\frac{1}{r\lambda(0)}e^{-rt} \\
\Rightarrow S(t) =& \left[-\frac{1}{r\lambda(0)}e^{-rt} + \frac{1}{r\lambda(0)}e^{-rt}\right] e^{it} \\
=& \frac{e^{it}}{r\lambda(0)}\left(e^{-rt} - e^{-rT}\right) \qquad \forall t.
\end{aligned}$$

In particular, this is true for $t = 0$, so that

$$\begin{aligned}
S(0) =& \frac{1}{r\lambda(0)}\left(1 - e^{-rT}\right) = S_0 \\
\iff \lambda(0) =& \frac{1}{rS_0}\left(1 - e^{-rT}\right) \\
\Rightarrow \hat{S}(t) =& \frac{e^{it}S_0}{1 - e^{-rT}}\left(e^{-rt} - e^{-rT}\right) \qquad \forall t
\end{aligned}$$

This is finally an explicit expression for the state variable S in time.
Of course, we want to also know consumption, so equation (6) gives us

$$\hat{C}(t) = \frac{rS_0}{(1-e^{-rT})\, e^{(r-i)t}} \qquad \forall t.$$

What is the economic significance of these expressions? Solutions obviously depend on r (interest rate) and i (discount rate).

- If, for instance, $r = i$, then $\hat{C}(t) = \frac{rS_0}{1-e^{-rT}}$, forall t. But t doesn't appear at all in this expression. Hence, under this assumption consumption is stable across time (and the consumer is running down initial savings, hence disinvesting).
- If instead $r < i$, of course $r - i < 0$ and hence
$$\frac{de^{(r-i)t}}{dt} < 0 \Rightarrow \frac{d\hat{C}(t)}{dt} > 0;$$
the consumer is forward-looking and wants to have a good retirement.
- Viceversa, if $r > i$,
$$\frac{d\hat{C}(t)}{dt} < 0:$$
the consumer *decreases* consumption in time.

Of course, in all three cases the consumer will arrive at T having disinvested everythin. But we may also look at what happens if $T = \infty$. This of course doesn't make sense for a single consumer, but still, it is interesting to verify (also because the assumption is not that meaningless in other contexts - i.e. considering a society instead than an individual).

$$\begin{aligned} T = \infty \Rightarrow \hat{C}(t) &= \frac{rS_0}{e^{(r-i)t}} \\ \Rightarrow \lim_{t\to\infty} \hat{C}(t) &= \begin{cases} rS_0 & \text{if } r = i \text{ (the consumer only consumes what he gets as interests)} \\ +\infty & \text{if } r < i \\ 0 & \text{if } r > i. \end{cases} \end{aligned}$$

This was just an example of the possibility to study *infinite orizons*.

4.4 The path to the steady state: diagrammatic analysis

Let's recall that $\hat{x}, \hat{u}$ were the solutions to

$$\max \int_0^T f(t,x,u)dt \qquad \text{subject to} \qquad \begin{aligned} x' &= g(t,x,u) \\ x(0) &= x_0 \\ x(T) &: \text{ some condition.} \end{aligned}$$

Now, if $T = \infty$, we can consider $\hat{x}(t), \hat{u}(t)$ for any t. Therefore, we can study

$$\lim_{t\to\infty} \hat{x}(t), \qquad \lim_{t+\infty} \hat{u}(t).$$

In that case:

- Hamiltonian conditions (1) - (3) remain valid (T didn't appear inside them),
- the transversality condition becomes, for example in case (3),

$$\lim_{t\to\infty} x(t) \leq \overline{x} \Rightarrow \lim_{t\to\infty} \lambda(t) \leq 0 \text{ and } \lim_{t\to\infty} \lambda(t)\,[x(t) - \overline{x}] = 0.$$

  (other cases are analog).

If $\lim \hat{x}(t)$ exists, then $t \to \infty$.

$$x^* := \lim_{t\to\infty} \hat{x}(t):$$

the limit will be called a *stationary state*. Questions:

1. how does x^* vary with the parameters? This is an exercise of *comparative statics*.
2. How does the path towards x^* vary with the parameters? This is an exercise of *comparative dynamics*.

We will face those question next time.

02/12/10

So far, we have derived the Hamiltonian conditions which are, according to the Pontryagin's theorem, *necessary* conditions which sometimes also become *sufficient*.

Recall that

$$\begin{aligned} V = \max \int_0^\infty & e^{rt} f(t,x,u)dt \\ \text{s.t. } x' &= g(t,x,u) \\ x_0 &= x_0 \\ x(\infty) &: \text{ some condition} \end{aligned}$$

$$\Rightarrow \frac{\partial V}{\partial_s}(\hat{x}(s)) = \lambda(s) \qquad \forall s \geq 0$$

and that this is the marginal value of the state variable at time $t = s$ *discounted back to time* $t = 0$, that is, at the moment of planning of the state variable.

Now recall that our Hamiltonian H was defined as

$$H = e^{-rt} f(t,x,u) + \lambda(t) g(t,x,u)$$

The discounting in λ is *implicit* in it. But we can rewrite the above as:

$$H = e^{-rt} \underbrace{\left[f(t,x,u) + \underbrace{e^{rt}\lambda(t)}_{m(t)} g(t,x,u) \right]}_{\text{current value Hamiltonian}}$$

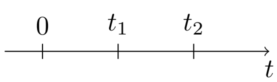
and

$$m(t) = e^{rt}\lambda(t) = \text{ current value multiplier,}$$

which is the marginal value of the state variable at time t in terms of values at t. We shall use this new multiplier in the solution to our problem, but before let's give a

Definition 103. (x^*, u^*) *is a* stationary *or* steady state *if for all* $t \geq 0$ *we have the following:*

$$x(t) = x^*, \; m(t) = m^* \Rightarrow x'(t) = m'(t) = 0$$



if $x(t) = x^*$, then $\lambda(t_1) \gneq \lambda(t_2)$, but it can still happen that $m_(t_1) = m(t_2)$.

Instead than seeing abstract justifications, let's get back to the problem (B):

$$\begin{aligned} \max \int_0^\infty & e^{-rt} U(f(E) - h(M))dt \\ \text{subject to } M' &= aE - bM \\ M(0) &= M_0 \\ \lim_{t \to \infty} M(t) &\leq \overline{M} \end{aligned}$$

As we had seen,

$$\begin{aligned} x =& M, \qquad u = E, \\ H =& f + \lambda g \\ =& e^{-rt}\left[U(\underbrace{f(E) - h(M)}_{C}) + m(aE - bM)\right] \end{aligned}$$

(recall $m = e^{rt}\lambda(t)$).
So the Hamiltonian conditions are

1. $$\frac{\partial H}{\partial E} = e^{-rt}\left[U'(C)f'(E) + ma\right] = 0;$$

   it is clear that the exponential is redundant:

   $$U'(C)f'(E) + ma \overset{(a)}{=} 0;$$

2. $$\begin{aligned} \lambda' =& -\frac{\partial H}{\partial M}, \qquad \lambda = e^{-rt}m \\ \Rightarrow& -r\cancel{e^{-rt}}m + \cancel{e^{-rt}}m' = -\cancel{e^{-rt}}\left[U'(C)(-h'(M)) - mb\right]. \end{aligned}$$

   So considering the *current value* allows us to get rid of the discounting factor. We can rewrite that as:

   $$m' \overset{(b)}{=} m(r + b) + U'(C)h'(M).$$

3. $$\frac{\partial H}{\partial \lambda} = M' \Rightarrow aE - bM = M'$$

   and we can use this now to express E:

   $$E \overset{(c)}{=} \frac{M' + bM}{a}.$$

So we have three equations for three variables (M, E and m). This is a system of differential equations which is not easy to solve, so we make a further simplifying assumption:

$$U(C) = C$$

(this is obviously the simplest simplification one can think of, but it's not particularly absurd - in the end well-being of countries is tipically measured with GDP).

Of course, this simplification implies $U'(C) = 1$, so we can simplify our expressions:

$$\begin{aligned} (a),(c) \Rightarrow& m \overset{(d)}{=} -\frac{f'\left(\frac{M'+bM}{a}\right)}{a} \\ (b) \Rightarrow& m \overset{(e)}{=} \frac{m' - h'(M)}{r + b}. \end{aligned}$$

This is already nicer, because we have *two* variables in two differential equations. . . and the system is *autonomous* (so one could hope to solve it).

At a stationary state, we must have $M' = m' = 0$. This means that

$$(d), (e) \Rightarrow$$

$$m = -\frac{f'\left(\frac{bM}{a}\right)}{a} \qquad \text{(M'=0)}$$

$$m = -\frac{h'(M)}{r+b}. \qquad \text{(m'=0)}$$

When we satisfy both equations (and only in that case), we are in a steady state.

We want now to try to draw qualitatively those functions. If for instance one equation has positive slope and one negative, it would be more plausible that they intersect:

$$\left.\frac{\partial m}{\partial M}\right|_{M'=0} = -\frac{f''(\frac{bM}{a}\frac{b}{a}}{a} \overset{f''<0}{>} 0$$

which has positive sign (because we have said that f'' is a concave function).

On the other hand,

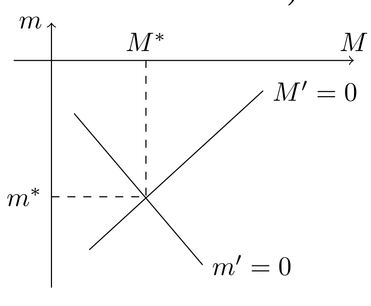
$$\left.\frac{\partial m}{\partial M}\right|_{m'=0} = -\frac{h''(M)}{r+b} \overset{h''>0}{<} 0$$

and this seems like good news.

Remark 104. *m, the costate variable, is negative. Does that mean anything? We know that*

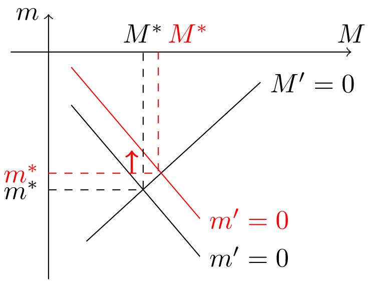
$$0 > \frac{\partial V}{\partial M_s} = \lambda(s) = e^{rs} m(s)$$

(where the first inequality can be seen as the intuition that when pollution increases, the maximum value decreases). This can help us draw the picture:



We can now make a couple of *comparative statics* exercises:

Example 105. *Let's consider an increase in r: the locus $M' = 0$ doesn't change, while $m' = 0$ does:*



Analogously, we could consider a change in a, which is the coefficient determining how polluting is technology. What would happen if technology became greener?

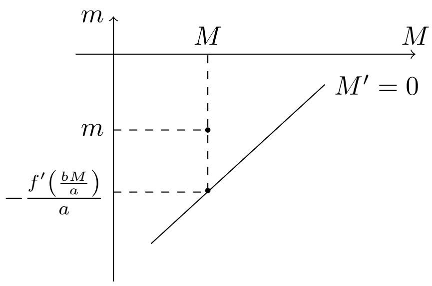
And in the same way, we can study what a change in b implies.

Let's notice that

$$E^* = \frac{bM^*}{a}.$$

The final lesson of this part is that we can say something economically meaningful even when we can't explicitly solve differential equations.

We will now address the *second* question we had formulated: *comparative dynamics*. Let's consider again the situation $M' = 0$.



If we take point $(M, m) \in \{M' = 0\}$, we have

$$m = -\frac{f'\left(\frac{bM}{a}\right)}{a};$$

if instead the point is *above*, we have

$$m > -\frac{f'\left(\frac{bM}{a}\right)}{a}$$

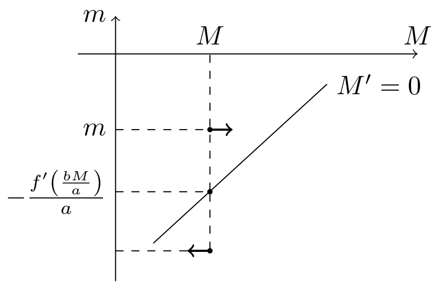
or more precisely

$$m \overset{(d)}{=} -\frac{f'\left(\frac{M'+bM}{a}\right)}{a} > -\frac{f'\left(\frac{bM}{a}\right)}{a}$$
$$\iff f'\left(\frac{M'+bM}{a}\right) < f'\left(\frac{bM}{a}\right);$$

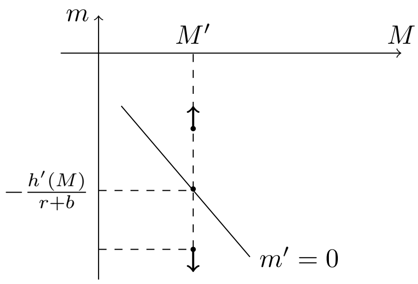
since we know f is concave, if the right hand term is bigger, it must have a bigger argument:

$$\frac{M'+bM}{a} > \frac{bM}{a} \Rightarrow M' > 0.$$

It is easy to verify that if we take a point *under* the line, the opposite holds:



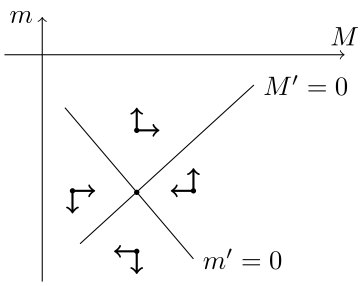
Consider now $\{m' = 0\}$:



If $(M, m) \in \{m' = 0\}$, we know $m = -\frac{h'(M)}{r+b}$.
If (M, m) is *above* $(m' = 0)$, then

$$m \overset{(e)}{=} \frac{m' - h'(M)}{r+b} > -\frac{h'(M)}{r+b} \iff m' > 0.$$

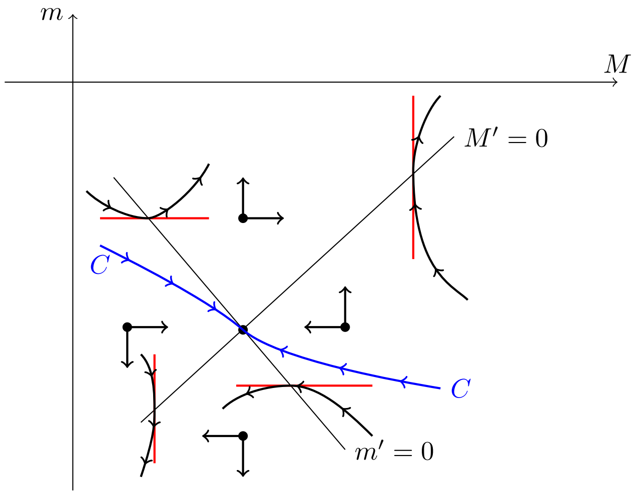
We are ready to put the two things together in a singl diagram:



Such a graph is called a *phase diagram*.

Looking at it, we may be a bit skeptical with respect to the equilibrium: no arrows point directly toward it: does it have any interesting meaning?

Let's first try to trace possible trajectories *when passing through the two lines*:



Again: what is the probability that this (blue) convergent path - which is (in this case) a monodimensional object in two dimensions - will ever by attained?

It can be shown that if the system of differential equations is stable, then the optimality conditions place the dynamic system on the convergent path, that is:

$$M_0 \mapsto m(0) \text{ s.t. } (M_0, m(0)) \in \overline{CC}.$$

4.4.1 Optimal control: an extension

We will now discuss an extension to the formulations seen so far[17]: consider

$$\begin{aligned} \max \int_0^T & f(x,u,t)dt \\ & \text{subject to } x' = g(x,u,t) \\ & \qquad\qquad x(0) = x_0 \\ & \qquad\qquad x(T): \text{ some condition} \\ & \qquad\qquad u(t) \in Z \subset \mathbb{R} \end{aligned}$$

We get the following Hamiltonian conditions: if (x^*, u^*)[18] is a solution, then $\exists \lambda^*(t)$ such that

$$H(t, x^*(t), u^*(t), \lambda^*(t)) \geq H(t, x^*(t), u, \lambda^*(t)) \qquad \forall u \in Z \tag{1}$$

(u^* is a maximizer). If we had an unconstrained problem, we would just put the partial derivative with respect to u equal to 0. Now, we must change approach (while conditions $(2)-(4)$ are exactly as before).

Exercise 106. *Let's solve Problem 32:*

[17] Used, for instance, in exercise 32 of the problem sets.

[18] This notation is different from the one used so far, in which we had $\hat{x}$ and $\hat{u}$

$$
\begin{aligned}
&\max \int_0^4 3x(t)dt \\
\textit{s.t. } & x' = x(t) + u(t) \\
& x(0) = 5,\ x(t) \textit{ free} \\
& u(t) \in [0,2]\ \forall t
\end{aligned}
$$

The Hamiltonian is:

$$H(t,x,u,\lambda) = 3x + \lambda(x+u)$$

and it gives the conditions:

$$
\begin{aligned}
H(t,x^*,u^*,\lambda^*) \geq & H(t,x^*,u,\lambda^*) \qquad \forall u \in [0,2] && (1)\\
\lambda^{*\prime} = -\frac{\partial H}{\partial x} = -3 - \lambda^* & && (2)\\
x^{*\prime} = \frac{\partial H}{\partial \lambda} = x^* + u^* & && (3)\\
\lambda^*(4) = 0 & && (4)\\
x^*(0) = 5 &
\end{aligned}
$$

Conditions are sufficient because we satisfy the hypothesis of Mangasarian theorem. Let's start imposing them:

$$(2) \Rightarrow \lambda'(t) = -\lambda(t) - 3$$

Remark 107. *In general, when we have*

$$y'(t) = ay(t) + b,$$

then $y(t) = ke^{at} - \frac{b}{a}$.

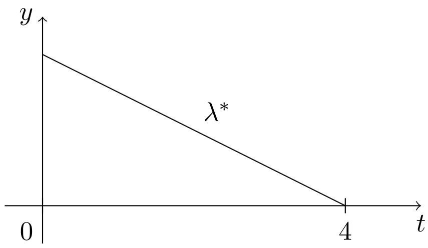
In fact:

$$
\begin{aligned}
y'(t) =& ake^{at} \\
=& a \underbrace{\left(ke^{at} - \frac{b}{a}\right)}_{y(t)} + b \\
=& ay(t) + b.
\end{aligned}
$$

In the present case, $a = -1$ *and* $b = -3$*. Therefore,*

$$
\begin{aligned}
\lambda(t) = ke^{-t} - 3 \qquad \forall t \\
\stackrel{(4)}{\Rightarrow} \lambda(4) = ke^{-4} - 3 = 0 \\
\iff k = 3e^4 \\
\Rightarrow \lambda(t) = 3e^{4-t} - 3
\end{aligned}
$$

This stated, we can draw this function:



Then,

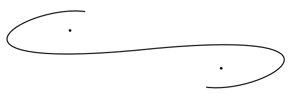
$$(1) \Rightarrow \max 3x + \lambda^* x + \lambda^* u \qquad u \in [0,2]:$$

we can maximize this simply by setting u *as large as possible...* $u^*(t) = 2$ *(for any value of* t*).*

$$\begin{aligned}(3) \Rightarrow & x'(t) = x(t) + u^* = x(t) + 2 \\ \Rightarrow & a = 1,\ b = 2 \\ \Rightarrow & x(t) = ke^t - 2 \qquad \forall t \in [0,4]\end{aligned}$$

In particular, this implies

$$\begin{aligned} x(0) = k - 2 \overset{(4)}{=} 5 \iff k = 7 \\ \Rightarrow x^*(t) = 7e^t - 2 \end{aligned}$$



[19]

[19] sic