Economics 897- Mathematics for Economists. Lecture Notes

Fabrizio Perri
University of Pennsylvania, McNeil Room 473.
Phone 898-0640 (Office),
545-8613 (Home)
e-mail: perri@econ.sas.upenn.edu
Lectures 10:30-12.00, 1.30-3.00 M-F
Office hours M,W After class

The second part of Econ 897 deals with Integration, Linear Algebra and Multivariate calculus. These notes are not a comprehensive exposition of these subjects but they contain the basic concepts you are supposed to pick up form this part of the course. For those who wish to do further work additional references will be given in class. They also contain exercises you are responsible for. Additional problems will be handed in in class. The final exam for this part of the class is scheduled for August 16 at 10.30 AM. These notes are based on similar notes prepared for past editions of Econ 897 by Alberto Trejos, Alex Citanna, Jacques Olivier and Marcos Lisboa. I thank them all but of course all remaining errors are entirely my own responsibility.

Part I

The Riemann Integral

Our basic problem in this section will be how to compute an integral. An integral is a number that has something to do with the area underlying a function. As we will discuss in class the concept of integral is closely related to that of summation (the symbol $\int$ represents a stretched S for sum) and to that of differentiation (one is in some sense the inverse of the other). First here we will give some basic definitions that will help us defining more precisely what we mean by an integral. Then we will give some conditions for the existence of integral and some properties of the integral itself. We'll then go through some basic theorems and techniques useful in computing the integrals and finally we'll consider some extensions of the concept.

1. Definitions

Consider for the rest of this section a function $f(x) : [a, b] \to \Re$, bounded on the interval $[a, b]$

Definition 1.1. *A partition of* $[a, b]$ *is a finite set of points:* $a = x_0 < x_1 < .... < x_{n-1} < x_n = b$.*We'll use the letter* P *to denote a partition.*

Definition 1.2. *The upper sum of* f *relative to* P *, denoted by* $U(f, P)$*, is*

$$\sum_{i=1}^{n} L_i(x_i - x_{i-1})$$

where $L_i = \sup_{x\in[x_{i-1},x_i]} f(x)$.
Analogously we define the lower sum $L(f, P) = \sum_{i=1}^{n} l_i(x_i - x_{i-1})$
where $L_i = \inf_{x\in[x_{i-1},x_i]} f(x)$.

Definition 1.3. *Let* $\wp(a, b)$ *be the set of all possible partitions over* $[a, b]$.*The upper Riemann integral of* f *over* $[a, b]$ *is defined as :*

$$\inf_{P\in\wp} U(f, P) = \overline{\int_a^b} f(x)dx$$

Analogously the lower Riemann integral of f over $[a, b]$ *is defined as*

$$\sup_{P\in\wp} L(f, P) = \underline{\int_a^b} f(x)dx$$

Definition 1.4. *The function $f(x)$ is (Riemann) integrable over $[a, b]$ if*

$$\overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx = \int_a^b f(x)dx$$

The number denoted by $\int_a^b f(x)dx$ is the definite Riemann integral of $f(x)$ over $[a, b]$.

These definitions are not very useful in computing integrals although in some cases they can be applied directly as in the following example:

Example 1.5. *Suppose we want to compute the integral of $f(x) = 1$ on the interval $[0, 1]$.It's easy to see that for every partition we take $U(f, P) = L(f, P) = 1$ so applying the last two definitions $\overline{\int_a^b} dx = \underline{\int_a^b} dx = \int_a^b dx = 1$. (You might say there is a better way of computing that integral and you might be right...).*

We'll now state and proof a theorem that will be useful later.

Theorem 1.6. *Take two arbitrary partitions P_1 and P_2 over $[a, b]$. Then*

$$U(f, P_1) \geq L(f, P_2) \qquad \forall P_1, P_2$$

Proof. *Consider the partition $P_3 = P_1 \cup P_2$. It's easy to see that $U(f, P_1) \geq U(f, P_3)$ (As the partition gets finer the upper sums decrease) and that $L(f, P_1) \leq L(f, P_3)$ (and the lower sums increase). From the definitions of upper and lower sums also follows that $U(f, P_3) \geq L(f, P_3)$. Combining the three inequalities we have*

$$U(f, P_1) \geq U(f, P_3) \geq L(f, P_3) \geq L(f, P_2)$$

■

The previous theorem states that every upper sum is larger than any lower sum. It's very intuitive if one thinks of the geometrical interpretation of lower and upper sums.

2. Conditions for integrability

In this section we present some conditions that help us determining whether the integral of a certain function exist.

Theorem 2.1. *(Riemann necessary and sufficient condition for integrability) .*
Consider the function $f(x) : [a, b] \to \Re$, bounded on the interval $[a, b]$. f is integrable if and only if $\forall \epsilon > 0$ $\exists P$ over $[a, b]$ s.t.

$$U(f, P) - L(f, P) < \epsilon$$

that is

$$\sum_{i=1}^{n} (L_i - l_i)(x_i - x_{i-1}) < \epsilon$$

Exercise 2.1. *Prove the previous theorem.*

The following theorems present some sufficient (but not necessary) conditions for integrability.

Theorem 2.2. *If $f(x) : [a, b] \to \Re$ is continuous then it is integrable.*

Exercise 2.2. *Prove the previous theorem.*

Theorem 2.3. *If $f(x) : [a, b] \to \Re$ has a finite number of discontinuities and it is bounded then it is integrable.*

Theorem 2.4. *If $f(x) : [a, b] \to \Re$ is monotone then it is integrable.*
Proof. *Consider the partition P composed by n equally distant points so that each sub-interval in the partition has length $\frac{b-a}{n}$. We then have*

$$U(f, P) - L(f, P) = \frac{b-a}{n} \sum_{i=1}^{n} (L_i - l_i)$$

Assume now w.l.o.g. that f is a monotonically increasing function. We can write

$$\frac{b-a}{n} \sum_{i=1}^{n} (L_i - l_i) =$$

$$= \frac{b-a}{n} (f(x_1) - f(a) + f(x_2) - f(x_1) + ... + f(b) - f(x_{n-1}))$$

$$= \frac{b-a}{n} (f(b) - f(a))$$

then given ϵ>0 pick n s.t.

$$n > \frac{b-a}{\epsilon}(f(b) - f(a))$$

and the Riemann condition is satisfied. ■

Example 2.5. *Consider now the following function on* $[0, 1]$ $f(x) = 1$ *if* x *is rational and* 0 *otherwise. The function is not integrable.*

3. Properties of the Riemann integral

We now state some useful properties of the Riemann integral. You are strongly recommended to proof them as an exercise.

As a convention we assume the following:

$$\int_a^a f(x)dx = 0$$

$$\int_a^b f(x)dx = -\int_b^a f(x)dx$$

We can now list some important properties:

Consider f and g integrable on $[a, b]$ then

$k_1 f + k_2 f$ is integrable and we have:

$$\int_a^b k_1 f(x) + k_2 g(x)dx = k_1 \int_a^b f(x)dx + k_2 \int_a^b g(x)dx \tag{3.1}$$

where k_1 and k_2 are given constants.

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \qquad c \in [a, b] \tag{3.2}$$

$$if\ f(x) \geq m\ \forall x \in [a, b] \rightarrow \int_a^b f(x)dx \geq m(b-a) \tag{3.3}$$

$$if\ f(x) \geq g(x)\ \forall x \in [a, b] \rightarrow \int_a^b f(x)dx \geq \int_a^b g(x)dx \tag{3.4}$$

$|f(x)|\, is$ integrable and we have

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|\, dx \tag{3.5}$$

Exercise 3.1. *Prove 3.1 and 3.5..*

4. The fundamental theorem of calculus

In this section we'll derive formally the relationship between the operation of integration and differentiation. We first need a new version of the mean value theorem and an additional definition:

Theorem 4.1. *(Mean value theorem for integral calculus).If $f(x)$ is continuous on $[a, b]$ then $\exists\ \xi \in [a, b]$ s.t.*

$$\int_a^b f(x)dx = f(\xi)(b-a)$$

Proof. *If f is continuous on the interval then it has a maximum and a minimum on it (why ?) M and m. and therefore by 3.3 we have*

$$m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$$

therefore we can find a number $\lambda \in [m, M]$ satisfying

$$\int_a^b f(x)dx = \lambda(b-a)$$

but since $\lambda \in [m, M]$ and f is continuous by the mean value theorem we know there is a number $\xi \in [a, b]$ s.t. $f(\xi) = \lambda$. ■

Definition 4.2. *Let f be a continuous function on $[a, b]$ we define its indefinite integral the function*

$$F(x) = \int_a^x f(t)dt \qquad x \in [a, b]$$

We are now ready to prove the central theorem of this section:

Theorem 4.3. *(Fundamental theorem of calculus or Torricelli-Barrow theorem).*
If $f(t)$ is continuous $F(x)$ is continuous, differentiable and $F'(x) = f(x)$.

Proof. Consider

$$\Delta F(x) = \int_a^{x+\Delta x} f(t)dt - \int_a^x f(t)dt = \int_x^{x+\Delta x} f(t)dt.$$

By the previous mean value theorem we have that :

$$\int_x^{x+\Delta x} f(t)dt = f(\xi)\Delta x \qquad \xi \in [x, x+\Delta x]$$

and therefore

$$F'(x) = \lim_{\Delta x \to 0} \frac{\Delta F(x)}{\Delta x} = \lim_{\Delta x \to 0} f(\xi) = \lim_{\xi \to x} f(\xi)$$

and by the continuity of f

$$F'(x) = f(x)$$

■

4.1. Application

Definition 4.4. *A primitive of $f(x)$ is a function $g(x)$ such that $\frac{d}{dx}g(x) = f(x)$ as derivative.*

Often a primitive of f is denoted by the symbol $\int f(x)dx$ (without the extremes of integration). Since two functions that have the same derivative can at most differ by a constant (Can you prove it ?) and since we have seen that $\frac{d}{dx}\int_a^x f(t)dt$ $= f(x)$ we can characterize all primitives of f , denoted by Φ, with the following equation

$$\Phi(x) = \int_a^x f(t)dt + c \qquad c \text{ constant}$$

Letting $x = a$ we find $c = \Phi(a)$ and letting $x = b$ we find

$$\int_a^b f(x)dx = \Phi(b) - \Phi(a)$$

sometimes we write the previous expression as

$$\Phi(x)|_a^b$$

so the key object we need to evaluate $\int_a^b f(x)dx$ is a primitive of $f(x)$.

5. Methods for finding primitives

While computing derivatives is a straightforward technique for every function finding the primitive of a function can be very hard. Here we report some integration methods that work with relatively simple functions.

5.1. Immediate integrals

In the following table we report some primitives of basic functions

Function	Primitive
$x^\alpha (\alpha \neq -1)$	$\frac{x^{\alpha+1}}{\alpha+1} + c$
x^{-1}	$\log(x) + c$
e^x	$e^x + c$
$\sin(x)$	$-\cos(x) + c$
$\cos(x)$	$\sin(x) + c$
$\frac{1}{1+x^2}$	$\arctan(x) + c$
$\frac{1}{\sqrt{1+x^2}}$	$\arcsin(x) + c$

These integrals can be used as a basis for a guess and verify method:

Exercise 5.1. *Find* $\int a^x dx$

The following two methods transform the original integral in another one that is (hopefully) easier to compute.

5.2. Integration by substitution

Theorem 5.1. *Let $f(x)$ a continuous function over an interval and let $x = \phi(t)$ a C1 function with $\phi'(t) \neq 0$. Then we have :*

$$\int f(x)dx = \left[\int f(\phi(t))\phi'(t)dt \right]_{x=\phi(t)}$$

Proof. Let $F(x)$ be a primitive of $f(x)$ then $F(\phi(t))$ is a primitive of $f(\phi(t))\phi'(t)$. Applying infact the chain rule we have that

$$F'(\phi(t)) = F'(\phi(t))\phi'(t) = f(\phi(t))\phi'(t)$$

therefore we have

$$\int f(x)dx = F(x) + c = F(\phi(t)) + c = \int f(\phi(t))\phi'(t)dt$$

■

5.2.1. Procedure

- Find a suitable $x = \phi(t)$
- Compute $dx = \phi'(t)dt$
- Substitute for x and dx in the original integral.

When you use the substitution method with a definite integral always remeber to change the integration bounds.

Exercise 5.2. *Compute* $\int_0^1 \frac{dx}{x^2+4}$ *using the substitution* $x = 2t$

5.3. Integration by parts

Let u and v two $C2$ functions. We have:

$$\int u(x)v'(x)dx = u(x)v(x) - \int u'(x)v(x)dx$$

u is called the finite factor while v is called the differential factor.

Exercise 5.3. *Prove the above equality.*

Exercise 5.4. *Compute* $\int \log x dx$ *using log(x) as finite factor and 1 as differential factor.*

6. Improper integrals

In this section we consider the integrability of unbounded functions or functions with unbounded support

Let f be a function continuous everywhere on $[a, b)$ and unbounded in a neighborhood of b and suppose we want to compute $\int_a^b f(x)dx$. We take a point $\xi \in [a, b)$ and we compute the following limit :

$$\lim_{\xi \to b} \int_a^{\xi} f(x)dx = \lim_{\xi \to b} F(\xi) - F(a)$$

if the limit exists finite then $f(x)$ is improperly integrable on $[a, b]$.

Exercise 6.1. *Compute* $\int_0^1 \frac{1}{\sqrt{x}} dx$

Similarly we can compute the (improper) integral of f continuous on $[a, \infty)$ by computing the limit (if it exists and is finite):

$$\lim_{\xi \to \infty} \int_a^{\xi} f(x)dx = \lim_{\xi \to \infty} F(\xi) - F(a)$$

Exercise 6.2. *Compute* $\int_0^\infty e^{-x}dx$.

The previous method for computing improper integrals are applicable when we know the primitive of the integrand function. If we don't we can at least check their existence using the following two theorems.

Theorem 6.1. *If f is continuous everywhere on $[a, b)$ and unbounded in a neighborhood of b, $\int_a^b f(x)dx$ exists if*

$$\lim_{x \to b} \frac{f(x)}{\frac{1}{x-b}} = 0$$

Theorem 6.2. *If f is continuous on $[a, \infty)$, $\int_a^b f(x)dx$ exists if*

$$\lim_{x \to \infty} \frac{f(x)}{\frac{1}{x}} = 0$$

Exercise 6.3. *Verify the existence of* $\int_{-\infty}^{+\infty} e^{-x^2}dx$.

7. The Riemann Stieltjes integral

In this section we consider a generalization of the Riemann integral that is particularly useful in statistics and econometrics. Consider a function f bounded on $[a, b]$ (integrand) and function g bounded and monotone on $[a, b]$ (integrator). Take once again a partition P and define:

$$U(f, g, P) = \sum_{i=1}^{n} (g(x_i) - g(x_{i-1})) \sup_{x \in [x_{i-1}, x_i]} f(x)$$

$$L(f, g, P) = \sum_{i=1}^{n} (g(x_i) - g(x_{i-1})) \inf_{x \in [x_{i-1}, x_i]} f(x)$$

The Riemann Stieltjes integral of $f(x)$ with respect to $g(x)$ is the common value (if it exists) defined by the following equality:

$$\inf_{P\in\wp} U(f,g,P) = \sup_{P\in\wp} L(f,g,P)$$

and is denoted by the symbol $\int_a^b f(x)dg$.

Theorems and conditions similar to those seen before apply to the Riemann-Stieltjes integral too. In particular the Riemann condition is necessary and sufficient for existence while monotonicity and continuity of the integrand are sufficient conditions.

If the integrator is a step function and the integrand is continous then the R.S. is easy to compute as shown in the following theorem:

Theorem 7.1. *Let g be a step function on $[a,b]$ with jumps $g_1, ...g_n$ at $x_1, ...x_n$ and let f be a continous function on $[a,b]$ then we have*

$$\int_a^b f(x)dg = \sum_{i=1}^{n} f(x_i)g_i$$

Notice though that if the integrand is discontinouos and the integrator has a finite number of discontinuities then the R.S. integral may fail to exist.

Exercise 7.1. *Verify the integrability of the following function over the interval [1,4]:*

$$f(x) = \begin{matrix} -1 & x \in [1,2) \\ 1 & x \in [2,4] \end{matrix}$$

w.r.t the integrator $g(x) = f(x)$.

The following theorem is useful in reducing the R.S. integral to a R. integral

Theorem 7.2. *If $f(x)$ is continuous on $[a,b]$ and g is differentiable with continuous derivative we have*

$$\int f(x)dg = \int f(x)g'(x)dx$$

8. The Leibniz's rule

We will now present a theorem that is useful when the integrand function depends also on a parameter that is also in the integrations bounds

Theorem 8.1. *Let f be a continous function with continous partial derivative with respect to a parameter a and let p and q be differentiable function. Consider the function*

$$F(a) = \int_{p(a)}^{q(a)} f(x, a)dx$$

then

$$F'(a) = \int_{p(a)}^{q(a)} \frac{df(x, a)}{da} dx + f(q(a), a)q'(a) - f(p(a), a)p'(a)$$

Part II
Linear Algebra

Linear algebra is the theory of linear functions. This theory has a wide range of applications and so we start considering a very general framework.

9. Vector Spaces

So far our space was $\Re$, that is the real line. Now we will enter a in a more complex environment called **vector space**. A vector space is a space of *vectors*, together with rules for adding vectors and multiplying them with elements belonging to a *field*.

Definition 9.1. *Let K be a subset of the complex numbers C. K is a field if it satisfies the following conditions:*

- If $x \in K$ and $y \in K$ then $x + y \in K$ and $xy \in K$
- If $x \in K$ then $-x \in K$ and if $x \neq 0$ then $x^{-1} \in K$
- 0 , $1 \in K$

When we do not specify a field we implicitly assume that the relevant field is $\Re$ that is the set of reals together with addition and multiplication defined in the standard way.

Exercise 9.1. *Determine if the following sets are fields: Q (the set of rationals) N (the natural numbers) , the interval on the real line* $[0, 1]$

Definition 9.2. *A vector space V over the field K is a set of objects together with the operations +: V x $V \rightarrow V$ (sum) and *: K x $V \rightarrow V$ (scalar multiplication) satisfying the following properties:*

1. Given u , v , w elements of V we have: $(u+v)+w = u+(v+w)$ (Associativity for sum)

2. $\exists$ O$\in V$ satisfying O $+u = u+$ O $= u$ $\forall u \in V$ (Identity for sum)

3. $\forall u \in V$ $\exists(-u) \in V$ s.t. $u + (-u) =$ O (Inverse element for sum)

4. Given u , v elements of V : $u + v = v + u$ (Commutativity)

5. Given u , v elements of V and $c \in K$ we have $c * (a + b) = c * a + c * b$

6. Given a, b elements of K and $u \in V$ we have $(a + b) * u = a * u + b * u$ (Distributive laws)

7. Given a, b elements of K and $u \in V$ we have $(a * b) * u = a * (a * u)$ (Associativity for scalar product)

8. $\forall u \in V$ we have $1 * u = u$ (identity for scalar product).

Exercise 9.2. *Given u, v elements of V such that $u + v = u$ show that $v =$O.*

Since we have not defined what a vector is at this stage the concept of vector space is a very general one and can encompass fairly complicated spaces as the space of infinite sequences or spaces of functions.

Exercise 9.3. *Consider the set of bounded and continuous functions $f : [0, 1] \rightarrow \Re$. Define addition and scalar product over $\Re$ and show that it is a vector space. Is the space of monotonic functions a vector space ?*

Another useful concept is that of subspace:

Definition 9.3. *A set $W \subseteq V$ (vector space) is a vector subspace if :*

1. Given u , v elements of W, $u + v \in W$

2. Given $u \in W$ and $c \in K$, $c * u \in W$

3. O$\in W$

Exercise 9.4. *Define a linear combination of n vectors ($v_1, v_2, ..., v_n$) elements of V to be the expression $\sum_{i=1}^{n} v_i * \alpha_i$ where $\alpha_i \in K$. Show that the set of all linear combinations of ($v_1, v_2, ..., v_n$) is a vector subspace.*

Exercise 9.5. *Consider the vector space C^n (The space of complex vectors) with the field C. Is $\Re^n$ a vector subspace of that field ?*

10. Linear independence, bases and dimension of a vector space.

In this section we develop the concept of bases that is a collection of vectors that can be used to construct an entire vector space. Notice that still we do not specify what a vector is so the following concepts apply to general vector spaces. (Notationally now we will omit the symbol * to denote the scalar product)

Definition 10.1. *The vectors* $(v_1, v_2, ..., v_n)$ *elements of* V *are linearly independent if*

$$\sum_{j=1}^{n} \alpha_i v_i = O \rightarrow \alpha_i = 0 \qquad \alpha_i \in K \qquad \forall i$$

The vectors are linearly dependent if there are n scalars $(\alpha_1, \alpha_2, \ldots\ \alpha_n)$ $\alpha_i \neq 0$ *for some i such that* $\sum_{j=1}^{n} \alpha_i v_i = O$.

Exercise 10.1. *Show that O cannot be a part of a collection of linear independent vectors.*

Exercise 10.2. *Show that n vectors are linearly dependent if and only if one of them can be expressed as linear combination of the others.*

Exercise 10.3. *Show that if* $(v_1, v_2, ..., v_n)$ *are linearly independent then* $(v_1, v_2, ..., v_{n-1})$ *are linearly independent too and that if* $(v_1, v_2, ..., v_n)$ *are linearly dependent then* $(v_1, v_2, ..., v_{n+1})$ *are linearly dependent too.*

Exercise 10.4. *In the space of functions show that* x *and* x^2 *are linearly independent and that* x *and* $3x$ *are linearly dependent.*

Definition 10.2. *The vectors* $(v_1, v_2, ..., v_n)$ *elements of* V *are said to generate* V *if* $\forall u \in V\ \exists(\alpha_1, \alpha_2, \ldots\ \alpha_n) \in K$ *such that* $u = \sum_{i=1}^{n} \alpha_i v_i$.

Definition 10.3. *The vectors* $(v_1, v_2, ..., v_n)$ *elements of* V *are a basis for* V *if they generate* V *and are linearly independent. If* $(v_1, v_2, ..., v_n)$ *are a basis for* V *and* $w \in V$ *then we have*

$$w_= \sum_{i=1}^{n} v_i a_i \qquad a_i \in \Re.$$

The numbers $a_1, ...a_n$ *are called the coordinates of* w *with respect to the basis* $(v_1, v_2, ..., v_n)$.

Theorem 10.4. *Let $(v_1, v_2, ..., v_n)$ be a basis for a vector space V and w be an element of V. Let $a_1, ...a_n$ be the coordinates of w with respect to the basis $(v_1, v_2, ..., v_n)$. Then $a_1, ...a_n$ are uniquely determined.* ***Proof.*** *Assume there is a different set of coordinates $(b_1, ...b_n)$ s.t. $w = \sum_{i=1}^{n} v_i b_i$. This implies that $O = \sum_{i=1}^{n} v_i(b_i - a_i)$ but since by assumption $\exists\ i$ s.t. $b_i \neq a_i$ then this contradicts the fact that the v_i are linearly independent.* ■

Theorem 10.5. *Let V be a vector space over K. Let $(v_1, v_2, \ldots\ v_m)$ be a basis for V. Then the vectors $(w_1, w_2, ..., w_n)$, $n > m$ are linearly dependent.* ***Proof.*** *The proof is by contradiction and by induction. Assume that $(w_1, w_2, ..., w_n)$ are linearly independent. Then $w_i \neq O\ \forall i$. Since $(v_1, v_2, \ldots\ v_m)$ constitute a basis we can write*

$$w_{1=} \sum_{i=1}^{n} v_i a_i$$

Since $w_1 \neq 0$ we know $\exists a_i \neq 0$. W.l.o.g. (we can always renumber the vectors) we assume $a_1 \neq 0$ so we can write

$$v_1 = \frac{1}{a_1}(w_1 - \sum_{i=2}^{n} v_i a_i)$$

So the vectors $(w_1, v_2, \ldots\ v_m)$ generate v_1 and since $(v_1, v_2, \ldots\ v_m)$ generate the entire space $(w_1, v_2, \ldots\ v_m)$ do the same. (Why ?). Now we want to show that if a set of vectors $(w_1..w_r, v_{r+1}, \ldots\ v_m)$, $1 \leq r < m$ generate the entire space so do the set $(w_1..w_{r+1}, v_{r+2}\ \ldots\ v_m)$. To this end it suffices to show that $(w_1..w_{r+1}, v_{r+2}\ \ldots\ v_m)$ generate v_{r+1}. Since by assumption $(w_1..w_r, v_{r+1}, \ldots\ v_m)$ generate the entire space we can write:

$$w_{r+1} = \sum_{i=1}^{r} b_i w_i + \sum_{i=r+1}^{m} c_i v_i$$

From our assumption that $(w_1, w_2, ..., w_n)$ are linearly independent and from exercise 10.3 we have that $\exists i$ s.t. $c_i \neq 0$. W.l.o.g. we assume $i = r + 1$ and we can write

$$v_{r+1} = \frac{1}{c_{r+1}} \left(w_{r+1} - \sum_{i=1}^{r} b_i w_i - \sum_{i=r+2}^{m} c_i v_i \right)$$

and this proves that $(w_1..w_{r+1}, v_{r+2}\ \ldots\ v_m)$ generate v_{r+1} and therefore the entire space. Applying this reasoning repeatedly we can show that $(w_1, w_2, ..., w_m)$

generate the entire space and therefore there are m scalars $(d_1, d_2, ..., d_m)$ such that

$$w_{m+1} = \sum_{i=1}^{m} d_i w_i$$

contradicting the initial assumption that $(w_1, w_2, ..., w_n)$ are linearly independent. ■

Corollary 10.6. *Let V be a vector space and suppose that a basis is composed by n vectors, then every other basis is composed by the same number of vectors.*

Definition 10.7. *Let V be a vector space with a basis composed by n elements. The vector space is said to have dimension n. If the vector spaces is composed only by the O element then it is said to be 0 dimensional while if has space has a set of linearly independent vectors for every n it is called infinite dimensional.*

Theorem 10.8. *Let V ba space of dimension n and $(v_1, ...v_n)$ be a set of linearly independent vectors. Then $(v_1, ...v_n)$ constitute a basis* ***Proof.*** *We have to show that $(v_1, ...v_n)$ generate V. Take $w \in V$. By the previous theorem $(v_1, ...v_n\ w)$ are linearly depndent so we have $a_0 w + a_1 v_1 + ..a_n v_n = O$. Since it cannot be $a_0 = 0$ (otherwise we would violate the hypotesis that $(v_1, ...v_n)$ are linearly independent) we can express w in function of the $(v_1, ...v_n)$.* ■

Exercise 10.5. *Find a basis and show that is indeed a basis in the space of second degree polynomials ($a + bx + cx^2$).*

11. The vector space $\Re^n$

We will now focus on a very important vector space, that is the space of ordered arrays of n real numbers. In this space a vector is defined as a (row) array x=$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$ $x_i \in \Re$. The zero vector is defined as O=$\begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}$. The sum of two vectors x,y elements of $\Re^n$ is defined as the vector $\begin{pmatrix} x_1 + y_1 & x_2 + y_2 & \cdots & x_n + y_n \end{pmatrix}$ and the scalar product k x is the vector$\begin{pmatrix} kx_1 & kx_2 & \cdots & kx_n \end{pmatrix}$. Each element of a vector is also called a component.

Exercise 11.1. *Show that $\Re^n$ together with sum and scalar product defined above is indeed a vector space.*

All the concepts we have seen in the previous section apply readily in $\Re^n$. For example it is to verify that a basis in $\Re^n$ is given by the set of vectors $(e_1, e_2, \cdots, e_n)$ where e_i is the vector composed by all 0 and 1 in the i^{th} position (this is also called canonical basis). From that follows that the dimension of $\Re^n$ is n.

Exercise 11.2. *Consider the following subsets of $\Re^2$ and determine if they are subspaces:*

1. The vectors with the first and second component equal (i.e. (-1,-1), (0,0),(1,1),..)
2. The vectors with positive components.
3. The vectors with integer components.
4. The vectors that solve the following equation: $3x_1 + x_2 = 0$.

Exercise 11.3. *Let (x_1, x_2) and (y_1, y_2) be two vectors in $\Re^2$. Show that they are linearly independent if and only if $x_1 y_2 - x_2 y_1 \neq 0$.*

We'll now define two additional operations in $\Re^{n.}$

Definition 11.1. *Given x,y vectors in $\Re^n$ we define the (Euclidean) inner product of the two vectors (xy) as the function :$\Re^n \times \Re^n \to \Re$ that associates to the vectors x and y the number $(x_1 y_1 + x_2 y_2 + \cdots x_n y_n)$. x and y are orthogonal if xy=0 (This concept has a geometric interpretation in $\Re^2$).*

The scalar product has the following properties easy to verify:

1. $(k_1 x + k_2 y)z = k_1 xz + k_2 yz$ with $k_1 k_2$ constants and x,y elements of $\Re^{n.}$.
2. $xy = yx$
3. $xx \geq 0$ and $xx = 0$ if and only if $x = O$.

Remark 1. *Any function that satisfy the properties above is called a scalar product.*

The following exercise shows why scalar product is so important in economics.

Exercise 11.4. *Italy each year produces 1000 tons of spaghetti and 2000 gallons of wine with prices of 20 and 30 lire per unit. As intermediate inputs Italians import 500 tons of grape, 500 tons of wheat and 100 tractors at the price of 10, 20 and 50 lire per unit. Express in vector notation, using the scalar product, the equation describing the national product of Italy.*

Exercise 11.5. *Given S subspace of* $\Re^n$ *consider the space S⊥ defined as* $\{x : x \in \Re^n, xs = 0, \forall s \in S\}$, *that is the space of vectors orthogonal to all the elements in S. Show that S⊥ is a subspace.*

Exercise 11.6. *Consider S⊂* $\Re^2$ *composed by the vectors (t,2t) t∈* $\Re$. *Find S⊥.*

Definition 11.2. *The norm of a vector x∈* $\Re^n$, *denoted as* $\|x\|$, *is a function that associates to every vector the square root of the scalar product of the vector with itself that is* $\|x\| \equiv \sqrt{xx} .\equiv \sqrt{\sum_{i=1}^n x_i^2}$

We are now ready to state and prove two important inequalities:

Theorem 11.3. *(Cauchy-Schwarz inequality) Given x,y elements of* $\Re^n$ *we have:*

$$|xy| \leq \|x\| \, \|y\|$$

and the inequality holds with equality if and only if x and y are colinear ($\exists \lambda \in \Re$ *s.t. x=λ y).* ***Proof.*** *Take* $\lambda \in \Re$ *and write:*

$$0 \leq \sum_{i=1}^n (x_i + \lambda y_i)^2 = \sum_{i=1}^n x_i^2 + 2\lambda \sum_{i=1}^n x_i y_i + \lambda^2 \sum_{i=1}^n y_i^2$$

the previous expression can be seen as a quadratic equation in λ *and since it must be always positive its* Δ *must be always less or equal to zero therefore:*

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

from which we have

$$|xy| \leq \|x\| \, \|y\|$$

Notice then that if x=λ y then $|xy| = \|x\| \, \|y\|$ *follows from the definition of norm and inner product while if* $|xy| = \|x\| \, \|y\|$ *then the above* $\Delta = 0$ *and therefore there exist a number* λ *s.t.* $\sum_{i=1}^n (x_i + \lambda y_i)^2 = 0$ *that in turn implies x and y are colinear since* $x_i + \lambda y_i = 0 \ \forall i$. ■

Corollary 11.4. *(Triangle inequality) Given x,y elements of $\Re^n$ we have:*

$$\|x+y\| \leq \|x\| + \|y\|$$

Exercise 11.7. *Prove the triangle inequality.*

12. Matrices

We will now consider a different object useful in linear algebra that is a *matrix.* Let K be a field and m and n two integers (m,n≥ 1). A matrix of dimension m×n is the following array of elements from the field K:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & & & \\ \vdots & & & & \\ a_{m1} & a_{m2} & & & a_{mn} \end{pmatrix}$$

The matrix has m rows and n columns and the first index of a_{jk} denotes its row while the second denotes its column. With the notation $A_{i.}$ we denote the i^{th} row of A while with $A_{.i}$ the i^{th} column. A matrix is said to be square if m=n. Addition and scalar multiplication for matrices are analogous to those defined for vectors in $\Re^n$ namely the sum of two m×n matrices A,B is an m×n matrix (C) with generic element $c_{jk} = a_{jk} + b_{jk}$ while the scalar product of a matrix A and a scalar k is the matrix B with generic element $b_{jk} = ka_{jk}$. It is straightforward to show that the set of matrices of a given dimension with addition and scalar product defined above is a vector space: in this space the zero element is given by the matrix in which each element is equal to 0. The space of matrices of dimension m×n over K is often denoted by $\mathrm{M}_{(m,n)}(K)$.

Exercise 12.1. *What is the dimension of the space of m×n matrices ? Give a basis for this space.*

Definition 12.1. *Let A be a (m×n) matrix. the (n×m) matrix B is called the transpose of A (denoted also as A′) if $b_{jk} = a_{kj}$. In other words the transpose of a given matrix is obtained by changing rows into columns and viceversa. A matrix A is symmetric if $A = A'$.*

Definition 12.2. *Let A be a (n×n) square matrix. The elements ($a_{11}, a_{22}, ..., a_{nn}$) are called diagonal elements or elements on the main diagonal. A matrix with all zeros except on the main diagonal is called a diagonal matrix. A diagonal matrix in which $a_{ii} = 1$ $\forall i$ is called an identity matrix and is denoted by I_n.*

We will now introduce the concept of matrix multiplication. Given A (m×n) matrix and B (n×q) we define the product between A and B the matrix C=AB with generic element given by $c_{jk} = \sum_{i=1}^{n} a_{ji}b_{ik} = a_{j1}b_{1j} + a_{j2}b_{2j} + ... + a_{jn}b_{nj}$. Notice that AB$\neq$ BA (Actually if AB is well defined BA may fail to be). Notice also that if A is ($m \times n$) then $AI_n = A$. From now on when we write AB it is assumed that the product exist. We list below two properties of the matrix product that are straightforward (but boring) to prove. Let A, B, C be matrices and k a scalar then

$$A(B + C) = AB + AC$$

$$A(kB) = k(AB)$$

$$A(BC) = (AB)C$$

Definition 12.3. *A square matrix A of dimension n is said to be invertible (or non singular) if $\exists$ a matrix B of dimension n s.t.*

$$AB = BA = I_n$$

B is called the inverse of A and is denoted by A^{-1}.

Exercise 12.2. *Show that if A is invertible then its inverse is unique.*

Exercise 12.3. *Find the inverse of the matrix:* $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.

Later, when we will introduce the concept of determinant we will present a general method for finding the inverse of a matrix.

Exercise 12.4. *Finally we give some properties of the operation of matrix transposition and inversion.*

1. $(AB)' = B'A'$ $(*)$, $(A + B)' = A' + B'$, $(A')' = A$, $(kA)' = kA'$
2. $(ABC)' = C'B'A'$ (*)

3. If A and B are invertible the (AB) is invertible and $(AB)^{-1} = B^{-1}A^{-1}$ (*)

4. $(A^{-1})' = (A')^{-1}$

Exercise 12.5. *Prove the starred properties above.*

Exercise 12.6. *We define the trace of a square matrix ($tr(A)$) as the sum of the elements on its main diagonal. Given two n dimensional matrices A, B prove that $tr(AB) = tr(BA)$.*

Exercise 12.7. *Find all 2×2 matrices satisfying $AA = O$. (nilpotent matrices)*

Exercise 12.8. *Find a 2×2 matrix satisfying $AA = A$. (idempotent matrix).*

13. Linear mappings

In this section we present the class of functions that are the object of study of linear algebra.

Definition 13.1. *Let V and W be vector spaces over the field K. A linear mapping is a map F:V→W that satisfy the following two properties*

1. $\forall u, v \in V \ F(u + v) = F(u) + F(v)$

2. $\forall u \in V$ *and* $c \in K \ F(cu) = cF(u)$.

Example of linear mapping are easy to find. For example the map P:$\Re^n \rightarrow \Re^m$ ($n > m$) is called a projection if P$(x_1, x_2, ... \ x_n) = (x_1 .... x_m)$; it's easy to show that a projection is linear.

Exercise 13.1. *Consider the mapping P:$\Re^n \rightarrow \Re^m$ defined as $P(x) = Ax$ where A is (m×n) matrix and x is a column vector in $\Re^n$. Show that it is linear.*

We will now prove a very useful property of linear mappings:

Theorem 13.2. *Let V and W be vector spaces and let O_W and O_V be the respective zero elements. Let F: V → W be a linear map. Then $F(O_V) = F(O_W)$.* ***Proof.*** *From linearity we have $F(v+ O_V) = F(v\ ) + F(O_V)\ \forall v \in V$. On the other hand $v+ O_V = v$ and therefore $F(v+ O_V) = F(v\ )$. We conclude that $F(v\ ) + F(O_V) = F(v\ )$ and therefore $F(O_V) = O_W$.* ■

We will now describe some important characteristics a linear mapping (F:$V \rightarrow W$) might have. First of all we define the image of F ($Im(F)$ as the set $\{F(v) : v \in V\}$. Then we say that the mapping F is :

- Injective(or 1 to 1) , if $\forall u, v \in V \ u \neq v \rightarrow$ F(u) $\neq$ F(v).
- Surjective (or onto) if $Im(F) = W$.
- Bijective if it is injective and surjective.

Definition 13.3. *The Kernel (or Null-space) of F ($Ker(F)$) is the set* $\{v : F(v) = 0\}$.

It is straightforward to prove that the Kernel of a linear map is a subspace.

Exercise 13.2. *Consider the map* $F : \Re \rightarrow \Re = x^2 - 1$. *Is the Kernel of this map a subspace ?*

Theorem 13.4. $Ker(F) = \{O\} \iff F$ *is injective****Proof.*** *Suppose* $Ker(F) = \{O\}$ *and suppose* v *and* w *are such* $F(v) = F(w)$. *We have that* $F(v) - F(w) = O$ *and by linearity* $F(v - w) = O = F($ $O)$. *Therefore* $v - w = O$ *and* $v = w$. *Suppose now* F *is injective therefore* $\exists! v$: $F($ $v)$ $=O$ *but by linearity* $v = O$, *therefore* $Ker(F) = O$.

Theorem 13.5. *Let F:*$V \rightarrow W$ *be an injective linear map. Then if* ($v_1, v_2, ..., v_n$) *are linearly independent in* V, ($F(v_1), F(v_2), ..., F(v_n)$) *are linearly independent in W.*

Theorem 13.6. *Let F:*$V \rightarrow W$ *be a linear map.* $Im(F)$ *is subspace of W.* $Im(F)$ *is also called the rank of F.*

Exercise 13.3. *Prove the previous 2 theorems.*

We are now ready to prove a very important theorem

Theorem 13.7. *Let F:*$V \rightarrow W$ *be a linear map. Then*

$$\dim V = \dim Ker(F) + \dim Im(F)$$

Proof. *Let's call* $s > 0$ *the dimension of* $Im(F)$, $q \geq 0$ *the dimension of* $Ker(F)$ *and* n *the dimension of V. Let* $(w_1, w_2, ..w_s)$ *be a basis of* $Im(F)$. *Then there* *will be elements* v_i *s.t.* $w_i = F(v_i)$. *Then let* $(u_1, u_{2,}...u_q)$ *be a basis for* $Ker(F)$. *We want to show that* $(v_1....\ v_s,\ u_1, ..., u_q)$ *is a basis for* V. *To this end we first have to show that they generate* V *and then that they are linearly independent. Consider* $v \in V$. *Since the* $w's$ *are a basis for* $Im(F)$ *there exist* s *numbers s.t* $F(v) = x_1w_1 + ... + x_sw_s = x_1F(v_1) + ...x_sF(v_s)$ *and by linearity we have:*

$$F(v - x_1v_1 - ..... - x_sv_s) = O$$

but this means $v - x_1v_1 - ..... - x_sv_s$ *lies in* $Ker(F)$ *and that we have* q *numbers satisfying*

$$v - x_1v_1 - ..... - x_sv_s = y_1u_1 + .. + y_su_q$$

implying

$$v = x_1v_1 + ... + x_sv_s + y_1u_1 + .. + y_su_q$$

and therefore $(v_1....\ v_s,\ u_1, ..., u_q)$ *generate* V. *To show that* $(v_1....\ v_s,\ u_1, ..., u_q)$ *are linearly independent consider the linear combination*

$$x_1v_1 + ... + x_sv_s + y_1u_1 + .. + y_su_q = O$$

Applying F *to both sides and using the facts* $F(u_i) = O$ *and* $F(v_i) = w_i$ *we have*

$$x_1w_1 + ... + x_sw_s = O$$

that implies $x_i = 0$ *since the* w *are a basis and therefore linearly independent. This in turn implies that*

$$y_1u_1 + .. + y_su_q = O$$

and since the $u's$ *are a basis we have* $y_i = 0$. *Hence* $(v_1....\ v_s,\ u_1, ..., u_q)$ *are linearly independent.* ■

Corollary 13.8. *Let* $F{:}V \to W$ *be a linear map and assume* $\dim(V) = \dim(W)$. *If* $Ker(F) = O$ *or if* $Im(F) = W$ *then* L *is bijective.*

13.1. Composition and inverse of linear maps

Let U, V, W be vector spaces and let $F : U \to V$ and $G : V \to W$ be mappings, we define the composite mapping $G \circ F : U \to W$, to be the mapping $G(F(t))$. We then define the identity map as the map $I_W : W \to W = w \forall w \in W$,

and finally we can define the inverse map:

Let $F : U \to V$. We say that F has an inverse if there exist a mapping $G : V \to U$ such that

$$F \circ G = I_U \text{ and } G \circ F = I_V$$

Theorem 13.9. *Let $F : U \to V$. F has an inverse if and only if it is bijective.*

The previous concepts apply to general mappings while now we will state two theorems that are relative to linear mappings.

Theorem 13.10. *Let U, V, W be vector spaces and let $F : U \to V$ and $G : V \to W$ be linear mappings, then the composite mapping $G \circ F : U \to W$ is linear.*

Theorem 13.11. *Let $F : U \to V$ be a linear mapping. If F has an inverse $G : V \to U$ then it is linear.*

Exercise 13.4. *Proof the previous theorem.*

Theorem 13.12. *A linear map that is surjective and has Kernel equal to $\{O\}$ has an inverse.*

A linear map that has an inverse is also called an *isomorphism*. If we can find an isomorphism between two vector spaces V and W the two spaces are called *isomorphic*.

Exercise 13.5. *Let U, V, W be vector spaces and let $F : U \to V$ and $G : V \to W$ be isomorphisms. Show that $G \circ F : U \to W$ an isomorphism and that $(G \circ F)^{-1} = F^{-1}G^{-1}$.*

14. Bases Matrices and Linear Maps

In an exercise in the previous section we have seen how a (m×n) matrix generates a linear map $F : \Re^n \to \Re^m$. It is also easy to prove the converse with the following theorem:

Theorem 14.1. *Let $F : \Re^n \to \Re^m$ be a linear map. Then $\exists! A \in M_{m \times n}$ s.t. $F(x) = Ax$, $x \in \Re^n$.**__Proof.__** Let $\{E_1 ....E_n\}$ be the canonical (column) basis in $\Re^n$ and $\{e_1 ....e_n\}$ be the canonical (column) basis in $\Re^m$. Any (column) vector $x \in \Re^n$ can be written as:*

$$x = x_1E_1 + ... + x_nE_n \qquad x_i \in \Re$$

using linearity we have that

$$F(x) = x_1 F(E_1) + ... + x_n F(E_n)$$

Since $L(E_i) \in \Re^m$ we can write it as :

$$\begin{aligned} F(E_1) &= a_{11}e_1 + ... + a_{m1}e_m \\ &.... \\ F(E_n) &= a_{1n}e_1 + ... + a_{mn}e_m \end{aligned}$$

and therefore

$$\begin{aligned} F(x) &= x_1(a_{11}e_1 + ... + a_{m1}e_m) + .. + x_n(a_{n1}e_1 + ... + a_{nm}e_m) \\ &= (a_{11}x_1 + ... + a_{1n}x_n)e_1 + ... + x_n(a_{m1}x_1 + ... + a_{mn}x_n)e_m \end{aligned}$$

or

$$F(x) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = Ax$$

To show that the matrix is unique assume there is another matrix B such that $F(x) = Bx$. Since it must be $B_{i.}x = A_{i.}x \ \forall x$ it follows also that $B_{i.} = A_{i.} \ \forall i$ and therefore $B = A$. ■

Now that we have established a one to one relation between matrices and linear mappings we can interpret the theorem relating the dimension of the domain of a function with the dimension of its kernel and its image. We can infact write a linear map as

$$f : V \to W = Ax = A_1 x_1 + ... + A_n x_n$$

where $A_1, ...A_n$ are the column vectors of A. So the image of A (or its rank) coincides with the space spanned by its column vectors. So if the image has dimension n it must be that the matrix has n linearly independent vectors. This conclusion is important for square matrices that are surjective linear mappings. We have seen that surjective linear mappings are invertible if and only if the dimension of their Kernel is 0. This implies that the linear mapping associated to a square matrix of order n is invertible if and only if the dimension of its image is n but this in turn is true if and only if the column vectors of A are linearly independent. We can therefore state the following theorem:

Theorem 14.2. *A linear mapping associated with a square matrix is invertible if and only if its columns are linearly independent.*

Exercise 14.1. *Find the matrix associated with the identity map from $\Re^3 \to \Re^3$ and with the projection mapping from $\Re^3$ to $\Re^2$.*

Exercise 14.2. *Consider $F : \Re^n \to \Re^m$. Show that a necessary and sufficient condition for F to be invertible is $n = m$ and that the columns of the matrix associated to the mapping are linearly independent.*

In the previous theorem we have seen how the matrix associated with a linear mapping depends on the basis chosen for the spaces mapped from the function. We will now consider a more general statement that will consider arbitrary bases. First is important to observe that a more general (n dimensional) vector space is isomorphic to $\Re^n$; to see this consider V a vector space and $(v_1 .... v_n)$ a basis for that space; then $\forall v \in V$ we have $v = x_1 v_1 + ... + x_n v_v$ with x_i numbers. Consider now the linear mapping $\Re^n \to V$ given by $(x_1, ...x_n) \to x_1 v_1 + ... + x_n v_v$ that is the map that associates to every vector its coordinates respect to a given system of bases.

Exercise 14.3. *Show that this last map is linear and it is an isomorphism.*

Let's denote with $X_\beta(v)$ the vector of coordinates of v with respect to the basis β. Then using the previous theorem and the previous observation we can prove the following:

Theorem 14.3. *Let V and W be vector spaces with basis β and β' respectively, and let $F : V \to W$ be a linear map, then there is a unique matrix, denoted by $M_\beta^{\beta'}(F)$ such that:*

$$X_{\beta'}(F(v)) = M_\beta^{\beta'}(F) X_\beta(v)$$

Corollary 14.4. *Let V a vector space and let β and β' bases for V and Id denote the identity mapping, then*

$$X_{\beta'}(v) = M_\beta^{\beta'}(Id) X_\beta(v).$$

this corollary tell us how we can change the bases in a vector spaces premultiplying vectors by a matrix. We will now show that this matrix is invertible. With simple matrix algebra we can proof the following:

Theorem 14.5. *Let V,W,U be vectors spaces and β, β', β'' be their basis. Let $F : V \to W$ and $G : W \to U$ be linear maps then we have*

$$M^{\beta'}_{\beta''}(G)M^{\beta}_{\beta'}(F) = M^{\beta}_{\beta''}(G \circ F)$$

Consider now V=W=U and $F = G = Id$ and $\beta'' = \beta$ then we have

$$M^{\beta'}_{\beta}(Id)M^{\beta}_{\beta'}(Id) = M^{\beta}_{\beta}(Id)$$

but since by definition that $M^{\beta}_{\beta}(Id) = I$ we have that

$$M^{\beta'}_{\beta}(Id)M^{\beta}_{\beta'}(Id) = I = M^{\beta}_{\beta'}(Id)M^{\beta'}_{\beta}(Id)$$

so the matrices that change coordinates between β and β' are invertible. We'll finally get to the concept of diagonalization of a linear map. Using the previous theorem we can prove that if $F : V \to V$ is a linear map and β , β' are bases for V then:

$$\begin{aligned} M^{\beta'}_{\beta'}(F) &= N^{-1}M^{\beta}_{\beta}(F)N \\ with \qquad N &= M^{\beta'}_{\beta}(Id) \end{aligned}$$

Exercise 14.4. *Prove the previous equality*

So if $F : V \to V$ is a linear map we say that a basis β diagonalize F if $M^{\beta}_{\beta}(F)$ is a diagonal matrix. F is diagonalizable if such a base exists. Finally using the previous equality we can easily prove that $F : V \to V$ is diagonalizable if and only if there exists an invertible matrix N such that $N^{-1}M^{\beta}_{\beta}(F)N = M'$ is a diagonal matrix. $M^{\beta}_{\beta}(F)$ and M' are also called similar matrices reflecting the fact that they are associated to the same linear map but under different bases. Notice that unfortunately not all linear maps are diagonalizable.

15. Determinants and Inverse Matrices

In this section we will introduce the notion of determinant, a number associated with a square matrix, that can tell us whether the matrix (and therefore the function associated to it) is invertible and is useful in computing the inverse matrix (if it exists).We have seen that a square matrix is invertible iff its columns are linearly independent. We will show in class that linear independence of vectors has a geometrical interpretation: when vectors are linearly independent a particular space has non zero measure (By measure here we mean area in $\Re^2$ and volume in $\Re^3$). The determinant of a square matrix give us the measure of that area. In particular the determinant will be zero whenever this space will have measure 0 or when the columns(rows) of a square matrix are linearly dependent, and therefore the matrix is not invertible.

Definition 15.1. *Let A be a square matrix of order n. We call determinant of A , or $\det(A)$, or $|A|$ the number:*

$$\sum (-1)^j a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

where the summation is extended to all the possible $n!$ permutations $(j_1, j_2, \cdots j_n)$ of the first n integers and j represent the number of inversions of that permutation with respect to the fundamental one given by $(1, 2, ..., n)$.

Exercise 15.1. *Compute, using the definition, the determinant of the following two matrices:* $\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ 0 & \ddots & \vdots \\ 0 & 0 & a_{nn} \end{pmatrix}.$

Exercise 15.2. *Prove that if c is a number and A is a square matrix of order n then* $\det(cA) = c^n \det A$

There are many properties of the determinants and here we will mention a few that we will use in deriving methods for computing the determinant.

1. $$\det A = \sum (-1)^j a_{1j_1} a_{2j_2} \cdots a_{nj_n} = \sum (-1)^i a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \qquad with \quad i = j$$

2. $$\det A = \det A'$$

3. If in a matrix we move a row(column) of a matrix p positions to the top or to the bottom (left or right) its determinant is multiplied by $(-1)^p$.

4. If we swap two rows or two colums the determinant changes its sign. Notice infact that swapping two rows imply moving the lower p positions above and the upper $p-1$ positions below so the determinant by the previous property is multiplied by $(-1)^{2p-1} = -1 \quad \forall p$

5. If two rows (columns) are equal the determinant is 0. If we switch the two equal rows by property 4 the determinant has to change sign but since the matrix remains the same also its determinant has to be the same. Therefore it must be $\det A = 0$

6. If the ith row (column) of a matrix A is the sum of n vectors ($v_1, ..v_j, ..v_n$) then the determinant of A is given by the sum of determinants of the n matrices equal to A but with the v_j in the ith row (column). That is if
$$A = \begin{pmatrix} A_{1\cdot} \\ \vdots \\ B_{i\cdot} + C_{i\cdot} \\ \vdots \end{pmatrix} \text{ then } \det A = \det \begin{pmatrix} A_{1\cdot} \\ \vdots \\ B_{i\cdot} \\ \vdots \end{pmatrix} + \det \begin{pmatrix} A_{1\cdot} \\ \vdots \\ C_{i\cdot} \\ \vdots \end{pmatrix}.$$

7. (Binet's theorem)
$$\det(AB) = \det(A)\det(B)$$

Exercise 15.3. *Prove that if A is a square matrix that has inverse A^{-1} then* $\det(A^{-1}) = 1/\det(A)$.

we are now ready to prove the first and second Laplace theorems that are useful in computing determinants and inverse matrices.

Definition 15.2. *Given a square matrix A the algebraic complement (or cofactor) of its element a_{ij} (call it A_{ij}) is given by the determinant of the matrix obtained erasing the ith row and the jth column from A (call it M_{ij}) times (-1)$^{i+j}$. So* $A_{ij} = (-1)^{i+j} M_{ij}$.

Theorem 15.3. *If the ith row (column) of a square matrix A is equal to the vector $a_{ij} \cdot e_j$ (The jth canonical bases times a scalar a_{ij}) then $\det A = (-1)^{i+j} a_{ij} M_{ij}$.* ***Proof.*** *Consider a particular A (call it A_0) that has the first row is equal to $a_{ij} \cdot e_1$. In this case from the definition of determinant we have that:*

$$\det A_0 = \sum (-1)^j a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$
$$= a_{ij} \sum (-1)^j a_{2j_2} \cdots a_{nj_n} \qquad j_2, ...j_n \neq 1$$
$$= a_{ij} M_{11}$$

Consider now the matrix A with the i^{th} equal to the vector e_j. This matrix can be reduced to the following matrix $\begin{pmatrix} a_{ij} & 0 & \cdots & 0 \\ 0 & a_{hk} & & \\ \vdots & & & \\ 0 & & & \end{pmatrix}$ *by moving the i^{th}row of i-1 positions and the j^{th} column of j-1 position. Using therefore the property 3 of determinants above, the fact that $i + j - 2$ is of the same class (odd or even) as $i + j$, and how we compute the determinant of A_0 we have:*

$$\det A = (-1)^{i+j} a_{ij} M_{ij} = a_{ij} A_{ij}$$

■

Theorem 15.4. *(First Laplace theorem)The determinant of a square matrix A is computed by summing the product of the elements of any line (row or column) by their algebraic complement.**Proof.** Consider for example the j^{th} row $(a_{i1}, ..., a_{in})$. It can be written as $(a_{i1}e_1 + ...a_{n1}e_n)$. Using property 6 of the determinants above and the previous theorem we have* $\det(A) = \sum_{j=1}^{n} A_{ij}a_{ij}$ ■

This theorem provide us with a recursion for computing determinants of matrix of order n.

Theorem 15.5. *(Second Laplace theorem) The sum of the products of the element of any line of a square matrix with the algebraic complement of the elements of a parallel line is equal to 0.**Proof.** Consider the expression*

$$\sum_{j=1}^{n} a_{pj} A_{qj} \qquad p \neq q$$

this is the expression for the the determinant of the matrix with the p^{th} and the q^{th} rows that are equal. By property 5 this is 0. ■

The two theorems can be summarized in the following expression (for the rows):

$$\begin{aligned} \sum_{j=1}^{n} a_{pj} A_{qj} &= \det(A) \qquad if\ p = q \\ &= 0 \qquad if\ p \neq q \end{aligned}$$

and a similar one for the columns:

$$\sum_{j=1}^{n} a_{jp}A_{jq} = \det(A) \qquad if\ p = q$$
$$= 0 \qquad if\ p \neq q$$

The two theorems can be used to find the inverse of a matrix. Define the adjugate matrix of a square matrix A ($adg(A)$) to be the matrix in which each element is substituted by its algebraic complement, that is

$$adg(A) = \begin{pmatrix} A_{11} & & A_{1n} \\ \vdots & & \\ A_{n1} & \cdots & A_{nn} \end{pmatrix}$$

It is easy to prove using the Laplace theorems the following equality:

$$[Adg(A)]'\,A = A\,[Adg(A)]' = \det(A)I$$

Exercise 15.4. *Prove the previous equality.*

that in turn, if the determinant is different from zero can be used to find the following expression for the inverse of A.

$$A^{-1} = \frac{1}{\det(A)}\,[Adg(A)]'$$

We conclude this section stating (without proof, but it would be a good exercise to give it a try) a theorem that syntesize the relation between determinants and invertibility of a matrix:

Theorem 15.6. *Let A be a square matrix. Then the following conditions are equivalent*

- *A is invertible*
- *the rows of A are linearly independent*
- *the columns of A are linearly independent*
- $\det(A) \neq 0$.

16. Systems of linear equations

In this section we will apply some of the theory we have seen so far to the task of solving systems of linear equations.In particular we will deal with the problem of existence, uniqueness and determinacy of solutions.

Let define the following system of linear equations

$$a_{11}x_1 + ... + a_{1n}x_n = b_1$$
$$\cdots$$
$$a_{m1}x_1 + ... + a_{mn}x_n = b_m$$

that can be represented as

$$Ax = b$$

where A is a $(m \times n)$ matrix, $x \in \Re^n$, $b \in \Re^m$. The system

$$Ax = \mathrm{O}$$

is called the homogeneus system associated with $Ax = b$. Notice that the matrix A is associated with a linear mapping $f : \Re^n \to \Re^m$. The first question we ask is about existence of solutions. To this purpose we have the following theorem

Theorem 16.1. *(Rouché Capelli) The system $Ax = b$ has a solution $\forall b$ if and only if the rank of A is equal to the rank of $A|b$.* ***Proof.*** *If there is a solution to the system then b belongs to the space spanned by the column vectors of A therefore $A|b$ and A have the same number of linearly independent vectors and therefore the same rank. If A and $A|b$ have the same rank then they have the same number of linearly independent vectors so $A|b$ must constitute a set of linearly dependent vectors and b can be expressed as a linear combination of the column vectors of A.* ■

Corollary 16.2. *Consider a system $Ax = O$ with $A \in M(m, n)$ with $n > m$ (Homogeneous with more unknowns than equation) then the system has always a solution (excluding the trivial one).*

Theorem 16.3. *Consider the system $Ax = O$ with $A \in M(m, m)$ (Homogeneous with unknowns = equations) then if the rank of A is m then the only solution is $x = O$.*

We have therefore seen that establishing the rank of a matrix (i.e. the number of its linearly independent column vectors) is important in determining the existence of solutions. Now we'll see some result concering the uniqueness of the solution.

Theorem 16.4. *Let x_0 be a solution (if it exists) of the system $Ax = b$, $A \in M(m, n)$ then we have:*

- *x_1 is a solution if and only if $x_1 = x_0 + z$ where $z \in Ker(F_A)$.*
- *$Ax = b$ has an unique solution if and only if the Rank of A is equal to n.*

Corollary 16.5. *Consider the system $Ax = b$ with $A \in M(m, m)$ (square matrix) then if the rank of A is m then it has an unique solution.*

Determining the rank of a matrix is therefore important in estabilishing the existence of solutions to the system associated with that matrix. We therefore state a theorem that help us determining the rank of a matrix (this theorem is called by Strang the fundamental theorem of linear algebra)

Theorem 16.6. *Let $F : \Re^n \to \Re^m$ be a linear map and A be the matrix associated to it. A' defines therefore a mapping $F' : \Re^m \to \Re^n$. Let's also denote with r the rank of A ($Dim(Im(F))$) then we have the following result:*
$Rank(A') = Rank(A) = r$

The previous theorem states that the number of linearly independent row of a matrix is the same as the number of linearly independent columns. An obvious consequence of the theorem is that if $A \in M(m, n)$ then $Rank(A) \leq \min(m, n)$. As we will discuss in class this theorem has important consequences in relating uniqueness and existence of solutions.

Finally we'll briefly mention methods for finding solutions, provided that they exist.

In a system $Ax = b$ where A is a $(n \times n)$ invertible matrix, $x \in \Re^n$, $b \in \Re^n$ the solution can be find simply by inverting the matrix so $x = A^{-1}b$ and a particular case of this procedure is the famous Cramer's rule that states

$$x_s = \frac{\det(A_{b_s})}{\det(A)}$$

where A_{b_s} is the matrix A with the s^{th}column substituted by the vector b.

When we deal with non square systems the most efficient method for solving them is the so called triangularization that will be discussed in class with the help of examples.

17. Eigenvalues and eigenvectors

In this section we will deal with the space C^n over the field C. Addition and scalar multiplication are defined (*mutatis mutandis*) as in $\Re^n$. The notion of scalar product though is slightly different. If x and y are elements of C^n we define their scalar (or hermitian) product as:

$$xy = \sum_{i=1}^{n} x_i \overline{y_i}$$

where $\overline{y_i}$ is the conjugate of y_i.

Exercise 17.1. *Prove that if x and y are elements of C^n $xy = \overline{yx}$ and that $xx \geq 0$.*

The concept of orthogonality is analogous for vectors in C^n.

Consider now the following linear dynamical system:

$$x(t+1) = Ax(t)$$

where A is a square matrix of order n and $x(t) \in \Re^n$. Economists are often interested in solutions to this system that satisfy

$$x(t+1) = Ax(t) = \lambda x(t)$$

$\forall t$, $\lambda \in C$ (The set of complex numbers). These solutions are interesting for two reasons; firstly they satisfy $x(t)/x(t+1) = \lambda$ for every t and secondly because the expression for $x(t)$ can be readily obteined as $x(0)\lambda^t$. The existence of this kind of problems suggest the following

Definition 17.1. *If A is a square (real) matrix $\lambda \in C$, $x \in C^n$, $x \neq O$ satisfying*

$$Ax = \lambda x$$

we say that λ is an eigenvalue of A and x is an eigenvector of A associated with the eigenvalue λ.

Exercise 17.2. *Prove that if A is a square matrix with eigenvalue λ then the space of eigenvectors associated with λ plus the O vector constitute a vector space (This is also called an eigenspace).*

We will now present some results that will help us in finding eigenvalues and eigenvectors of a matrix.

Theorem 17.2. *Let A a square matrix and let λ be a scalar. λ is an eigevalue of A if and only if $(\lambda I - A)$ is not invertible ($\det(\lambda I - A) = 0$).* ***Proof.*** *Assume λ is an eigenvalue of A. Then there exists $x \in C^n, x \neq 0$ satisfying $\lambda x - Ax = 0$ or $(\lambda I - A)x = 0$ so the linear mapping defined by $(\lambda I - A)$ has non empty Kernel and is therefore non invertible. Assume now $(\lambda I - A)$ is non invertible, then it has non empty Kernel and therefore there is an $x \neq 0$ satisfying $(\lambda I - A)x = 0 \longrightarrow \lambda x = Ax$ and that therefore λ is an eigenvalue of A.* ■

The previous theorem gives us an easy way of finding the eigenvalues of a square matrix, that is to find the zeros of the equation in $\lambda \det(\lambda I - A) = 0$. This equation is called the characteristic equation of A. Remembering the definition of determinant it's easy to see that the equation is a polynomial of degree n (The only term that contains λ^n is the one relative to the fundamental permutation), and that by the fundamental theorem of the algebra it has n (real or complex) roots.

Exercise 17.3. *Show that* 0 *is an eigenvalue of* A *if and only if* $\det(A) = 0$.

Exercise 17.4. *Find the matrix that has the following characteristic equation:*

$$\lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n = 0$$

Exercise 17.5. *Find eigenvalues and eigenvectors of* $\begin{pmatrix} 0 & 2 \\ 1 & 4 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$.

Eigenvalues and eigenvectors can be efficiently used to perform diagonalization (when possible) or triangularization of matrices. Let A be a square matrix of order n and let $\lambda_1, \lambda_2, ...\lambda_n$ be its eigenvalues. Let also $x_1, ...x_n$ be n correspondent eigenvectors so that

$$Ax_i = \lambda_i x_i \qquad i = 1, ..n$$

these relations can be written in matrix form as

$$AX = X\Lambda$$

where $\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix}$ and $X = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}$. We defined before the concept of similarity for matrices so now we can state the following theorem:

Theorem 17.3. *A square matrix A is similar to a diagonal matrix Λ , that is*

$$A = S^{-1}\Lambda S$$

with S of the same order of A if and only if A has n linearly independent eigenvectors.

Exercise 17.6. *Show that the following matrix:* $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ *is not diagonalizable.*

As we have seen not every matrix is diagonalizable. A sufficient condition for diagonalization is provided by the following theorem:

Theorem 17.4. *Eigenvectors associated to distinct eigenvalues are linearly independent.*

Finally we mention that an important result in linear algebra guarantees for every square matrix the existence of a triangular matrix similar to it with the eigenvalues on the main diagonal.

18. Diagonalization of symmetric matrices and quadratic forms

A function $Q : \Re^n \to \Re = (a_1x_1 + a_2x_2 + ...a_nx_n)^2$ is called a quadratic form. Every quadratic form can be expressed as

$$Q(x) = x'Ax$$

where A is a symmetric matrix. Notice that we can always find a matrix A that is symmetric. Suppose infact A is not symmetric then we have

$$\begin{aligned} Q(x) &= x'Ax = \frac{1}{2}x'Ax + \frac{1}{2}x'Ax = \frac{1}{2}x'Ax + \frac{1}{2}\left(x'Ax\right)' = \\ \frac{1}{2}x'Ax + \frac{1}{2}x'A'x &= x'(\frac{A+A'}{2})x \end{aligned}$$

and $\frac{A+A'}{2}$ is symmetric.As an example suppose $x \in \Re^2$ consider the quadratic form $Q(x) = x_1^2 + 10x_1x_2 + x_2^2$. It can be written as

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & 5 \\ 5 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Definition 18.1. *A quadratic form $Q(x)$ is definite positive(negative) if $\forall x \neq O$ we have $Q(x) > 0$ ($Q(x) < 0$). It is semidefinite positive (semidefinite negative) if $\forall x$ we have $Q(x) \geq 0$ ($Q(x) \leq 0$) with equality with some $x \neq O$.*

The nature of a quadratic form is linked to the sign of its eigenvalues and therefore we provide the following theorem for the eigenvalues of symmetric matrices.

Theorem 18.2. *The eigenvalues of a square symmetric matrix are all real.*

Theorem 18.3. *The eigenvectors associated to distinct eigenvalues for a symmetric matrix are linearly independent and orthogonal.*

Before we introduce the next result we mention that n vectors ($v_1, v_2, ...v_n$) are said orthonormal if we have $v_i v_j = 0$ $i \neq j$ and$\|v_i\| = 1$ $\forall i$. We then define an orthonormal (square) matrix a matrix that has orthonormal columns.Let F be an orthonormal matrix then we have $FF' = I$ and consequantly $F' = F^{-1}$.

Theorem 18.4. *(Spectral theorem) If A is a square symmetric matrix then we can always find an orthonormal matrix F (that has n orthonormal eigenvectors of A as columns) such that*

$$A = F\Lambda F'$$

where Λ is the diagonal matrix formed with the eigenvalues of A.

From the previous theorem is straightforward to prove the following:

Theorem 18.5. *Necessary and sufficient condition for a quadratic form $Q(x) = x'Ax$ to be:*

- *Positive definite is to have all positive eigenvalues*
- *Positive semidefinite is to have all non negative eigenvalues with at least one equal to 0*

- *Negative definite is to have all negative eigenvalues*
- *Negative semidefinite is to have all non positive eigenvalues with at least one equal to 0.*

Part III
Multivariate calculus

In the first part of 897 you have seen functions defined from $\Re$ to $\Re$. Unfortunately in economics most of the times they are not enough. Consider for example a technology that relates hours worked and capital employed to output produced, or suppose that an individual gets utility from a basket of different goods. In these cases we need to consider functions defined on more complex domains. In particular in this section we will study the class of functions defined from $\Re^n$ to $\Re^m$. Passing from $\Re$ to $\Re^n$ is not completely straightforward since $\Re$ is a completely ordered set (that is given two points in $\Re$ either one is bigger or of the other or they are equal) while $\Re^n$ is not (Consider the vectors in $\Re^2$ (0,1) and (1,0)). When we move from one point to the other in $\Re$ there is only one direction we can go, in $\Re^n$ there are infinitely many.

19. Basic topology in $\Re^n$

We have already described $\Re^n$ and some of its properties in the previous section. In particular remember the concept of norm that is the function that is defined as $\|x\| \equiv \sqrt{xx} .\equiv \sqrt{\sum_{i=1}^n x_i^2}$. With the help of the norm we define the concept of distance between two vectors in $\Re^n$.

Definition 19.1. *Given x and y elements of $\Re^n$ their distance is defined as the function $d : \Re^n \times \Re^n \to \Re$ s.t.*

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

d is also called the Euclidean distance.

In the rest of this part the notation $d(x, y)$ or $\|x - y\|$ to denote the distance between two points will be used interchangeably.

Exercise 19.1. *Check that the Euclidean distance has the following properties.*

- $d(x, x) = 0$
- $d(x, y) > 0 \; if \; x \neq y.$
- $d(x, y) = d(y, x)$
- $d(x, y) \leq d(x, z) + d(z, y)$

where x, y, z are elements of $\Re^n$.

Remark 2. *Every function that satisfy the properties above is said to be a distance.*

Remark 3. *Any set that is equipped with a distance is called a metric space. For example Philadelphia, New York and San Francisco together with the road distance constitute a metric space.*

Definition 19.2. *Let $x \in \Re^n$ and let $\epsilon > 0$ be a number then:*

- *the set $N_\epsilon(x) = \{y \in \Re^n : d(x, y) < \epsilon\}$ is called an open neighborhood (open ball)*
- *the set $\overline{N}_\epsilon(x) = \{y \in \Re^n : d(x, y) \leq \epsilon\}$ is called a closed neighborhood (closed ball)*

Definition 19.3. *Let S be a subset of $\Re^n$ and assume $x \in S$. We say that x is an interior point of S if $\exists \epsilon > 0$ s.t. $N_\epsilon(x) \subseteq S$.*

Definition 19.4. *A set $S \subseteq \Re^n$ is open if every point in the set is interior to S.*

Definition 19.5. *A set $S \subseteq \Re^n$ is closed if its complement $\Re^n/S$ is open*

Definition 19.6. *Let S be a subset of $\Re^n$ and assume $x \in \Re^n$. We say that x is an adherent point of S if $\forall \epsilon > 0$ $N_\epsilon(x)$ contains at least one point of S.*

Definition 19.7. *Let S be a subset of $\Re^n$ and assume $x \in \Re^n$. We say that x is an accumulation point of S if $\forall \epsilon > 0$ $N_\epsilon(x)$ contains at least one point of S different from x.*

Notice that accumulation and adherent points for a set do not need to belong to that set. Clearly all points that belong to a set are adherent points but this is not true for accumulation points. If a point belong to a set but it's not its accumulation point is called an *isolated point*.

Theorem 19.8. *If x is an accumulation point of S then every open ball centered in x contains infinitely many points of S.*

We finally characterize close sets in a different way

Theorem 19.9. *A set $S \subseteq \Re^n$ is closed if and only if it contains the set of all its adherent points (Also called the closure of S and denoted by $\overline{S}$.)*

Theorem 19.10. *A set $S \subseteq \Re^n$ is closed if and only if it contains the set of all its accumulation points (Also called the derived set of S and denoted by S')*

Notice finally that a set can be nor close neither open (Consider the set $[0,1)$ in $\Re$) and can be both closed and open (Consider $\Re^n$).

Exercise 19.2. *Take A, B open subsets of $\Re^n$, show that $A \cup B$ and $A \cap B$ are open too.*

Definition 19.11. *A set $S \subseteq \Re^n$ is bounded if $\exists \epsilon > 0$ s.t $S \subset N_\epsilon(x)$ for some $x \in \Re^n$.*

Theorem 19.12. *(Bolzano-Weierstrass) If S is a bounded set in $\Re^n$ that contains infinitely many points then there is at least a point in $\Re^n$ that is an accumulation point for S.*

Definition 19.13. *A collection of sets F is said to be a cover of a set S if $S \subseteq \cup_{A \in F} A$. The cover is open if every set in the collection is open.*

Definition 19.14. *A set $S \subseteq \Re^n$ is compact if and only if every open covering contains a finite subcover, that is a finite subcollection that also cover S.*

Exercise 19.3. *Show, using the definition, that the set $(0,1]$ in $\Re$ is not compact.*

Theorem 19.15. *A set $S \subseteq \Re^n$ is compact if and only if it is closed and bounded.*

Theorem 19.16. *A set $S \subseteq \Re^n$ is compact if and only if every infinite subset of S has an accumulation point in S.*

Theorem 19.17. *Let T be a closed subset of a compact metric space. Then T is compact*

20. Convergent sequences in $\Re^n$

Definition 20.1. *A sequence $\{x_n\}$ of points in $\Re^n$ is said to converge to $p \in \Re^n$ ($x_n \to p$) if*

$$\forall \epsilon > 0 \qquad \exists N \text{ s.t. } d(x_n, p) < \epsilon \quad \forall n > N$$

Remark 4. *The previous definition imply $x_n \to p$ in $\Re^n$ if and only if $d(x_n, p) \to 0$ in $\Re$.*

Exercise 20.1. *Prove that if $\{x_n\}$ is a convergent sequence in $\Re^n$ then its limit is unique.*

Definition 20.2. *A subsequence of $\{x_n\}$ is a sequence $\{s_{k(n)}\}$ whose n^{th} term is $x_{k(n)}$ where $k : N \to N$ satisfies $k(m) > k(n)$ if $m > n$.*

Theorem 20.3. *A sequence converges to p if and only if every subsequence converges to p.*

Notice that a sequence $\{x_n\}$ converges to p if and only if the univariate sequences $\{x_{in}\} \to p_i$ $i = 1, ...n$.

Definition 20.4. *A sequence $\{x_n\}$ of points in $\Re^n$ is called a Cauchy sequence if*

$$\forall \epsilon > 0 \qquad \exists N \text{ s.t. } d(x_n, x_m) < \epsilon \quad \forall n, m > N$$

Exercise 20.2. *Show that every convergent sequence is a Cauchy sequence.*

Exercise 20.3. *Find an example of a subset in $\Re^n$ in which a Cauchy sequence is not a convergent sequence.*

Definition 20.5. *A metric space in which every Cauchy sequence converges is called a complete metric space.*

Theorem 20.6. *Let $S \subseteq \Re^n$ and let p be an accumulation point of S. Then there is a sequence $\{x_n\} \in S$ and s.t. $\{x_n\} \to p$.* ***Proof.*** *Since p is an accumulation point for S for every integer n there is a point $x_n \in S$ s.t. $d(x_n, p) < \frac{1}{n}$. Letting n go to infinity we have that $d(x_n, p) \to 0$ and using the remark above $x_n \to p$* ■

Theorem 20.7. *Let $S \subseteq \Re^n$ then S is compact if and only if every sequence $\{x_n\} \subset S$ has a convergent subsequence.*

21. Multivariate functions: Limits and continuity

In this section we will focus on functions from $\Re^n$ to $\Re$.

Definition 21.1. *Let $S \subseteq \Re^n$ and $T \subseteq \Re$ we define a function $f : S \to T$ a mapping that associates a number $f(x) \in T \subseteq \Re$ to every vector $x \in S \subseteq \Re^n$.*

Definition 21.2. *Let $S \subseteq \Re^n$ and $T \subseteq \Re$, let $f : S \to T$, let p be an accumulation point of S ,and let b be point in $\Re$. If $\forall \epsilon > 0$ $\exists \delta > 0$ such that $d(x, p) < \epsilon$, $x \in S, x \neq p$ implies $|f(x) - b| < \delta$ then we write*

$$\lim_{x \to p} f(x) = b$$

An alternative definition is the following.

Definition 21.3. *If $\forall N_\epsilon(b)$ $\exists N_\delta(p)$ such that*

$$x \in N_\delta(p) \cap S \Longrightarrow f(x) \in N_\epsilon(p), \quad x \neq p$$

then we say

$$\lim_{x \to p} f(x) = b$$

Notice that in the definition of limit we do not require neither p to be in the range of f nor b to be to be in its image.

Theorem 21.4. *Assume that p is an accumulation point of S and assume $b \in T$, then*

$$\lim_{x \to p} f(x) = b$$

if and only if for every sequence $\{x_n\} \to p$, $\{x_n\} \in S/\{p\}$

$$\lim_{n \to \infty} f(x_n) = b$$

The usual properties of limits you've already seen in $\Re$ apply readily for limits of functions from $\Re^n$ to $\Re$. In particular if $\lim_{x \to p} f(x) = a$ and $\lim_{x \to p} f(x) = b$ then

- $\lim_{x \to p}[f(x) + g(x)] = a + b$

- $\lim_{x \to p} kf(x) = ka \; \forall k \in \Re$
- $\lim_{x \to p} f(x)g(x) = ab$
- $\lim_{x \to p} \frac{f(x)}{g(x)} = \frac{a}{b}$ if $b \neq 0$

Exercise 21.1. *Prove that if* $\lim_{x \to p} f(x)$ *exists then it is unique.*

We are now ready to introduce the concept of continuity of a function f:

Definition 21.5. *Let* $S \subseteq \Re^n$ *and* $T \subseteq \Re$, *let* $f : S \to T$. f *is said to be continuous at* $p \in S$ *if* $\forall \; \epsilon > 0$

$$\exists \delta > 0 : |x - p| < \delta \Longrightarrow d(f(x), f(p)) < \epsilon$$

if f *is continuous at every point in* S *then is said to be continuous on* S.

Notice that if p is an accumulation point in S then we f is continuous if and only if

$$\lim_{x \to p} f(x) = f(p)$$

Remember also that if the function is defined on an isolated point is always continuous at that point. As it was in $\Re$ we have that the composite of a continuous functions is continuous as you are asked to show in the following exercise:

Exercise 21.2. *Let* $S \subseteq \Re^n$ *and* $T, U \subseteq \Re$ *and let* $f : S \to T$ *and* $g : f(S) \to U$. *Show that if* f *is continuous at* p *and* g *is continuous at* $f(p)$ *then* $g \circ f$ *is continuous at* p.

Checking the continuity of a function in $\Re^n$ is not an easy task as it was in $\Re$. The reason for that, as we already mentioned, is that in $\Re^n$ we can approach a point along infinitely many directions and we have to make sure that whichever direction we take the continuity property is preserved. As an example of that consider the following function:

$$f(x, y) = \begin{matrix} \frac{x^2}{y} & y \neq 0 \\ 0 & y = 0 \end{matrix}$$

If we check the continuity of $f(x, y)$ at the origin and we restrict ourselves to the directions given by $y = mx$ ($m \neq 0$) we have that $\lim_{x \to 0} f(x, mx) = 0$, but along the restriction $y = x^2$ we have $\lim_{x \to 0} f(x, x^2) = 1$ and therefore f is not continuous in the origin.

In class we will examine some methods including the use of polar coordinates that can help us in this task.

Theorem 21.6. *Let $K \subset \Re^m$ be a compact set and $f\colon K \to \Re$ be a continuous function. Then $f(K)$ is also compact*

22. Multivariable differential calculus

22.1. Partial derivatives

The partial derivative of a function of n variables gives us information on the behavior of the function when only one of the variables changes. More formally if $S \subseteq \Re^n$ is an open set, x_0 an element of S and $f : S \to \Re$ we define the partial derivative of f at x_0 with respect to the ith variable (denoted by$\frac{\partial f(x_0)}{\partial x_i}$ or $f_{x_i}(x_0)$) the following limit (if it exists)

$$\frac{\partial f(x_0)}{\partial x_i} = \lim_{h \to 0} \frac{f(x_0 + he_i) - f(x_0)}{h}$$

where e_i is the i^{th}canonical basis. We then say that f is *derivable* at x_0 if it has partial derivatives with respect to each variable. Then if f is derivable we can define its gradient that is the vector of its partial derivatives:

$$Df(x_0) = [f_{x_1}(x_0), \ldots . f_{x_n}(x_0)]$$

The calculation of partial derivatives does not give any problems since we can apply the same rules of univariate calculus. In particular when we derive f with respect to the ith variable we simply treat the other $n - 1$ variables as constant.

If consider a function $f(x, y) : \Re^2 \to \Re$ we can give a geometrical interpretation of the partial derivative. In particular the partial derivative with respect to x in the point x_0, y_0 is the slope of the intersection of the surface $z = f(x, y)$ with the vertical plane $y = y_0$.

Not surprisingly the derivability of a function does not imply continuity. Infact the existence of partial derivative imply continuity along certain directions but we have seen that this is not enough to guarantee the continuity of the function.

Partial derivatives of higher order are simply defined as the partial derivatives of the partial derivatives:

$$\frac{\partial f(x_0)}{\partial x_j \partial x_i} = f_{x_i x_j}(x_0) = \frac{\partial f_{x_i}(x_0)}{\partial x_j}$$

and if $f : S \subseteq \Re^n \rightarrow \Re$ has second order partial derivatives we can define its *Hessian Matrix* as the square matrix of order n that has on the i^{th} row and j^{th} column the partial derivative with respect to the j^{th} variable of partial derivative with respect to the i^{th} variable, that is

$$\begin{pmatrix} f_{x_1 x_1} & \cdots & f_{x_n x_1} \\ \vdots & & \vdots \\ f_{x_1 x_n} & \cdots & f_{x_n x_n} \end{pmatrix}$$

We will also state (without proving) the following:

Theorem 22.1. *(Schwarz's) Let $f : S \subseteq \Re^n \rightarrow \Re$ and x_0 element of S. If $f_{x_i x_j}$ and $f_{x_j x_i}$ exist in a neighborhood of x_0 and are continuous at x_0 then we have:*

$$f_{x_i x_j}(x_0) = f_{x_j x_i}(x_0)$$

Exercise 22.1. *Consider the function*

$$f(x,y) = \begin{matrix} \frac{x^3 y}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{matrix}$$

Show that $f_{xy} \neq f_{yx}$. Why Schwarz's theorem does not apply ?

22.2. Directional derivatives

Let $f : S \subseteq \Re^n \rightarrow \Re$ and x_0 interior point of S. Suppose we are interested in evaluating how f changes when we move away from the point x_0 toward the point x_0+d. Each point on the segment joining x_0 and x_0+d can be expressed as $x_0+\rho d$ (where $\rho \in \Re$) and therefore it make sense to define the directional derivative in x_0 along the direction d (denoted by the symbol $f'(\ x_0, d)$) as:

$$f'(x_0, d) = \lim_{\rho \rightarrow 0} \frac{f(x_0 + \rho d) - f(x_0)}{\rho}$$

Remark 5. *A direction is not univocously determined by a vector. In particular all the vectors of the form* $hd = h(d_1, ..., d_n)$ *with* $h \in \Re$ *represent the same direction. It's therefore common to choose the vector* d *s.t.* $\|d\| = 1$. *This implies that for functions* $f(x,y) : \Re^2 \to \Re$ *all the possible directions can be described by the vectors* $(\cos\theta, \sin\theta)$ *with* $\theta \in [0, 2\pi]$.

Remark 6. *If* $d = e_i$ *then the directional derivative coincides with the partial derivative with respect to the* i^{th} *variable.*

Remark 7. *The existence of directional derivatives along any direction do not imply continuity of the function (but this again should not be surprising) but imply derivability. Derivability (that is existence of all partial derivatives) do not imply the existence of directional derivatives as is shown by the following function*

$$f(x,y) = \begin{matrix} x+y & x = 0 \text{ or } y = 0 \\ 3 & \text{otherwise} \end{matrix}$$

but later we will give conditions under which the directional derivative can be expressed as a function of the partial derivative.

Exercise 22.2. *Let* $f : S \subseteq \Re^n \to \Re$ *be a linear function. Show that its directional derivative exists and does not depend on the point where it is calculated but only on the direction vector* d.

Exercise 22.3. *Compute the directional derivative of the function* $f(x,y) = x+y$ *in the origin along a direction (an angle)* θ. *Which is the value of* θ *that gives the higher value of the derivative ?*

22.3. Differentiability

The existence of partial and directional derivatives were not enough to guarantee continuity of a function. We will introduce the concept of differentiability.

Definition 22.2. *Let* $f : S \subseteq \Re^n \to \Re$ *and* x_0 *interior point of* S. f *is said to be differentiable at* x_0 *if it exists a linear function* $T : \Re^n \to \Re$ *such that we can write*

$$f(x_0 + h) = f(x_0) + T_{x_0}(h) + o(\|h\|)$$
$$As \quad h \to 0 \qquad h \in \Re^n$$

Exercise 22.4. *Show that if f is differentiable at x_0 then it is continuous at x_0.*

Theorem 22.3. *Let $f : S \subseteq \Re^n \rightarrow \Re$ and x_0 interior point of S. Let f be differentiable at x_0. Then it has directional derivatives $f'(x_0, d)$ for every $d \in \Re^n$ and we have*

$$T_{x_0}(d) = f^{'}(x_0, d)$$

The linear function $T_{x_0}()$ is called total derivative of f at x_0, while the expression $T_{x_0}(d)$ is also called differential of f. ***Proof.*** *In the definition of differential take $h = md$, $m \in \Re$ and divide by m so:*

$$\frac{f(x_0 + md) - f(x_0)}{m} = \frac{T_{x_0}(md) + o(\|md\|)}{m} = T_{x_0}(d) + \frac{\|d\|\, o(m)}{m}$$

then let m go to 0 and the result follows. ■

Theorem 22.4. *Let $f : S \subseteq \Re^n \rightarrow \Re$ and x_0 interior point of S. Let f be differentiable at x_0. Then the vector associated to the linear function $T_{x_0}(h)$ is the gradient of f in x_0. therefore we can write*

$$f(x_0 + h) = f(x_0) + Df(x_0)h + o(\|h\|)$$

Proof. *Let's first write h as $h_1e_1 + ... + h_ne_n$, then using the linearity of T we have*

$$T_{x_0}(h_1e_1 + ... + h_ne_n) = \sum_i h_i T_{x_0}(e_i) =$$

$$\sum_i h_i f'(x_0, e_i) = \sum_i h_i f_{x_i}(x_0) = Df(x_0)h$$

■

This last theorem gives us a first order approximation of a multivariate function using its partial derivatives. Notice also that a consequence of the theorem is that if the function is differentiable the directional derivative along a certain direction d can be easily computed as the scalar product of the gradient with the vector representing the direction.

Remark 8. *Continuity and derivability are necessary but not sufficient conditions for differentiability as you are asked to show in the next exercise.*

Exercise 22.5. *Show that the following function:*

$$f(x,y) = \begin{array}{cc} \frac{x^2y}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{array}$$

is continuous and derivable but not differentiable in the origin .Hint. to show that is not differentiable use the expression

$$f(x_0 + h) = f(x_0) + Df(x_0)h + o(\|h\|)$$

A sufficient condition for differentiability is stated in the following :

Theorem 22.5. *If f has continuous partial derivatives in a neighborhood of x_0 then it is differentiable in x_0.*

Remark 9. *If we consider $z = f(x,y)$ we can rewrite the expression of the differential in x_0, y_0 as*

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

and that can be interpreted as the equation of the plane tangent to the surface in x_0, y_0.

22.4. Taylor's formula

Before we introduce this formula we need to specify the higher order differentials. If f:$\Re^n \rightarrow \Re$ with partial derivatives of m^{th} order in x and $t \in \Re^n$ we write:

$$f''(x;t) = \sum_{i=1}^{n} \sum_{j=1}^{n} D_{i,j} f(x) t_i t_j$$

$$f'''(x;t) = \sum_{k=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} D_{k,i,j} f(x) t_i t_j t_k$$

and similar expressions for the m^{th} order.

these formulae can be interpreted as higher order directional derivatives or as differential of differentials.

Theorem 22.6. *Let f be a function differentiable m times on an open set S and let a , b together with all the points on the line segment $L(a, b)$ be elements of S. Then there is a point $z \in L(a, b)$ such that:*

$$f(b) = f(a) + \sum_{k=1}^{m-1} \frac{1}{k!} f^{(k)}(a, b-a) + \frac{1}{m!} f^{(m)}(z, b-a)$$

Remark 10. *An alternative way of writing Taylor formula is with the remainder according to Peano that is*

$$f(b) = f(a) + \sum_{k=1}^{m} \frac{1}{k!} f^{(k)}(a, b-a) + o(\|b-a\|^m)$$

23. Vector valued functions

In this section we will generalize the concept of a function from $\Re^n$ to $\Re$ to the one of function from $\Re^n$ to $\Re^m$.

Definition 23.1. *Let $S \subseteq \Re^n$ and $T \subseteq \Re^m$ we define a function $\underline{f}$: $S \to T$ a mapping that associates a vector $\underline{f}(x) = (y_1, ..y_m) \in T \subseteq \Re$ to every vector $x = (x_1, ...x_n) \in S \subseteq \Re^n$.*

A vector valued function from $\Re^n$ to $\Re^m$ can be thought as a set of m functions from $\Re^n$ to $\Re$ therefore sometimes we write $\underline{f}$: $\Re^n \to \Re^m = (f1, ..f2)$ where $fi : \Re^n \to \Re$, $i = 1, ...m$ are called the components of f. An example of vector valued function is given by the following

$$\underline{f} : \Re^2 \to \Re^2 = \begin{array}{c} f1(x_1, x_2) = x_1 + x_2 \\ f2(x_1, x_2) = x_1 x_2 \end{array}$$

Definition 23.2. *Let $S \subseteq \Re^n$ and $T \subseteq \Re^m$, let $f : S \to T$, let p be an accumulation point of S ,and let b be point in $\Re^m$. If $\forall \epsilon > 0$ $\exists \delta > 0$ such that $\|x - p\| < \epsilon$, $x \in S, x \neq p$ implies $\|f(x) - b\| < \delta$ then we write $\lim_{x \to p} f(x) = b$*

Definition 23.3. *Let $S \subseteq \Re^n$ and $T \subseteq \Re^m$. A function $\underline{f}$: $S \to T$ is continuous at $p \in S$ if $\forall \epsilon$ >0*

$$\exists \delta > 0 : \|x - p\| < \delta \Longrightarrow \|f(x) - f(p)\| < \epsilon$$

Continuity of a vector valued function can be established by establishing the continuity of its components as shown in the next theorem:

Theorem 23.4. *Let $S \subseteq \Re^n$ and $T \subseteq \Re^m$, let $\underline{f}$: $S \rightarrow T$, let p be a point in S . Then $\underline{f}$ is continuous in p if and only if all its components $(f1, ...fm)$ are continuous in p.* ***Proof.*** *If $\underline{f}$ is continuous at p then $\forall \epsilon > 0$*

$$\exists \delta > 0 : \|x - p\| < \delta \Rightarrow \|f(x) - f(p)\| < \epsilon \Rightarrow |fi(x) - fi(p)| < \epsilon \quad \forall i$$

thus implying the continuity of each component. On the other hand if each component is continuous then we have $\forall \epsilon/m > 0$

$$\exists \delta_i > 0 : \|x - p\| < \delta_i \Rightarrow |fi(x) - fi(p)| < \epsilon/m, \qquad i = 1, .., n$$

Take then $\delta_\epsilon = \min(\delta_1, ...\delta_m)$. We then have that $\forall \epsilon/m$

$$\exists \delta_\epsilon > 0 : \|x - p\| < \delta_\epsilon \Rightarrow |fi(x) - fi(p)| < \epsilon/m, \qquad i = 1, .., n$$

therefore $\forall \epsilon$

$$\exists \delta_\epsilon > 0 : \|x - p\| < \delta_\epsilon \Rightarrow \sum_{i=1}^{m} |fi(x) - fi(p)| < \epsilon \tag{23.1}$$

but notice now that

$$\sum_{i=1}^{m} |fi(x) - fi(p)| \geq \|f(x) - f(p)\|$$

(to see that just take the square of both sides) and therefore 23.1 implies that $\forall \epsilon$

$$\exists \delta_\epsilon > 0 : \|x - p\| < \delta_\epsilon \Rightarrow \|f(x) - f(p)\| < \epsilon$$

that is the continuity of $\underline{f}$. ∎

For vector valued functions the directional, partial derivatives and differentiability are defined as we did for real valued functions. In particular let $\underline{f}$: $S \subseteq \Re^n \rightarrow \Re^m$ and x_0 interior point of S. We define the partial derivative of $\underline{f}$ with respect to the i^{th} variable

$$\frac{\partial \underline{f}(x_0)}{\partial x_i} = \lim_{h \to 0} \frac{\underline{f}(x_0 + he_i) - \underline{f}(x_0)}{h}$$

and the directional derivative of$\underline{f}$ along a direction d

$$\underline{f}'(x_0, d) = \lim_{\rho \to 0} \frac{\underline{f}(x_0 + \rho d) - \underline{f}(x_0)}{\rho}$$

In particular if $\underline{f} = (f1, ..., fm)$ then $\underline{f}'(x_0, d)$ exists if and only if $fi'(x_0, d)$ exists for every i and we have that $\underline{f}'(x_0, d) = (f1'(x_0, d), ..., fm'(x_0, d))$ and if $d = e_k$ then $\underline{f}'(x_0, e_k) = \underline{f}_k(x_0) = (f1_k(x_0), ..., fm_k(x_0))$. In other words the partial derivative of a vector valued function with respect to the k^{th} variable is the vector of partial derivatives of the components of $\underline{f}$ with respect to the same variable. Analogously $\underline{f}$ is said to be differentiable at x_0 if it exists a linear function $T : \Re^n \to \Re^m$ such that we can write

$$\begin{aligned} \underline{f}(x_0 + h) &= \underline{f}(x_0) + T_{x_0}(h) + o(\|h\|) \\ As \quad h &\to 0 \qquad h \in \Re^n \qquad o(\|h\|) \in \Re^m \end{aligned}$$

or

$$\begin{aligned} \underline{f}(x_0 + h) &= \underline{f}(x_0) + T_{x_0}(h) + \|h\| \, E_{x_0}(h) \\ where \quad E_{x_0}(h) &\to 0 \quad as \quad h \to 0 \\ h &\in \Re^n \qquad E_{x_0}(h) \in \Re^m \end{aligned}$$

For functions from $\Re^n$ to $\Re^1$ we had that the linear function $T_{x_0}(x) = Df(x_0)x$ was associated to the gradient of the function computed in the point. For vector valued functions the linear function $T(x)$ is associated to a $(m \times n)$ matrix called the Jacobian matrix. Before we derive the form of the Jacobian matrix we state (The proof is analogous to that for functions from $\Re^n \to \Re$) the following:

Theorem 23.5. *Let $\underline{f}$: $S \subseteq \Re^n \to \Re^m$ and x_0 interior point of S. Let f be differentiable at x_0. Then it has directional derivatives $\underline{f}'(x_0, d)$ for every $d \in \Re^n$ and we have*

$$T_{x_0}(d) = \underline{f}^{'}(x_0, d)$$

and

$$T_{x_0}(d) = \sum_{i=1}^{n} d_k D_k \underline{f}(x_0)$$

A consequence of the linearity of T_{x_0} is that

$$T_{x_0}(d) = T_{x_0}(d_1e_1 + ... + d_ne_n) = d_1T_{x_0}(e_1) + ... + d_nT_{x_0}(e_n) = \sum_{i=1}^{n} d_iT_{x_0}(e_i)$$

but from the previous theorem

$$T_{x_0}(e_k) \quad = \quad D_k\underline{f}(x_0) = \sum_{i=1}^{m} D_kfi(x_0)u_i$$
$$\textit{where } u_i \textit{ are the canonical bases in } \Re^m$$

therefore we can write

$$T_{x_0}(d) = \sum_{i=1}^{n} d_i \sum_{j=1}^{m} fj_i(x_0)u_j$$

or

$$T_{x_0}(d) = \begin{matrix} d_1f1_1(x_0) + ... + d_nf1_n(x_0) \\ \vdots \\ d_1fm_1(x_0) + ... + d_nfm_n(x_0) \end{matrix}$$

so that the matrix associated with the linear function $T_{x_0}(d)$, denoted by $J\underline{f}(x_0)$ is

$$J\underline{f}(x_0) = \begin{matrix} f1_1(x_0) & \cdots & f1_n(x_0) \\ \vdots & & \vdots \\ fm_1(x_0) & \cdots & fm_n(x_0) \end{matrix}$$

notice that the k^{th} row of the Jacobian is simply the gradient of $fk(x_0)$ and that in the special case of a function from $\Re^n$ to $\Re$ the Jacobian consist in only one row and is the gradient of the function.

23.1. The chain rule

We will see now how for vector valued function the Jacobian matrix of a composite function can be found by multiplying the Jacobian matrices of two functions:

Theorem 23.6. *Let $\underline{f}$: $\Re^n \to \Re^m$ and $\underline{g}$: $\Re^q \to \Re^n$ and let $\underline{g}$ be differentiable at a with total derivative $\underline{g}'_a()$ and $b = \underline{g}(a)$. Let also $\underline{f}$ be differentiable at b with total derivative $\underline{f}'_b()$. Then the composite function $\underline{f \circ g}\ (a) = \underline{h}(a)$ is differentiable at* *a and its total derivative is given by the composition of the linear functions* $\underline{g}'_a()$ *and* $\underline{f}'_b()$ *that is*

$$\underline{h}'_a() = \underline{f \circ g}'_a() = \underline{f}'_b() \circ \underline{g}'_a() = \underline{f}'_b[\underline{g}'_a()]$$

Proof. *We want to show that*

$$\underline{h}(a+y) - \underline{h}(a) = \underline{f}'_b[\underline{g}'_a(y)] + \|y\|\, E(y) \tag{23.2}$$
$$\textit{where } E(y) \to 0 \text{ as } y \to 0$$

Let $b = \underline{g}(a)$, *and* $v = \underline{g}(a+y) - \underline{g}(a)$. *Then, since if differentiable,*

$$\begin{aligned}\underline{h}(a+y) - \underline{h}(a) = \underline{f}(\underline{g}(a+y)) - \underline{f}(\underline{g}(a)) = \underline{f}(b+v) - \underline{f}(b) = \\ \underline{f}'_b(v) + \|v\|\, E_b(v) \\ \textit{where } E_b(v) \to 0 \textit{ as } v \to 0\end{aligned} \tag{23.3}$$

and since $\underline{g}$ *is differentiable we have*

$$v = \underline{g}(a+y) - \underline{g}(a) = \underline{g}'_a(y) + \|y\|\, E_a(y) \tag{23.4}$$
$$\textit{where } E_a(y) \to 0 \textit{ as } y \to 0$$

substituting 23.4 in 23.3 we get

$$\begin{aligned}\underline{h}(a+y) - \underline{h}(a) = \underline{f}'_b[\underline{g}'_a(y)] + \|y\|\, \underline{f}'_b(E_a(y)) + \|v\|\, E_b(v) \\ = \underline{f}'_b[\underline{g}'_a(y)] + \|y\|\, E(y) \\ \textit{where } E(y) = \underline{f}'_b(E_a(y)) + \frac{\|v\|}{\|y\|} E_b(v)\end{aligned}$$

so we need to show that $E(y) = \underline{f}'_b(E_a(y)) + \frac{\|v\|}{\|y\|} E_b(v)$ *goes to* 0 *as* y *goes to* 0. $\underline{f}'_b(E_a(y))$ *goes to* 0 *as* $y \to 0$ *since* $E_a(y) \to 0$ *and* $\underline{f}'_b()$ *is linear. When* y *goes to* 0 v *goes to* 0 *as well so* $E_b(v)$ *goes to* 0 *too. We therefore need to show that* $\frac{\|v\|}{\|y\|}$ *doesn't go to infinity as* y *goes to* 0. *Using the triangular inequality on 23.4 we have*

$$\|v\| \leq \left\|\underline{g}'_a(y)\right\| + \|y\|\, \|E_a(y)\|$$

but using first the triangular inequality and then the Cauchy Schwarz inequality we have

$$\left\|\underline{g}'_a(y)\right\| = \left\|\sum_{i=1}^{m} [Dfi(a) \cdot y] e_i\right\| \leq \sum_{i=1}^{m} \|[Dfi(a) \cdot y] e_i\| = \sum_{i=1}^{m} |Dfi(a) \cdot y| \leq \|y\| \sum_{i=1}^{m} \|Dfi(a)\|$$

where $\sum_{i=1}^{m} \|Dfi(a)\|$ *is bounded by a number* M *since* f *is differentiable. We can therefore write*

$$\|v\| \leq M \|y\| + \|y\| \|E_a(y)\|$$

and then

$$\frac{\|v\|}{\|y\|} \leq M + \|E_a(y)\|$$

We can finally conclude that as $y \to 0$ $E(y) \to 0$ *thereby proving 23.2.* ■

Remark 11. *We have seen that the matrix associated to the total derivative is the Jacobian matrix of the function in the point, that is*

$$\underline{g}'_a(y) = J\underline{g}(a) \cdot y$$

a consequence of that and that of the chain rule is that $J(\underline{f \circ g})(a) = J\underline{f}(b)J\underline{g}(a)$ *that is the Jacobian matrix of the composite of* f *and* g *is equal to the matrix multiplication of the Jacobian matrices of* f *and* g.

24. Applications of multivariate differential calculus

We will now see three important applications of multivariate differential calculus namely the mean value theorem, the inverse function theorem and finally the famous implicit function theorem.

Theorem 24.1. *(Mean Value Theorem) Let* $S \subseteq \Re^n$ *be an open set and let* $\underline{f}\colon S \to \Re^m$ *be differentiable in* S*. Let also* x *and* y *be two points in* S *such that the entire segment joining them (* $L(x, y) = \{tx + (1 - t)y : 0 \leq t \leq 1\}$ *) is in* S*. Then for every vector* $a \in \Re^m$ *there is a point* $z \in L(x, y)$ *s.t.*

$$x \cdot (\underline{f}(x) - \underline{f}(y)) = a \cdot \underline{f}_z(x - y) = a \cdot Jf(z) \cdot (x - y)$$

Remark 12. *If* $\underline{f}\colon S \to \Re$ *then we can pick* $a = 1$ *and the theorem states*

$$f(x) - f(y) = Df_z(x - y) = Df(z) \cdot (x - y)$$

Theorem 24.2. *(Inverse function theorem) Let* $\underline{f} : S \to \Re^n$ *be a* C^1 *(Continuous with continuous partial derivatives) function with* S *open subset of* $\Re^n$ *and let* T *be the image of* S *under* $\underline{f}$*. If the determinant of the Jacobian matrix* $|Jf(a)|$*is different from* 0 *for some point* $a \in S$ *then there are two open sets* $X \subseteq S$ *and* $Y \subseteq T$ *and an unique function* $\underline{g}$*. that satisfy:*

- $a \in X$ *and* $\underline{f}(a) \in Y$
- $\underline{f}$ *is bijective from* X *to* Y
- $\underline{g}$ *is bijective from* Y *to* X *and satisfies* $\underline{g}(\underline{f}(x)) = x$, $\forall x \in X$
- $\underline{g}$ *is* C^1
- $J\underline{g}(y) = J^{-1}(f(g(y))$

The inverse theorem gives us condition under which the following system of (not necessarily linear) equations:

$$\begin{array}{rcl} y_1 & = & f1(x_1, .., x_n) \\ \cdots & & \cdots \\ y_n & = & fn(x_1, .., x_n) \end{array}$$

can be (locally) solved for x_1,.. x_n in function of y_n and also guarantees that locally the solutions are unique, continuous and continuously differentiable. Suppose now we have a more general system of equations

$$\begin{array}{rcl} 0 & = & f1(x_1, .., x_n, y_1, ...y_k) \\ \cdots & & \cdots \\ 0 & = & fn(x_1, .., x_n, y_1, ...y_k) \end{array}$$

and we ask whether we can solve for $x_1, ...x_n$ in function of $y_1, ...y_k$.The tool we need is the implicit function theorem. In the following we will use the letters x and y to denote vectors in $\Re^n$ and $\Re^k$ respectively and (x, y) will denote a vector in $\Re^{n+k}$.

Theorem 24.3. *(Implicit function theorem or Dini's theorem).Let* $\underline{f} : S \to \Re^n$, *where* S *is an open subset of* $\Re^{n+k}$, *be a* C^1 *function. Let* (x_0, y_0) *be a point for which* $\underline{f}(x_{0,}y_0) = 0$ *and for which* $\det(J_x f(x_0, y_0) \neq 0$ *where*

$$J_x\underline{f}(x_0, y_0) = \begin{array}{ccc} D_{x_1} f1(x_0, y_0) & \cdots & D_{x_n} f1(x_0, y_0) \\ \vdots & & \vdots \\ D_{x_1} fn(x_0, y_0) & \cdots & D_{x_n} fn(x_0, y_0) \end{array}$$

then there exist an open set $W \subseteq \Re^k$ *and an open set* $V \subseteq \Re^{n+k}$ *with* $y_0 \in W$ *and* $(x_0, y_0) \in V$ *and a unique function* $\underline{g}$: $W \to \Re^n$ *satisfying:*

- $(\underline{g}(y), y) \in V$ and $\underline{f}(\underline{g}(y), y) = 0 \ \forall y \in W$
- $\underline{g}$ is C^1
- $J\underline{g}(y_0) = -\left[J_x \underline{f}(x_0, y_0)\right]^{-1} J_y \underline{f}(x_0, y_0)$

25. Homogeneous functions

In this section we will present some results on a particular class of functions particularly useful in economics

Definition 25.1. *A set $A \subseteq \Re^n$ is said a cone if $x \in A \to \rho x \in A$ with $\rho \geq 0$.*

Definition 25.2. *A function $f : A \to \Re$ where A is a cone subset of $\Re^n$ is said (positively) homogenous of degree a in A if $\forall x \in A$ and $\forall \ \rho > 0$*

$$f(\rho x) = \rho^a f(x)$$

Remark 13. *Let f be a function homogeneous of degree $a \neq 0$. Then it must be*

$$f(0) = 0$$

Theorem 25.3. *Let $f : A \to \Re$ where A is a cone subset of $\Re^n$ be a derivable and homogeneous function of degree a. Then its partial derivatives are homogeneous of degree $a - 1$.*

Theorem 25.4. *(Euler's) Let $f : A \to \Re$ where A is a cone subset of $\Re^n$ be a differentiable function on A. f is homogeneous of degree a on A if and only if we have*

$$xDf(x) = af(x)$$