Chapter 2

Functions of several variables

Most financial models include several variables. Portfolio choice depends on at least two variables, the expected return and the variance of returns of the portfolio. When n assets are traded on the market, the utility of a portfolio depends on the n weights of the assets in the portfolio. Functions of several variables also appear naturally in option pricing. The standard Black-Scholes option pricing model (1973) allows to evaluate an option contract as a function of six variables, namely the price of the underlying asset, the time to maturity, the volatility of the underlying asset, the strike price, the risk-free rate and the dividends paid on the underlying.

In this chapter, we start by concepts generalizing chapter 2 of part I which was focused on single-variable functions. First, we need to generalize some results of chapter 1 of part I. Section 2.1 defines a metric space and the notion of distance (also called metric) on a metric space. The concept of distance is very general, but in the present chapter we essentially apply it to finite-dimensional spaces like $\mathbb{R}^p$, to study functions depending on p variables. Section 2.2 presents continuity and differentiability of functions depending on several variables and some important results like multidimensional Taylor's formulas. These formulas are interesting when it comes to approximating functions by polynomials ot to stating optimality conditions (see chapters 3

Download free eBooks at bookboon.com

and 4). Finally, section 3 deals with implicit differentiation and homogeneous functions.

2.1 Metric spaces

If you are interested in overseas races, like the Vendée Globe Challenge, you want to know the ranking of boats at regular time intervals. On the website of the race[1] you can download a map where the boats are represented on the ocean and you can also see the total remaining distance. If you think to the problem a few minutes, you see that it is not a trivial matter to calculate a distance between two points A and B on a sphere (a reasonable approximation for our planet). It becomes even more difficult when constraints are added to the problem (boats are supposed to stay on the water!). The distance is not the same as the crow flies or for people who possibly need to climb mountains or stay on oceans. Think to people walking in New York City, or in major U.S towns. As streets are orthogonal to each other it is not very useful to know that, the distance between A to B as the crow flies is 5 miles.

In mathematical terms, a distance should be a mapping linking the pair (A, B) to a positive number and satisfying some reasonable properties. But this mapping should also be sufficiently general to adapt to many different contexts.

2.1.1 Metric on a set

Definition 83 *A **distance (metric)** on a set E is a mapping d from $E \times E$ to $\mathbb{R}_+$ satisfying:*

1) $\forall (x, y) \in E \times E,\ d(x, y) = d(y, x)$ *(symmetry).*

[1] http://www.vendeeglobe.org

Download free eBooks at bookboon.com

2) $\forall (x,y) \in E \times E,\ d(x,y) = 0$ *if* $x = y$ *and* $d(x,y) > 0$ *otherwise (positivity).*

3) $\forall (x,y,z) \in E \times E \times E,\ d(x,z) \leq d(x,y) + d(y,z)$ *(triangular inequality).*

The pair (E,d) *is called a* ***metric space.***

Part (2) says that if the distance between two elements is zero, they are identical. Though this property seems very intuitive, we are going to provide examples showing that this intuition can be misleading. Part (3) says, in everyday language, that the shortest route between two points x and z is a straight line. One more time, it seems intuitive when a distance on real numbers is defined by $d(x,y) = |x-y|$. In this case the distance between x and z is the length of the segment joining the two points. But recall the Vendée Globe Challenge! On a sphere, there are no straight lines!

Consider for example the 2-dimensional space $\mathbb{R}^2$; the usual metric on this space, called the Euclidian distance, is defined by:

$$d_0(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{2.1}$$

with $x = (x_1, x_2)$ and $y = (y_1, y_2)$.

This metric measures the "physical" distance between x and y. The reader can easily check this is the case by using the Pythagorean theorem.

Of course, the Euclidian metric is easily generalized to p-dimensional spaces as follows:

$$d_0(x,y) = \sqrt{\sum_{k=1}^{p} (x_k - y_k)^2} \tag{2.2}$$

with $x = (x_1, x_2, ..., x_p)$ and $y = (y_1, y_2, ..., y_p)$ two elements of $\mathbb{R}^p$.

However, our "sailing" example shows that there are many ways to measure distances on a sphere or in New York City. In this latter case, the real

Download free eBooks at bookboon.com

distance should be defined as d^{*2}:

$$d^*(x, y) = |x_1 - y_1| + |x_2 - y_2|$$

The reader can check that d^* satisfies the three properties of definition 83 and that any norm $\|.\|$ on a vector space E induces a metric d on E. d is defined as:

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

where x and y are two vectors in E. Consequently, a normed vector space is also a metric space (E, d) when d is the metric induced by the norm on E.

[2] We assume that streets are either parallel or orthogonal to axes in the two-dimensional space.

Download free eBooks at bookboon.com

Click on the ad to read more

Example 84 *Forecasts by financial analysts*

Financial analysts publish earnings and dividends forecasts, and target prices as well. These forecasts are important for fund managers, banks and investment advisors. Some firms also calculate a market consensus to summarize the forecasts of a set of analysts. For a given firm, the most simple summary consists in averaging the forecasts of analysts. However, such an average provides no clue about the dispersion of forecasts. Imagine that there are two firms F_1 and F_2, and two analysts A_1 and A_2. The following matrix shows the earnings forecasts:

$$M = \begin{bmatrix} & A_1 & A_2 \\ F_1 & 11 & 9 \\ F_2 & 2 & 18 \end{bmatrix} \tag{2.3}$$

The consensus (average) is 10 for the two stocks, but the forecasts are much more dispersed for the second stock. Using the consensus to take investment decisions is more risky and error-prone if individual forecasts are more dispersed. More generally, assume that the two analysts provide forecasts on N stocks and denote $(p_1^1, ..., p_N^1)$ et $(p_1^2, ..., p_N^2)$ these forecasts. For each company, i, the consensus is defined as the average:

$$\overline{p}_i = \frac{1}{2}\left(p_i^1 + p_i^2\right)$$

The risk of the consensus forecast for firm i can then be defined as the distance d_i between the vector of individual forecasts (p_i^1, p_i^2) and the pair of consensus forecasts $(\overline{p}_i, \overline{p}_i)$ that would obtain if the two analysts were predicting the same earnings.

$$d_i = \sqrt{\left(p_i^1 - \overline{p}_i\right)^2 + \left(p_i^2 - \overline{p}_i\right)^2}$$

Of course if the forecasts are actually equal, the distance is 0, meaning that there is no divergence between analysts. The geometric interpretation of this result is simply that identical forecasts lie on the first bisector of the two-

Download free eBooks at bookboon.com

dimensional space where the first (second) axis represents the forecasts of the first (second) analyst. d_i is in fact the distance between firm i and the first bisector. Of course, d_i is an oversimplified measure of divergence but it is important to note that all dispersion measures are built in the same spirit.

In our example, the benchmark is the bisector where all forecasts concerning a given firm are identical. However, suppose that analysts build their forecasts using two types of factors. First, there are macroeconomic factors (or common factors) influencing all firms, and firm-specific factors. In such a framework, it could be better to neutralize the divergence on common factors to measure the forecasting risk of firm i. For example, if the first analyst is much more optimistic than the second about common factors, her forecasts will be higher on a large part of the firms under scrutiny. The consequence is that most p_i will be under the first bisector and this bisector is not the good benchmark! It is necessary to find another benchmark taking into account the divergence about macroeconomic factors. One of the popular methods to do so is Principal Component Analysis which allows to find the line D minimizing the following quantity:

$$\sum_{i=1}^{N} d(p_i, D)^2$$

We do not elaborate in more details this example but the reader should remember that this topic is extensively studied in finance research.

2.1.2 Open sets in metric spaces

Open and closed sets in $\mathbb{R}$ have been introduced in chapter 1 of part I. These concepts are still valid in metric spaces with minor changes in the definitions.

Definition 85 *Let (E, d) denote a metric space. An **open ball** centered at x with radius r, denoted $B^{\circ}(x, r)$, is the set of elements $y \in E$ satisfying*

Download free eBooks at bookboon.com

$d(x,y) < r$. *This set can be formally written as:*

$$B^{\circ}(x,r) = \{y \in E \quad \textit{such that } d(x,y) < r\}$$

A ***closed ball*** *centered at x with radius r, denoted $\overline{B}(x,r)$, is the set of elements $y \in E$ satisfying $d(x,y) \leq r$. This set can be formally written as:*

$$\overline{B}(x,r) = \{y \in E \quad \textit{such that } d(x,y) \leq r\}$$

In the set $\mathbb{R}$ of real numbers, equipped with the usual metric, the open ball centered at $x \in \mathbb{R}$ with radius r is simply the interval $]x-r; x+r[$

The corresponding closed ball is the closed interval $[x-r; x+r]$. As a consequence, the concept of an open ball in a metric space is the natural generalization of open intervals in $\mathbb{R}$.

Definition 86 *Let G be a subset of a metric space E.*

$x \in G$ is ***interior*** *to G if there exists an open ball centered at x with radius $r > 0$, satisfying $B^{\circ}(x,r) \subset G$.*

G is an ***open*** *set in E if any element in G is interior to G, that is:*

$$\forall x \in G, \exists r \in \mathbb{R}_+^* \textit{ such that } B^{\circ}(x,r) \subset G$$

As before, we deduce immediately the definition of a closed set.

Definition 87 *A subset F of a metric space E is* ***closed*** *if the complement $F^c = \{x \in E \quad \textit{such that } x \notin F\}$ is an open set.*

The disc $G \subset \mathbb{R}^2$ that appears on figure 2.1 is open if the boundary circle is not in G; it is closed otherwise.

The following proposition is valid in any metric space.

Proposition 88 *a) Any union of open sets in E is an open set and any finite union of closed sets in E is a closed set in E.*

Download free eBooks at bookboon.com

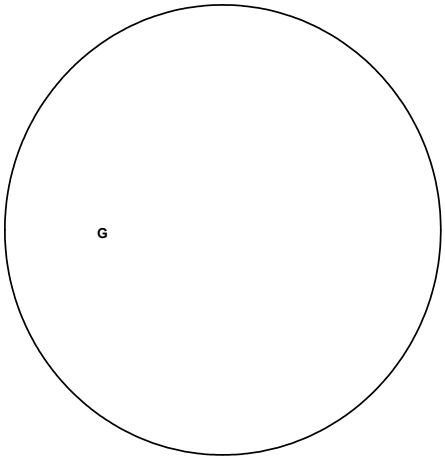


Figure 2.1: Disc in the plane

b) Any intersection of closed sets in E is a closed set and any finite intersection of open sets in E is an open set.

c) E and $\emptyset$ are simultaneously open and closed.

Proposition 88 shows that there is an asymmetry between open and closed subsets. In (a), the union is considered over any number (finite or not) of open subsets but only over a finite number of closed subsets. In (b), the intersection is over any number of closed subsets but over a finite number of open subsets. The difference may be illustrated by the following example; consider the sequence of open intervals $\left(\right]-\frac{1}{n};\frac{1}{n}\left[\right., n \in \mathbb{N}^*)$. We get:

$$\bigcap_{n\in\mathbb{N}^*} \left]-\frac{1}{n};\frac{1}{n}\right[= \{0\}$$

The set $\{0\}$ is a closed subset of $\mathbb{R}$ but is written as an infinite intersection of open subsets.

The other topological concepts are generalizations (more or less intuitive) of what we presented in chapter 1 of part I for the set $\mathbb{R}$. We briefly recall

Download free eBooks at bookboon.com

these definitions for the sake of completeness.

Definition 89 *a) The* ***closure*** *of a subset H of a metric space (E,d), denoted $\overline{H}$, is the smallest closed subset such that $H \subset \overline{H}$. It is also the intersection of all closed subsets containing H.*

b) The ***interior*** *of a subset H of a metric space (E,d), denoted H°, is the largest open subset such that $H^{\circ} \subset H$, or the union of all open subsets included in H.*

c) The ***exterior*** *of a subset H of a metric space (E,d), is the interior of the complement of H in E.*

d) The ***frontier*** *of a subset H of a metric space (E,d), is the set of elements in E that are neither in the interior nor in the exterior of H.*

Download free eBooks at bookboon.com

Click on the ad to read more

All these definitions were already given in the framework of the metric space $\mathbb{R}$. We need now to add a more abstract concept which will be useful later on.

Definition 90 *a) A subset H in E is a **dense subset** of E if $\overline{H} = E$ where $\overline{H}$ is the closure of H.*

*b) A metric space (E, d) is **separable** if E contains a dense countable subset*[3]*.*

Another way to say the same thing is that for any x in E, there exists a sequence in H converging to x. In short:

$$\forall x \in E, \exists (y_n \in H, n \in \mathbb{N}), \lim_{n \to +\infty} d(y_n, x) = 0 \tag{2.4}$$

or

$$\forall x \in E, \forall \varepsilon > 0, \exists x^* \in H, d(x^*, x) < \varepsilon \tag{2.5}$$

Property 2.5 shows that saying a set is dense has something to do with an approximation. For any element x of E and any distance ε, it is possible to find an element of H as close as desired (at a distance lower than ε) of x.

Example 91 *The most standard (which also proves the most useful) example is the set $\mathbb{Q}$ of rational numbers which is dense in $\mathbb{R}$. For example, in any calculator, numbers like π or e are approximated by rational numbers (that is ratios of integers), with the desired level of accuracy. Doing so is relatively safe because $\mathbb{Q}$ is dense in $\mathbb{R}$.*

In $\mathbb{R}$, a set is *bounded* if it is included in an interval with finite ends. This definition can be easily adapted to metric spaces, using balls instead of intervals.

[3]Recall that a set A is countable if one can "count" its elements. In other words A is countable if there exists a bijection between the set $\mathbb{N}$ of positive integers and A.

Download free eBooks at bookboon.com

Definition 92 *A subset G of a metric space (E, d) is said* ***bounded*** *if it is included in a ball $B(x, r)$ with $r < +\infty$.*

A counterintuitive result is that boundedness depends on the metric. A set can be bounded for a given metric and unbounded for another metric. For example, there exists on $\mathbb{R}$ a metric called the *discrete metric* , defined by :

$$d^*(x, y) = \begin{cases} 1 \text{ if } x \neq y \\ 0 \text{ otherwise} \end{cases}$$

The interval $]-\infty; 5]$ is bounded under d^* because it is included in $B^*(0, 1)$. This mapping d^* gives almost no information about the location of points in the metric space. We can only say if two elements are identical or not when we know the distance between them (0 or 1).

Finally, a compact set in $\mathbb{R}$ was a bounded and closed subset. It is still true in R^p, but it is false in general metric spaces[4]. In chapter 2 of part I we saw that a function defined on a compact set reaches its bounds and possesses a maximum and a minimum. This proposition is still valid for functions of several variables considered in this chapter.

2.1.3 Sequences in metric spaces

To define the convergence of a sequence in $\mathbb{R}$, we used absolute values $|x_n - x|$, where x denoted the limit. The same concept in metric spaces uses distances, in particular in R^p. Nothing surprising here because the mapping $(x, y) \to |x - y|$ is a metric on $\mathbb{R}$.

Definition 93 *Let (E, d) denote a metric space, $(x_n, n \in \mathbb{N})$ a sequence of elements of E and x an element of E.*

[4]In a general metric space, a set A is **compact** if, from any sequence of elements of A, it is possible to extract a convergent sub-sequence. In this book we will not need such a general definition.

Download free eBooks at bookboon.com

$(x_n, n \in \mathbb{N})$ ***converges*** *to* x *if* $\lim_{n\to+\infty} d(x_n, x) = 0$. *We write* $\lim_{n\to+\infty} x_n = x$.

$(x_n, n \in \mathbb{N})$ *is called a* ***Cauchy sequence*** *if* $\lim_{i,j\to+\infty} d(x_i, x_j) = 0$.

Proposition 48 in chapter 1 of part 1, related to the convergence of Cauchy sequences in $\mathbb{R}$, does not remain valid in general metric spaces, but remains true in $E = \mathbb{R}^p$.

2.2 Continuity and differentiability

This section introduces continuity and differentiability of functions defined on a subset of $\mathbb{R}^p$ and taking their values in $\mathbb{R}$. The set $\mathbb{R}^p$ is endowed with the Euclidean metric, unless otherwise stated. Tools of the previous section will allow to generalize the notions of limit, continuity and differentiability presented in part I for functions of one variable. Taylor's formula is also generalized.

2.2.1 Limits and continuity

Definition 94 *Let* f *be a function defined on an open set* $D \subset \mathbb{R}^p$. f *has a* ***limit*** $b \in \mathbb{R}$, *at* $a \in D$ *if, for any sequence* $(x_n, n \in \mathbb{N})$ *in* D *that converges to* a, *the sequence* $(f(x_n), n \in \mathbb{N})$ *converges to* b. *We write:*

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

This definition is almost identical to the definition of a limit in chapter 2 of part I. However, here the convergence of $(x_n, n \in \mathbb{N})$ refers to definition 93.

Definition 95 *Let* f *be a function defined on an open set* $D \subset \mathbb{R}^p$. f *is*

Download free eBooks at bookboon.com

continuous *at* $x^* = \left(x_1^*, ..., x_p^*\right) \in D$ *if:*

$$\lim_{x \to x^*} f(x) = f(x^*)$$

Here too, the definition is very close to the definition of continuity for functions of one variable. The only difference lies in the use of a metric on R^p. Moreover, it turns out that left and right continuity are meaningless in multidimensional spaces.

Remember that $\lim_{x \to x^*} f(x) = f(x^*)$ can also be written as:

$$\lim_{x \to x^*} |f(x) - f(x^*)| = 0 \tag{2.6}$$

Therefore, continuity could be defined in a much more general way for functions f defined on an open subset D of a metric space (E, d) and taking values in another metric space (F, d_F). In this general framework, f is continuous at $x^* \in E$ if:

$$\lim_{x \to x^*} d_F\left(f(x), f(x^*)\right) = 0$$

We just replaced the metric on $\mathbb{R}$ (defined by the absolute value $|f(x) - f(x^*)|$) by $d_F\left(f(x), f(x^*)\right)$. In particular, we encounter this situation when considering functions defined on R^p and taking values in R^m.

2.2.2 Partial derivatives

One of the essential differences between functions of one and of several variables lies in the concept of derivative. Remember that for $f : \mathbb{R} \to \mathbb{R}$, the derivative at x_0 is defined as follows:

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

There is no obvious generalization for functions of several variables because $x_0 \in \mathbb{R}^p$, $p > 1$. In fact, h should also have p components, and dividing

Download free eBooks at bookboon.com

by h would be meaningless. Therefore, for a function depending on p variables, we define p partial derivatives, each one being defined as a derivative with respect to one variable, the other variables being assumed constant.

Definition 96 *The **partial derivative** of f at x^*, with respect to the i-th variable, is the limit, if it exists, defined by:*

$$\lim_{h\to 0}\frac{f(x_1^*,..,x_i^*+h,...,x_p^*)-f(x^*)}{h}$$

Alternative notations are $\frac{\partial f}{\partial x_i}(x^*)$ or sometimes $f_{x_i}(x^*)$.

We know how to derive a single-variable function. A function of p variables is a function of one variable when $p-1$ variables are held constant. So the definition works as if $p-1$ variables where not changing. In fact, let g be the function defined by:

$$g(x)=f(x_1^*,...,x_{i-1}^*,x,x_{i+1}^*,...,x_p^*)$$

the derivative of g at x_i^* writes:

$$\begin{aligned} g'(x_i^*) &= \lim_{h\to 0}\frac{g(x_i^*+h)-g(x_i^*)}{h} \\ &= \lim_{h\to 0}\frac{f(x_1^*,..,x_i^*+h,...,x_n^*)-f(x^*)}{h}=\frac{\partial f}{\partial x_i}(x^*) \end{aligned}$$

What we just wrote for the i-th variable can be written the same way for the other $p-1$ variables. Consequently, we get p partial derivatives (when the corresponding limits exist).

In financial and economic models, it is common to assume that partial derivatives are continuous functions. Such functions are said C^1-functions.

Example 97 *Let f be a function on $\mathbb{R}_+\times\mathbb{R}_+$ taking values in $\mathbb{R}$ and defined by:*

$$x=(x_1,x_2)\to f(x)=\sqrt{x_1x_2}$$

Download free eBooks at bookboon.com

At any $x^* \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ *this function has partial derivatives defined by:*

$$\begin{aligned}\frac{\partial f}{\partial x_1}(x^*) &= \frac{1}{2}\sqrt{\frac{x_2^*}{x_1^*}} \\ \frac{\partial f}{\partial x_2}(x^*) &= \frac{1}{2}\sqrt{\frac{x_1^*}{x_2^*}}\end{aligned}$$

In fact we can write:

$$f(x) = \sqrt{x_1}\sqrt{x_2}$$

To compute $\frac{\partial f}{\partial x_1}(x^*)$*, we consider that* $\sqrt{x_2}$ *is a number equal to* $\sqrt{x_2^*}$ *(let us denote* c *this number) and we compute the derivative at* x_1^* *of the one-variable* $g(x_1) = c\sqrt{x_1}$*. This derivative is equal to:*

$$g'(x_1^*) = c \times \frac{1}{2\sqrt{x_1^*}}$$

Replace now c *by its value, that is* $\sqrt{x_2^*}$*. The result is:*

$$g'(x_1^*) = \frac{\partial f}{\partial x_1}(x^*) = \frac{1}{2}\sqrt{\frac{x_2^*}{x_1^*}}$$

The computation of $\frac{\partial f}{\partial x_2}(x^*)$ *can be done in the same way, replacing* $\sqrt{x_1}$ *by a number* b *equal to* $\sqrt{x_1^*}$*. The single-variable function is now denoted* m *and defined by* $m(x_2) = b\sqrt{x_2}$*.* m' *is then calculated as usual.*

$$m'(x_2) = b \times \frac{1}{2\sqrt{x_2^*}} = \frac{\partial f}{\partial x_2}(x^*) = \frac{1}{2}\sqrt{\frac{x_1^*}{x_2^*}}$$

Definition 98 *The* $p-$*dimensional vector* $\frac{\partial f}{\partial x_i}(x^*), i = 1, ..., p,$ *is called the*

Download free eBooks at bookboon.com

gradient *of* f *at* x^*. *It is denoted* $\nabla f(x^*)$ *(spell nabla for* ∇*):*

$$\nabla f(x^*) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x^*) \\ ... \\ \frac{\partial f}{\partial x_p}(x^*) \end{pmatrix}$$

$\nabla f(x^*)$ is an element of the vector space R^p. Therefore, $\nabla f(x^*)$ denotes a matrix with p rows and 1 column, containing the partial derivatives of f valued at x^*.

Download free eBooks at bookboon.com
Click on the ad to read more

2.2.3 Derivatives of compound functions

Calculation of partial derivatives of compound functions obeys the same rules as the ones used for functions of one variable, but the formulation is a bit more complex.

The proposition hereafter presents the case of functions depending on two variables.

Proposition 99 *Let f, g, h three continuously differentiable functions, defined on a open set $D \subset \mathbb{R}^2$. We have:*

$$\frac{\partial}{\partial x}\left[f\left(g(x,y),h(x,y)\right)\right] = \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x}$$

where $u = g(x,y)$ and $v = h(x,y)$. The partial derivative with respect to the second variable y is defined accordingly (replacing ∂x by ∂y).

Example 100 *Let f, g and h be defined as follows:*

$$f(u,v) = \exp(uv) \tag{2.7}$$
$$g(x,y) = x + y \tag{2.8}$$
$$h(x,y) = x - y \tag{2.9}$$

First, calculation of $\frac{\partial f}{\partial u}$ and $\frac{\partial f}{\partial v}$:

$$\frac{\partial f}{\partial u} = vf(u,v) \tag{2.10}$$
$$\frac{\partial f}{\partial v} = uf(u,v) \tag{2.11}$$

Second, calculation of $\frac{\partial u}{\partial x}$ and $\frac{\partial v}{\partial x}$:

$$\frac{\partial u}{\partial x} = 1 = \frac{\partial v}{\partial x} \tag{2.12}$$

Download free eBooks at bookboon.com

Finally

$$\frac{\partial}{\partial x}\left[f\left(g(x,y),h(x,y)\right)\right] = (x-y)\exp(x^2-y^2)+(x+y)\exp(x^2-y^2) \quad (2.13)$$
$$= 2x\exp(x^2-y^2) \quad (2.14)$$

Of course, in this example it would have been easier to directly replace u and v by their values and start with $f(x,y)=\exp(x^2-y^2)$

2.2.4 Differential of a function depending on several variables

In chapter 2 of part I, a C^1-function was approximated at $x+h$ by $f(x)+hf'(x)$ (first-order approximation). The mapping $h \to hf'(x)$ is linear. The differential of a function of several variables carries the same idea. However, starting from $x \in R^p$, we can move in different directions. In other words, writing $x+h$ refers to a vector like

$$x+h=\begin{pmatrix} x_1+h_1 \\ ... \\ x_p+h_p \end{pmatrix}$$

We then refer to a displacement in the direction of vector h.

The fact that partial derivatives exist is not sufficient to ensure that a function of several variables is continuous. There exist some pathological cases where the function possesses partial derivatives according to definition 95 but is not continuous. To solve this difficult question we need a little bit more, that is differentiability. We provide hereafter the definition of this word but in the sequel of the book we will in fact use a stronger (but much more intuitive) assumption.

Definition 101 *A function f defined on an open subset D of $\mathbb{R}^p$ is **differ-***

Download free eBooks at bookboon.com

entiable *at $x \in D$ if there exists $\alpha \in \mathbb{R}^p$ such that:*

$$f(x+h) = f(x) + \sum_{i=1}^{p} \alpha_i \frac{\partial f}{\partial x_i}(x) + \|h\| \varepsilon(h)$$

with $\lim_{h \to \mathbf{0}} \varepsilon(h) = 0$.

Remember that $\|h\|$ denotes the norm of vector h. In the limit, we used bold characters for the null vector to emphasize the fact that h is a vector, not a number. In the remaining of the text, we use proposition 102 to simplify the formulation of propositions.

Proposition 102 *Any C^1 function at x is differentiable at x.*

In the following definitions and propositions we assume that functions are C^1 over the interior of their domain. Mathematicians would say that weaker assumptions are better, but assuming C^1-functions is general enough for economics and finance.

Definition 103 *Let f be a C^1-function defined on an open set $D \subset \mathbb{R}^p$. The* ***differential*** *of f at x^* is the linear form, denoted df_{x^*}, defined on $\mathbb{R}^p$ as follows:*

$$df_{x^*}(h) = \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(x^*) h_i$$

with $h^T = (h_1, ..., h_p)$.

Using notations of chapter 1, $df_{x^*}(h)$ writes as the following inner product in R^p :

$$df_{x^*}(h) = < \nabla f(x^*), h >$$

Proposition 79 of chapter 1 (Riesz representation theorem) allows to say that $\nabla f(x^*)$ represents the linear mapping df_{x^*} because, for any h:

$$df_{x^*}(h) = < \nabla f(x^*), h >$$

Download free eBooks at bookboon.com

The cases $p = 1$ and $p = 2$ reveal the intuition behind the definition of differentials. Assume that the components of h are close to 0; $df_{x^*}(h)$ then approximates the difference $f(x^* + h) - f(x^*)$. If $p = 1, hf'(x^*)$ is a first-order approximation of $f(x^* + h) - f(x^*)$. The derivative $f'(x^*)$ also denotes the slope of the tangent to the curve representing f. If $p = 2$, the geometric interpretation of the differential is the same; the mapping $h \to df_{x^*}(h) = h_1 \frac{\partial f}{\partial x_1}(x^*) + h_2 \frac{\partial f}{\partial x_2}(x^*)$ approximates the surface f at x^* by the two-dimensional space tangent to the surface at x^*. As f is a C^1-function, it is differentiable; therefore this approximation is valid when the norm of h is small, that is when $x^* + h$ is close to x^* in the space $\mathbb{R}^p$.

Example 104 ***Interpretation of differentials***

Come back to the function $f(x) = \sqrt{x_1 x_2}$ *and define* $df_{x^*}(h)$ *for* $(x^*)^T = (1; 1)$. *Example 97 indicates that:*

$$df_{x^*}(h) = \frac{1}{2}(h_1 + h_2)$$

The set of all points such that $f(x) = 1$ *is called a* ***level curve*** *of* f[5]. *This set contains* x^* *and the equation* $\sqrt{x_1 x_2} = 1$ *implies that elements in this set satisfy:*

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

On figure 2.2, x_1 (x_2) *is the coordinate on the horizontal (vertical) axis. The slope at* $(x_1, x_2) = (1, 1)$ *is* -1 *because the derivative of* $g(x_1) = 1/x_1$ *at* $x_1 = 1$ *is equal to* -1. *Moreover, the coordinates of the gradient of* f *are* $(1/2; 1/2)$. *The arrow on the figure gives the direction of the gradient; it lies on the line* $x_1 = x_2$. *This gradient is orthogonal to the tangent to the level curve. This remark is not a surprise because the level curve is the curve along which the function* $f(x_1, x_2)$ *is constant, equal to 1. Therefore, moving from*

[5]The general definition is the following: a level curve $c \in \mathbb{R}$ of a function f is the set of elements x satisfying $f(x) = c$.

Download free eBooks at bookboon.com

x^* to $x^* + h$ along this curve keeps the differential equal to 0, that is:

$$\frac{\partial f}{\partial x_1}h_1 + \frac{\partial f}{\partial x_2}h_2 = 0$$

Using the inner product, this equation writes:

$$< \nabla f(x^*), h >= 0$$

In the neighborhood of x^*, this relationship means that the gradient and the tangent to the level curve are orthogonal.

If f represents the utility function of an investor, the level curve $f(x_1, x_2) = 1$ defines the pairs of consumed quantities generating the same level (equal to 1) of utility. The set of all these pairs is called an ***indifference curve****.*

At $(1, 1)$, the investor is indifferent if the quantity of one good marginally increases while the quantity of the other marginally decreases. At $y^* = (2, 1)$, the story is different. The differential writes:

$$df_{y^*}(h) = \sqrt{2}h_1 + \frac{1}{\sqrt{2}}h_2$$

For $df_{y^*}(h)$ to be zero and the investor be indifferent to a substitution between the two goods, it is necessary that he obtains twice as much of good 2 than the quantity of good 1 he gives up. This ratio is the well-known ***marginal rate of substitution*** *between the goods****.***

Differentials follow the same rules as derivatives, as can be seen in the following proposition.

Proposition 105 *Let f and g be two C^1-functions, defined on an open set $D \subset \mathbb{R}^p$, and denote x^* an element of D. We have:*

1) $d\,(f+g)_{x^*} = df_{x^*} + dg_{x^*}$

2) $d\,(af)_{x^*} = adf_{x^*}$ for any $a \in \mathbb{R}$

3) $d(fg)_{x^*} = f(x^*)dg_{x^*} + g(x^*)df_{x^*}$

Download free eBooks at bookboon.com

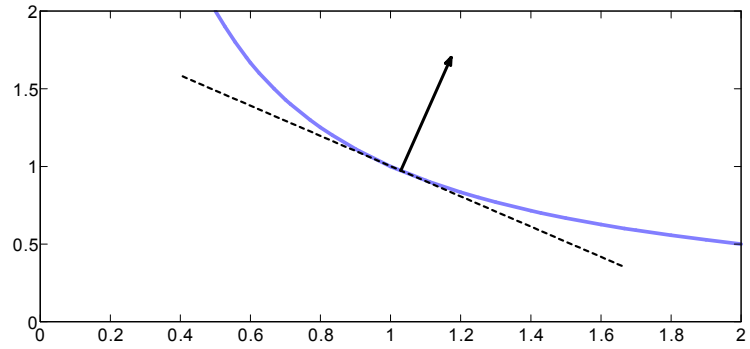


Figure 2.2: Gradient of $f(x_1, x_2) = x_1 x_2$

4) If $g(x^*) \neq 0$, *then* $d\left(\frac{f}{g}\right)_{x^*} = \frac{g(x^*)df_{x^*} - f(x^*)dg_{x^*}}{g(x^*)^2}$

Download free eBooks at bookboon.com

Click on the ad to read more

(1) and (2) are obvious as consequences of the definition of partial derivatives. To prove (3), just write:

$$\begin{aligned} d(fg)_{x^*} &= \sum_{i=1}^{p} \frac{\partial (fg)}{\partial x_i}(x^*)h_i = \sum_{i=1}^{p} \left(g(x^*)\frac{\partial f}{\partial x_i}(x^*) + f(x^*)\frac{\partial g}{\partial x_i}(x^*) \right) h_i \\ &= g(x^*)\sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(x^*)h_i + f(x^*)\sum_{i=1}^{p} \frac{\partial g}{\partial x_i}(x^*)h_i = f(x^*)dg_{x^*} + g(x^*)df_{x^*} \end{aligned}$$

The proof of (4) uses the same method, applying the rules of derivation for ratios of functions.

Alternate notations In most academic papers and economic textbooks, authors do not write $h \rightarrow df_x(h)$, even if it is the right way to understand that df_x is a linear mapping. In most cases, authors write:

$$df(x) = \frac{\partial f}{\partial x_1}dx_1 + ... + \frac{\partial f}{\partial x_p}dx_p$$

This simplified notation means that $dx_1, dx_2, ..., dx_p$ correspond to $h_i, i = 1, ..., p$ and $df(x)$ means $df_x(h)$, the differential of f evaluated at x.

2.2.5 The mean value theorem

In chapter 2 of part I, Rolle's theorem says that if a function g, defined on $[a; b]$, is differentiable on $]a; b[$, there exists $c \in]a; b[$ satisfying $g(b) - g(a) = (b - a)g'(c)$.

A similar result is valid for two-variable functions. However, one needs to be prudent in interpreting the result because of the existence of several partial derivatives.

The following example shows that the intuition is the same as in the single-variable case.

Denote f the function defined by :

Download free eBooks at bookboon.com

$$f(x,y) = x^2 - y^2$$

The graph of f over the square $[1;3] \times [1;3]$ appears on figure 2.3.

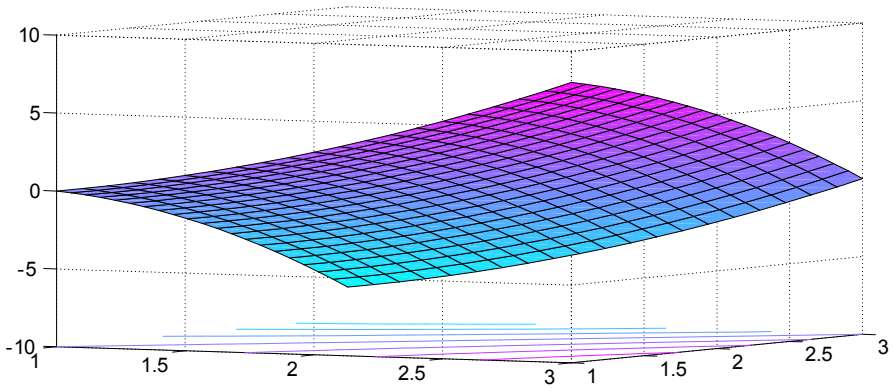


Figure 2.3: The function $f(x,y) = x^2 - y^2$

f satisfies $f(1,1) = 0$ and $f(3,2) = 5$. We can decompose $f(3,2) - f(1,1)$ as follows:

$$f(3,2) - f(1,1) = f(3,2) - f(1,2) + f(1,2) - f(1,1) \tag{2.15}$$

Let $h(x) = f(x,2)$ and $k(y) = f(1,y)$. Equation (2.15) writes:

$$f(3,2) - f(1,1) = h(3) - h(1) + k(2) - k(1)$$

We now apply Rolle's theorem (chapter 2, part I) to the functions h and k. Therefore, there exist $c_1 \in]1;3[$ and $c_2 \in]1;2[$ such that:

$$\begin{aligned} h(3) - h(1) &= (3-1) \times h'(c_1) \\ k(2) - k(1) &= (2-1) \times k'(c_2) \end{aligned}$$

Download free eBooks at bookboon.com

$h'(c_1)$ is the partial derivative of f with respect to x, evaluated at $(c_1; 2)$. $k'(c_2)$ is the partial derivative of f with respect to y, evaluated at $(1, c_2)$.

The mean value theorem hereafter formalizes the idea of the above example.

Proposition 106 *Let f be a C^1-function, defined on $D =]a_1; b_1[\times]a_2; b_2[\subset \mathbb{R}^2$, and $(x_1, y_1), (x_2, y_2)$ be two elements of D.*

There exists $(z_1, z_2) \in D$ such that:

$$f(x_2, y_2) - f(x_1, y_1) = (x_2 - x_1)\frac{\partial f}{\partial x}(z_1, y_2) + (y_2 - y_1)\frac{\partial f}{\partial y}(x_1, z_2)$$

In this proposition, we restrict the domain to a rectangle of $\mathbb{R}^2$. This assumption is not the most general but the key point is that (z_1, z_2) is in D.

2.2.6 Second-order partial derivatives

In the preceding section we showed that a p-variable function has p first-order partial derivatives. The calculation of second-order partial derivatives needs to derive any of the p first-order derivatives with respect to any of the p variables. More precisely, each partial derivative $\frac{\partial f}{\partial x_i}(x), i = 1, 2, ..., p$ can be derived with respect to each variable $x_j, j = 1, 2, ..., p$. As a consequence, the function possesses p^2 second-order partial derivatives. They are organized in a (p, p) matrix called the Hessian matrix or, in short, the Hessian of f.

Definition 107 *Let f a C^1-function, defined on an open set $D \subset \mathbb{R}^p$. The **Hessian matrix** (or **Hessian**) of f at x^*, denoted $H_f(x^*)$, is the (p, p) matrix defined by:*

$$H_f(x^*) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x^*)\right]_{\substack{i=1,...,p \\ j=1,...,p}}$$

where $\frac{\partial^2 f}{\partial x_i \partial x_j}(x^*) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}(x^*)\right)$ is the partial derivative with respect to x_i (when it exists) of the partial derivative of f with respect to x_j.

Download free eBooks at bookboon.com

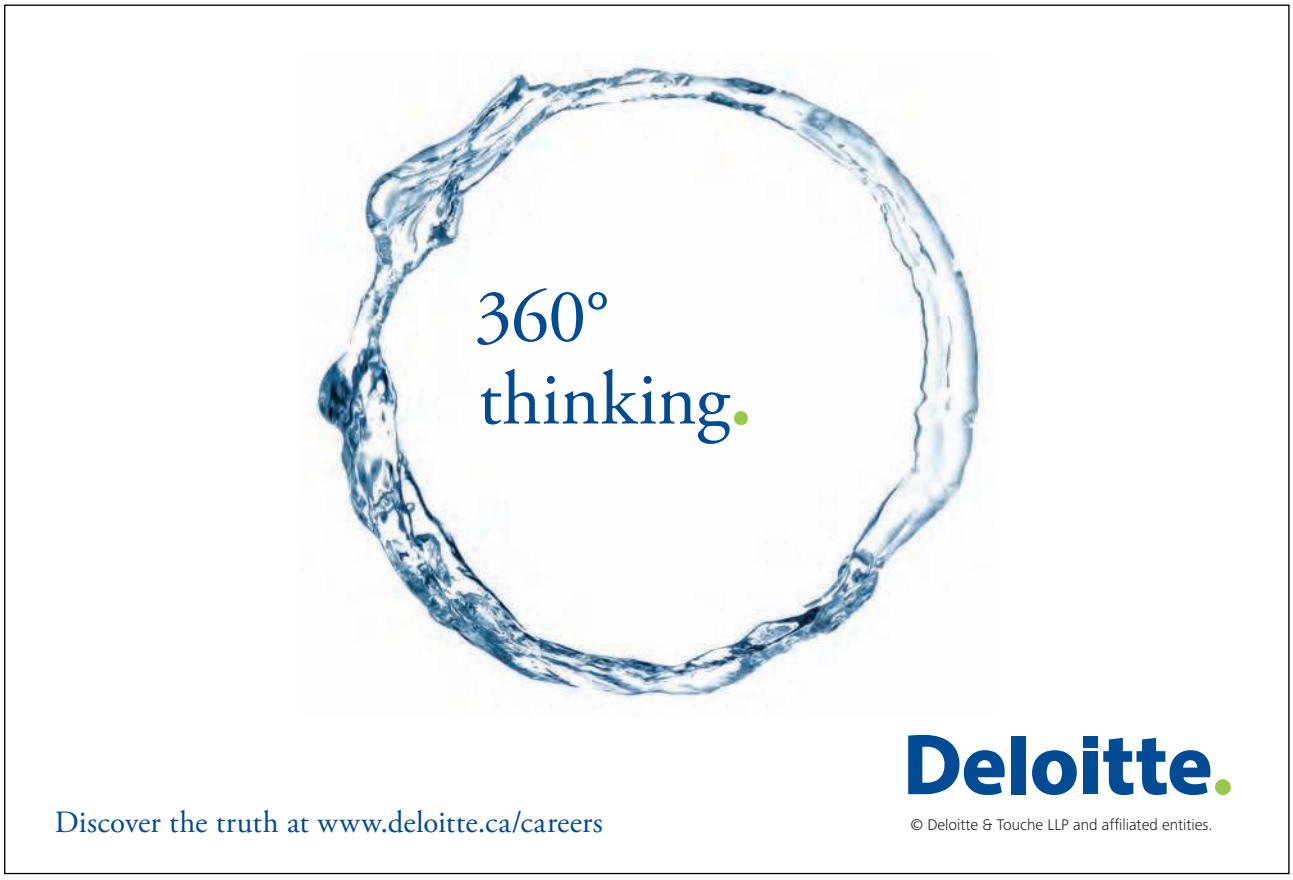
Diagonal elements of $H_f(x^*)$ are denoted:

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

Proposition 108 *If the second-order partial derivatives of a function f are continuous at x^* (f is called a C^2-function), the Hessian matrix is symmetric, that is:*

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

This proposition shows that when calculating the second-order derivatives, the order you choose to derive does not matter, the final result, either $\frac{\partial^2 f}{\partial x_i \partial x_j}(x^*)$ or $\frac{\partial^2 f}{\partial x_j \partial x_i}(x^*)$, is the same. The Hessian matrix is important when solving optimization problems, especially to get sufficient conditions of optimality (chapters 3 and 4). Fortunately, in finance problems, the Hessian matrix is always symmetric.



Download free eBooks at bookboon.com

Click on the ad to read more

Example 109 *Coming back to the function defined in example 97, that is*

$$f(x) = \sqrt{x_1 x_2}$$

we calculated:

$$\begin{aligned}
\frac{\partial f}{\partial x_1}(x^*) &= \frac{1}{2}\sqrt{\frac{x_2^*}{x_1^*}} = \frac{1}{2}\left(x_1^*\right)^{-\frac{1}{2}}\left(x_2^*\right)^{\frac{1}{2}} \\
\frac{\partial f}{\partial x_2}(x^*) &= \frac{1}{2}\sqrt{\frac{x_1^*}{x_2^*}} = \frac{1}{2}\left(x_1^*\right)^{\frac{1}{2}}\left(x_2^*\right)^{-\frac{1}{2}}
\end{aligned}$$

The Hessian matrix is then obtained as follows:

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_1^2}(x^*) &= -\frac{1}{4}\left(x_1^*\right)^{-\frac{3}{2}}\left(x_2^*\right)^{\frac{1}{2}} = -\frac{1}{4x_1^*}\sqrt{\frac{x_2^*}{x_1^*}} \\
\frac{\partial^2 f}{\partial x_1 \partial x_2}(x^*) &= \frac{1}{4}\sqrt{\frac{1}{x_1^* x_2^*}} \\
\frac{\partial^2 f}{\partial x_2^2}(x^*) &= -\frac{1}{4}\left(x_1^*\right)^{\frac{1}{2}}\left(x_2^*\right)^{-\frac{3}{2}} = -\frac{1}{4x_2^*}\sqrt{\frac{x_1^*}{x_2^*}} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1}(x^*) &= \frac{1}{4}\sqrt{\frac{1}{x_1^* x_2^*}}
\end{aligned}$$

In short we write $H_f(x^*)$:

$$H_f(x^*) = \begin{bmatrix} -\frac{1}{4x_1^*}\sqrt{\frac{x_2^*}{x_1^*}} & \frac{1}{4}\sqrt{\frac{1}{x_1^* x_2^*}} \\ \frac{1}{4}\sqrt{\frac{1}{x_1^* x_2^*}} & -\frac{1}{4x_2^*}\sqrt{\frac{x_1^*}{x_2^*}} \end{bmatrix} = \frac{1}{4\sqrt{x_1^* x_2^*}}\begin{bmatrix} -\frac{x_2^*}{x_1^*} & 1 \\ 1 & -\frac{x_1^*}{x_2^*} \end{bmatrix}$$

2.2.7 Taylor's formula

When a function f depends on a single variable, we know (chapter 2, part I) that the graph of f can be approximated by a straight line or by a curve rep-

Download free eBooks at bookboon.com

resenting a polynomial. Taylor's formula allows to calculate the coefficients of this polynomial.

Differentials allow to approximate C^1-functions of several variables at the first order. To approximate functions at higher orders, the right tool is a Taylor's series expansion. We restrict our presentation to second-order approximations because such a choice covers 99.9% of economic and financial models.

Definition 110 *Let β and γ two functions depending on a single variable h. We say that β has the same order of magnitude as γ in the neighborhood of 0, and we write $\beta = O(\gamma)$, if $\lim_{h\to 0} \left|\frac{\beta(h)}{\gamma(h)}\right| < +\infty$. In the same way, β is negligible with respect to γ in the neighborhood of 0 if $\lim_{h\to 0} \frac{\beta(h)}{\gamma(h)} = 0$. In this case, we note $\beta = o(\gamma)$. These two notations O and o are called* ***Landau notations.***

Using Landau notations allows to simplify formulas. $\beta = O(\gamma)$ means that $\beta(h)$ and $\gamma(h)$ are comparable in the following sense. The function β is not infinitely larger (smaller) than the function γ when h tends to 0.

$\beta = o(\gamma)$ means that β is negligible with respect to γ when h tends to 0. If such a situation occurs, that is $\beta = o(\gamma)$, the sum $\beta(h)+\gamma(h)$ is approximated by $\gamma(h)$ because $\beta(h)$ is negligible. Of course this approximation is valid only if h is close to 0.

Proposition 111 ***Taylor's formula***

Let f denote a C^2-function, defined on an open set $D \subset \mathbb{R}^p$, and $(x, x^*) \in D \times D$ such that the line joining x and x^* is in D. We have:

$$\begin{aligned} f(x) &= f(x^*) + \sum_{i=1}^{p} (x_i - x_i^*) \frac{\partial f}{\partial x_i}(x^*) \\ &\quad + \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} (x_i - x_i^*)(x_j - x_j^*) \frac{\partial^2 f}{\partial x_i \partial x_j}(x^*) + o\left(\sum_{i=1}^{p} (x_i - x_i^*)^2 \right) \end{aligned}$$

Download free eBooks at bookboon.com

The expression $o\left(\sum_{i=1}^{n}(x_i - x_i^*)^2\right)$ means that, when the distance between x and x^* tend to zero, all terms of order greater than 2 are negligible with respect to first and second-order terms that appear in the formula as coefficients of the partial derivatives. This Taylor's formula allows to approximate a function of p variables by a second-degree polynomial.

Matrix notation Using the Hessian matrix and the gradient of f shortens the above formula as follows:

$$f(x) = f(x^*) + < \nabla f(x^*), x - x^* > + \frac{1}{2}(x - x^*)^T H_f(x^*)(x - x^*) + o(\|x - x^*\|^2)$$

This alternative formulation is based on notions presented in chapter 4 of the part I and in chapter 1 of the present book, (inner product of vectors or product of matrices).

Example 112 *Let f be a function defined on $\mathbb{R}^2$ taking values in $\mathbb{R}_+$ and defined by :*

$$f(x, y) = \exp\left(-\frac{1}{2}\left(x^2 + y^2\right)\right)$$

The partial derivatives of f are equal to:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= -x\exp\left(-\frac{1}{2}\left(x^2 + y^2\right)\right); \frac{\partial f}{\partial y} = -y\exp\left(-\frac{1}{2}\left(x^2 + y^2\right)\right) \\
\frac{\partial^2 f}{\partial x^2} &= \left(x^2 - 1\right)\exp\left(-\frac{1}{2}\left(x^2 + y^2\right)\right); \frac{\partial^2 f}{\partial y^2} = \left(y^2 - 1\right)\exp\left(-\frac{1}{2}\left(x^2 + y^2\right)\right) \\
\frac{\partial^2 f}{\partial x \partial y} &= xy\exp\left(-\frac{1}{2}\left(x^2 + y^2\right)\right)
\end{aligned}$$

Applying Taylor's formula at $(0, 0)$ leads to write:

$$f(h_1, h_2) = 1 - \frac{1}{2}\left[h_1^2 + h_2^2\right] + o\left(h_1^2 + h_2^2\right)$$

Download free eBooks at bookboon.com

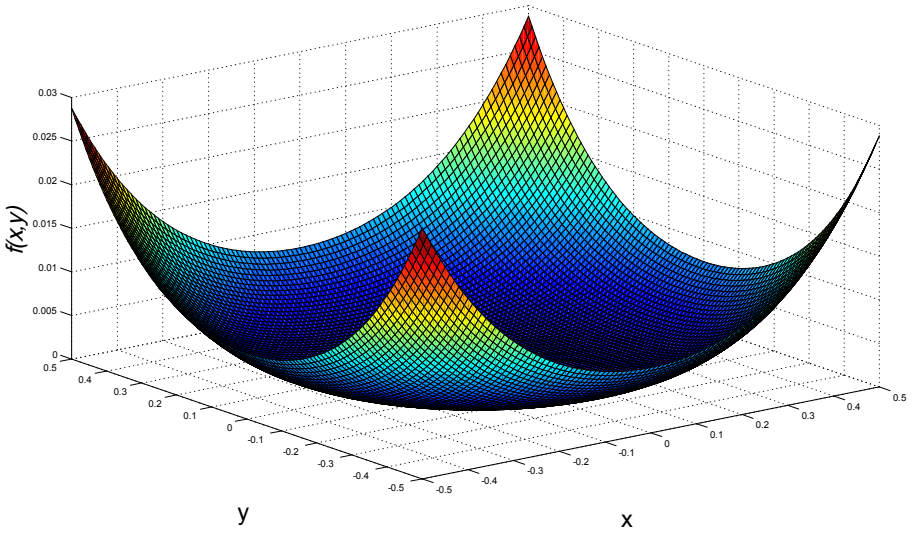


Figure 2.4: Approximation error

Figure 2.4 shows the difference between f and the second-degree polynomial when h_1 and h_2 move between $-0,5$ and $0,5$. This difference is represented by $o\left(h_1^2+h_2^2\right)$. We observe that, even "far" from (0,0), the approximation is quite good. The error is not larger than 0.03 with the function being valued 1 at 0. Of course, this case is specific; choosing a more complicated function could lead to approximations of lower quality.

2.2.8 Convex and concave functions

We presented in the chapter 2 of part I the definition of a convex single-variable function f defined on an interval $I \subset \mathbb{R}$. You will see in the following that the definition is almost the same when f depends on p variables, except for the domain of definition D. Of course, D must be included in $\mathbb{R}^p$ but we have to be sure that the definition of a convex function is meaningful. It is the reason why we first introduce convex sets in a vector space like $\mathbb{R}^p$.

Download free eBooks at bookboon.com

Definition 113 *Let C be a subset of $\mathbb{R}^p$. C is* ***convex*** *if:*

$$\forall (x, y) \in C \times C, \forall \alpha \in [0; 1], \alpha x + (1 - \alpha) y \in C$$

The geometric interpretation of this definition is simple If any two elements x and y are in the same convex set, all the segment joining x and y is also included in C. Remark in passing that if $p = 1$, C is an interval.

$\mathbb{R}^p$ being a vector space, the combination $\alpha x + (1 - \alpha) y$ in definition 113 is a linear combination of the vectors x and y. This linear combination has two specific features; the coefficients α and $1 - \alpha$ are positive and their sum equals 1. Such a combination of vectors is called a **convex combination**.

SIMPLY CLEVER
ŠKODA

Do you like cars? Would you like to be a part of a successful brand?
We will appreciate and reward both your enthusiasm and talent.
Send us your CV. You will be surprised where it can take you.
Send us your CV on
www.employerforlife.com
Download free eBooks at bookboon.com
Click on the ad to read more

We can now rigorously define convex and concave functions.

Definition 114 *1) Let f be a function defined on a convex domain $D \subset \mathbb{R}^p$. f is a **convex function** on D if, for any $\alpha \in [0;1]$ and any couple $(x,y) \in D \times D$, we have:*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

*2) Under the same assumptions on D, f is a **concave function** on D if the inequality is reversed.*

*3) The function f is **strictly convex (concave)** if the inequality is strict in part 1 (2) of the definition.*

Assuming convex or concave functions is very common in finance or economic models. Utility functions are usually concave and cost functions are convex. These assumptions make easier solving optimization problems. These issues are addressed in chapters 3 and 4.

In chapter 2 of part I, we characterized convex (concave) functions by positive (negative) second-order derivatives. For a function f depending on p variables, the corresponding result uses second-order partial derivatives, by means of a condition on the Hessian matrix of f.

Proposition 115 *Let f be a C^2-function defined on a convex open domain $D \subset \mathbb{R}^p$.*

1) f is convex (concave) on D if and only if the Hessian matrix $H_f(x)$ is positive (negative) semi-definite at any $x \in D$.

2) If $H_f(x)$ is positive (negative) definite, f is strictly convex (concave).

Recall that positive semi-definite matrices have been presented in chapter 4 of part I (definition 52).

Download free eBooks at bookboon.com

Example 116 *Consider a two-goods economy; an agent is characterized by the following utility function U defined on $D = \mathbb{R}_+^* \times \mathbb{R}_+^*$:*

$$U(x) = \ln(x_1 x_2)$$

where $x = (x_1, x_2)$ is the vector of consumed quantities and $U(x)$ measures the welfare generated by consumption. U is strictly concave on D. In fact, we have:

$$\frac{\partial U}{\partial x_1} = \frac{1}{x_1} \qquad \frac{\partial U}{\partial x_2} = \frac{1}{x_2}$$
$$\frac{\partial^2 U}{\partial x_1^2} = -\frac{1}{x_1^2} \qquad \frac{\partial^2 U}{\partial x_2^2} = -\frac{1}{x_2^2}$$
$$\frac{\partial^2 U}{\partial x_1 \partial x_2} = 0$$

It follows:

$$H_U(x) = \begin{pmatrix} -\frac{1}{x_1^2} & 0 \\ 0 & -\frac{1}{x_2^2} \end{pmatrix}$$

We can check that $H_U(x)$ is negative definite by computing $y^T H_U(x) y$, where y is a non zero vector in $\mathbb{R}^2$.

$$y^T H_U(x) = (y_1, y_2) \begin{pmatrix} -\frac{1}{x_1^2} & 0 \\ 0 & -\frac{1}{x_2^2} \end{pmatrix} = \left(-\frac{y_1}{x_1^2}, -\frac{y_2}{x_2^2} \right)$$
$$y^T H_U(x) y = \left(-\frac{y_1}{x_1^2}, -\frac{y_2}{x_2^2} \right) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = -\frac{y_1^2}{x_1^2} - \frac{y_2^2}{x_2^2}$$

We get $y^T H_U(x) y < 0$ showing that U is strictly concave. The interpretation of the concavity of U is the same as the one provided for single-variable functions. The utility obtained by consuming one more unit of a given good decreases with the quantity already consumed.

Download free eBooks at bookboon.com

2.3 Implicit and homogeneous functions

2.3.1 The implicit function theorem

Several economic variables are often linked by complex relationships so that it is impossible to express these relationships explicitly. The most well-known example of such a relationship is the definition of the internal rate of return or, equivalently, of the yield of a coupon-bearing bond. The yield r of a bond is linked to the price p and to the future payoffs $F_1, ..., F_T$ of the bond (coupons plus reimbursment price). Though economically intuitive, it is impossible to express r as an explicit function of the variables $(p, F_1, ..., F_T)$. In the same spirit, the utility provided by the consumption of a bundle of goods $(x_1, ..., x_p)$ is measured by a utility function $U(x_1, ..., x_p)$ taking values in $\mathbb{R}$. For a given utility level u, the equation $U(x_1, ..., x_p) = u$ creates a relationship between x_1 and the $p-1$ other variables. In general, no explicit formulation exists for this relationship.

In this section, we develop some results allowing to measure the sensitivity of a given variable with respect to variations in other variables. We start by the most simple case where a function F only depends on two variables.

Definition 117 *Let F be a function defined on an open subset $D \subset \mathbb{R}^2$. The equation $F(x, y) = 0$ defines an implicit function if there exists a function $g(x) = y$, defined on an interval and taking values in an interval such that $F(x, g(x)) = 0$.*

Of course the relationship between x and y is said implicit when g cannot be defined explicitly.

Proposition 118 ***Implicit function theorem (2 variables)***

If $F : \mathbb{R}^2 \to \mathbb{R}$ is C^1 and defines an implicit function g by means of the relationship $F(x, y) = 0$, we have :

$$g'(x) = \frac{\partial y}{\partial x} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}$$

Download free eBooks at bookboon.com

The notation $\frac{\partial y}{\partial x}$ may seem surprising: it is not really a rigorous way to write such a derivative but this formulation is common. It is the reason why we use this expression. In the same spirit, the notations $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ do not look precise enough because we do not specify the values at which the partial derivatives are calculated. But these notations are commonly used as long as they do not introduce confusion or ambiguity. In our example, we know that the partial derivatives are calculated at (x, y) such that $F(x, y) = 0$.

Theorem 118 is useful to perform comparative statics. Being given the value of a function of two variables (production function, utility function, net present value, etc.), comparative statics tests the impact of the variation of one variable on the value of the other variable.

Download free eBooks at bookboon.com

Click on the ad to read more

Example 119 ***The internal rate of return***

An investment project needs an initial outflow followed (in general) by inflows at future dates. Let F_0 denotes the initial negative cashflow and $F_t, t = 1, ..., T$ the future positive cashflows, where T denotes the maturity date of the project. The ***net present value*** *of this project, discounted at a rate r, is defined by:*

$$NPV(r) = \sum_{t=0}^{T} \frac{F_t}{(1+r)^t}$$

The ***internal rate of return*** *(IRR) is the rate r^* satisfying $NPV(r^*) = 0$.*

This equation defines an implicit relationship between the discount rate r^* and any given cashflow of the project. This relationship is called an implicit function because you cannot write r^* as follows:

$$r^* = f(T, F_t, t = 0, ..., T)$$

But the implicit function theorem allows to calculate the sensitivity of r^* with respect to variations in any given cashflow. For example:

$$\frac{\partial r}{\partial F_0} = -\frac{\frac{\partial NPV}{\partial F_0}}{\frac{\partial NPV}{\partial r}} = -\frac{1}{\frac{\partial NPV}{\partial r}} \tag{2.16}$$

We can also write:

$$\frac{\partial NPV}{\partial r} = -\sum_{t=0}^{T} \frac{tF_t}{(1+r)^{t+1}} \tag{2.17}$$

Equations (2.16) and (2.17) lead to:

$$\frac{\partial r}{\partial F_0} = \frac{1}{\sum_{t=0}^{T} tF_t(1+r)^{-t-1}}$$

A too superficial look at this formula could let the reader think that the IRR

Download free eBooks at bookboon.com

is an increasing function of the initial cost of the project because the derivative is positive. But remember that $F_0 < 0$. *As a consequence, a marginal increase in* F_0 *is in fact a marginal decrease of the cost of the project, everything else equal. Proposition 118 can be generalized to p-variable functions almost without modifications. Any equation* $F(x) = 0$ *where* $x = (x_1, ..., x_p)$ *defines an implicit function between components* x_j *and* x_k *for any* (j, k) *in* $\{1, ..., p\}^2$.

Proposition 120 ***Implicit function theorem (p variables)***

Let F *be a* C^1*-function, defined on an open set* $D \subset \mathbb{R}^p$. *Assume that* F *defines an implicit function linking* x_j *and* x_k *by means of the equation* $F(x) = 0$. *It follows that:*

$$\frac{\partial x_k}{\partial x_j} = -\frac{\frac{\partial F}{\partial x_j}}{\frac{\partial F}{\partial x_k}}$$

x_k is the k-th variable and x_j the j-th variable. This proposition is not very different from proposition 118 because the other variables do not play any role. It is as if we were dealing with a two-variable function, the $(p-2)$ others being kept constant.

2.3.2 Homogeneous functions and Euler theorem

Homogeneous functions are common in economics. The most well known example is the production function of a firm. When all production factors are doubled, the usual assumption is to consider that production will double. In such a situation, the function is said homogeneous of degree 1. A second example in finance is the price of an option contract when considered as a function of two variables, the strike price and the underlying price. When you double the two, the price of the option doubles. The definition below is the generalization of this intuitive example.

Definition 121 *Let* f *be a function defined on a set* $D \subset \mathbb{R}^p$, *taking values in* $\mathbb{R}$, *and let* D^* *denote a subset de* D. f *is* ***homogeneous*** *of degree* α *on*

Download free eBooks at bookboon.com

D^* if:

$$\forall x \in D^*, \forall \lambda \in \mathbb{R}_+, \lambda x \in D \text{ and } f(\lambda x) = \lambda^\alpha f(x)$$

When a function is homogeneous of degree $\alpha > 1$, doubling the inputs more than doubles the output. For production functions, it is the sign of economies of scale. The unit cost of production decreases when produced quantities increase.

A homogeneous function of degree 1 satisfies $f(2x) = 2f(x)$; this equality is also true for a linear function. However, if linear forms are homogeneous of degree 1 the reciprocal is false. The function $f(x_1, x_2) = \sqrt{x_1 x_2}$ is a simple counterexample. Of course, f is not linear but is homogeneous of degree 1 because $f(2x_1, 2x_2) = \sqrt{2x_1 \times 2x_1} = 2\sqrt{x_1 x_2} = 2f(x_1, x_2)$.

The following proposition shows that a homogeneous function has homogeneous partial derivatives. Only the degree of homogeneity changes.

Proposition 122 *Let f be a C^1 function defined on an open set $D \subset \mathbb{R}^p$, homogeneous of degree α on $D^* \subset D$. The functions $\partial f / \partial x_i$ are homogeneous of degree $\alpha - 1$ on D^*.*

Proof. We can write:

$$f(\lambda x) = \lambda^\alpha f(x) \Rightarrow \frac{\partial}{\partial x_i}\left[f(\lambda x)\right] = \frac{\partial}{\partial x_i}\left[\lambda^\alpha f(x)\right] \tag{2.18}$$

Let h be defined by $x \to f(\lambda x)$. h writes $f \circ g$ where $g(x) = \lambda x$. Consequently:

$$\frac{\partial}{\partial x_i}\left[f(\lambda x)\right] = \lambda \frac{\partial f}{\partial x_i}(\lambda x)$$

The linearity of derivations leads to:

$$\frac{\partial}{\partial x_i}\left[\lambda^\alpha f(x)\right] = \lambda^\alpha \frac{\partial f}{\partial x_i}(x)$$

It implies:

$$\frac{\partial f}{\partial x_i}(\lambda x) = \lambda^{\alpha-1} \frac{\partial f}{\partial x_i}(x)$$

Download free eBooks at bookboon.com

This equality shows that $\frac{\partial f}{\partial x_i}$ is homogeneous of degree $\alpha - 1$. ■

When functions are homogeneous of degree 1 (for example $f(x_1, x_2) = \sqrt{x_1 x_2}$), the proposition means that partial derivatives are homogeneous of degree 0. In fact we have:

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

Multiplying x_1 and x_2 by a non-zero number does not change the value of $\frac{\partial f}{\partial x_1}$. If f is the utility function of an investor, the proposition shows that the marginal utility provided by the consumption of a marginal quantity of good 1 is the same when the quantity already consumed is (x_1, x_2) or when it is $(\lambda x_1, \lambda x_2)$ with $\lambda > 0$. Geometrically, this result is not surprising because, along the line $x_1 = x_2$, f is a linear function. In fact, $f(x, x) = x$ for any x.

The specific features of homogeneous functions lead to the Euler theorem that links the value of a homogeneous function at a given point to the values of its partial derivatives at this same point.

Proposition 123 ***The Euler theorem***

Let f be a C^1 function defined on $\left(\mathbb{R}_+^*\right)^p$, homogeneous of degree α. At any $x \in \left(\mathbb{R}_+^*\right)^p$, we have:

$$\sum_{i=1}^{p} x_i \frac{\partial f}{\partial x_i}(x) = \alpha f(x)$$

We only provide hereafter a sketch of the proof. By definition of homogeneity we know that:

$$f(\lambda x) = \lambda^\alpha f(x) \tag{2.19}$$

Each side of equation 2.19 is a function of λ (that is the point!). The derivatives of the two sides with respect to λ must be equal. Assume that h is a small real number such that:

$$f((\lambda + h)\, x) \simeq f(\lambda x) + \sum_{i=1}^{p} h x_i \frac{\partial f}{\partial x_i}(\lambda x)$$

Download free eBooks at bookboon.com

We neglect the second-order terms whose order of magnitude is h^2 because they are negligible in the following limit.

$$\frac{\partial f}{\partial \lambda}(\lambda x) = \lim_{h \to 0} \frac{f((\lambda + h)\, x) - f(\lambda x)}{h} \simeq \sum_{i=1}^{p} x_i \frac{\partial f}{\partial x_i}(\lambda x) = \sum_{i=1}^{p} x_i \lambda^{\alpha - 1} \frac{\partial f}{\partial x_i}(x)$$

The last equality is obtained because $\frac{\partial f}{\partial x_i}$ is homogeneous of order $\alpha - 1$ by proposition 122.

The derivative of the right-hand side of equation (2.19) writes:

$$\alpha \lambda^{\alpha - 1} f(x)$$

As a consequence we obtain:

$$\sum_{i=1}^{p} x_i \lambda^{\alpha - 1} \frac{\partial f}{\partial x_i}(x) = \alpha \lambda^{\alpha - 1} f(x)$$

Simplifying by $\lambda^{\alpha - 1}$ leads to the result:

$$\sum_{i=1}^{p} x_i \frac{\partial f}{\partial x_i}(x) = \alpha f(x)$$

Download free eBooks at bookboon.com

Click on the ad to read more