QE Optimization, WS 2016/17

Part 2. Differential Calculus for Functions of n Variables

(about 5 Lectures)

Supporting Literature: ***Angel de la Fuente***, **"Mathematical Methods and Models for Economists",** Chapter 2

C. Simon, L. Blume, **"Mathematics for Economists"**

Contents

In this chapter we consider functions

$$f\colon \mathbb{R}^n \to \mathbb{R}$$

of $n \geq 1$ variables (**multivariate** functions). Such functions are the basic building block of formal economic models.

2 Differential Calculus for Functions of n Variables

2.1 Partial Derivatives

Everywhere below: $U \subseteq \mathbb{R}^n$ will be an **open** set in the space $(\mathbb{R}^n, \|\cdot\|)$ (with the Euclidean norm $\|\cdot\|$) and $f\colon U \to \mathbb{R}$,

$$U \ni (x_1, \ldots, x_n) \to f(x_1, \ldots, x_n) \in \mathbb{R}.$$

Definition 2.1.1. *The function f is **partially differentiable** with respect to the **i-th coordinate** (or variable) x_i, at a given point $x \in U$, if the following limit exists*

$$\begin{aligned} D_i f(x) :&= \lim_{h\to 0} \frac{f(x+he_i) - f(x)}{h} \\ &= \lim_{h\to 0} \frac{1}{h} \left[f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_n) \right], \end{aligned}$$

where $e_i := \{\underbrace{0, \ldots, 0}_{i-1}, 1, 0, \ldots, 0\}$ is the basis vector in $\mathbb{R}^n$.

Since U is *open*, there exists an *open ball* $B_\varepsilon(x) \subseteq U$. In the definition of $\mathop{lim}\limits_{h\to 0}$ one considers only "small" h with $|h| < \varepsilon$.

$D_i f(x)$ is called the ***i*-th partial derivative** of f at point x.

Notation: We also write $D_{x_i} f(x)$, $\partial_i f(x)$, $\partial f(x)/\partial x_i$.

The partial derivative $D_i f(x)$ can be interpreted as a usual derivative w.r.t. the i-th coordinate, whereby all the other $n-1$ coordinates are kept fixed. Namely, in the $\varepsilon-$neighbourhood of x_i, let us define a function

$$(x_i - \varepsilon, x_i + \varepsilon) \ni \xi \to g_i(\xi) := f(x_{1,i-1}, \xi, x_{i+1}, \ldots, x_n).$$

Then by Definition 2.1.1,

$$D_i f(x) := \lim_{h\to 0} \frac{g_i(x_i + h) - g_i(x_i)}{h} = g_i'(x_i).$$

Definition 2.1.2. *A function $f : U \to \mathbb{R}$ is called **partially differentiable** if $D_i f(x)$ exists for all $x \in U$ and all $1 \le i \le n$. Furthermore, f is called **continuously partially differentiable**, if all partial derivatives*

$$D_i f : U \to \mathbb{R},\ 1 \le i \le n$$

are continuous functions.

Example 2.1.3.

(i) ***Distance function***

$$r(x) := |x| = \sqrt{x_1^2 + \ldots + x_n^2}, \quad x \in \mathbb{R}^n$$

Let us show that $r(x)$ *is partially differentiable at all points* $x \in \mathbb{R}^n \backslash \{0\}$.

$$\xi \to g_i(\xi) := \sqrt{x_1^2 + \ldots + \xi^2 + \ldots + x_n^2} \in \mathbb{R}.$$

Use the chain rule for the derivatives of real-valued functions (cf. standard courses in Calculus) $\Longrightarrow$

$$\frac{\partial r}{\partial x_i}(x) = \frac{1}{2} \frac{2x_i}{\sqrt{x_1^2 + \ldots + \xi^2 +_n^2}} = \frac{x_i}{r(x)}.$$

Generalization: *Let* $f : \mathbb{R}_+ \to \mathbb{R}$ *be differentiable, then* $\mathbb{R}^n \ni x \to f(r(x))$ *is partially differentiable at all points* $x \in \mathbb{R}^n \backslash \{0\}$ *and*

$$\frac{\partial}{\partial x_i} f(r) = f'(r) \cdot \frac{\partial r}{\partial x_i} = f'(r) \cdot \frac{x_i}{r}.$$

(ii) ***Cobb–Douglas production function with n inputs***

$$f(x) := x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} \textit{ for } \alpha_i > 0, \quad 1 \le i \le n,$$

defined on

$$U := \{(x_1, \ldots, x_n) |\ x_i > 0,\ 1 \le i \le n\}.$$

Calculate the so-called ***marginal-product function*** *of input* i

$$\frac{\partial f}{\partial x_i}(x) = \alpha_i x_1^{\alpha_1} {}_i^{\alpha_i - 1} \ldots x_n^{\alpha_n} = \alpha_i \frac{f(x)}{x_i}.$$

Mathematicians will say: multiplicative functions with separable variables, polynomials. Economists are especially interested in the case $\alpha_i \in (0, 1)$.

This is an example of **homogeneous** *functions of order (degree)* $a = \alpha_1 + \ldots + \alpha_n$, *which means*

$$f(\lambda x) = \lambda^a f(x),\ \forall \lambda > 0,\ x \in U.$$

Moreover, the Cobb–Douglas function is $\log -$*linear:*

$$\log f(x) = \alpha_1 \log x_1 + \ldots + \alpha_n \log x_n.$$

(iii) ***Quasilinear utility function:***

$$f(m,x) := m + u(x)$$

with $m \in \mathbb{R}_+$ *(i.e.,* $m \geq 0$*) and some* $u : \mathbb{R} \to \mathbb{R}$*.*

$$\frac{\partial f}{\partial m} = 1, \ \frac{\partial f}{\partial x} = u'(x).$$

(iv) ***Constant elasticity of substitution (CES) production function with n inputs, which describes aggregate consumption for n types of goods.***

$$f(x_1, \ldots, x_n) := \left(\delta_1 x_1^{\alpha} + \ldots + \delta_n x_n^{\alpha}\right)^{1/\alpha},$$
$$\text{with } \alpha > 0, \quad \delta_i > 0 \text{ and } \sum\nolimits_{1 \leq i \leq n} \delta_i = 1,$$

defined on the open domain

$$U := \{(x_1, \ldots, x_n) \ \mid \ x_i > 0, \ 1 \leq i \leq n\}.$$

We calculate the ***marginal-product function***

$$\begin{aligned} \frac{\partial f}{\partial x_i}(x) &= \frac{1}{\alpha}\left(\delta_1 x_1^{\alpha} + \ldots + \delta_n x_n^{\alpha}\right)^{\frac{1}{\alpha}-1} \cdot \alpha \delta_i x_i^{\alpha-1} \\ &= \delta_i x_i^{\alpha-1}\left(\delta_1 x_1^{\alpha} + \ldots + \delta_n x_n^{\alpha}\right)^{\frac{1-\alpha}{\alpha}}. \end{aligned}$$

Note that f *is homogeneous :* $f(\lambda x) = \lambda f(x)$*.*

Definition 2.1.4. *Let* $U \subseteq \mathbb{R}^n$ *be open and* $f : U \to \mathbb{R}$ *be partially differentiable. Then, the vector*

$$\nabla f(x) := \mathrm{grad} f(x) := \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right) \in \mathbb{R}^n$$

is called the ***gradient*** *of* f *at point* $x \in U$*.*

Example 2.1.5.

(i) *Distance function* $r(x)$

$$\mathrm{grad} r(x) = \frac{x}{r(x)} \in \mathbb{R}^n, \quad x \in U := \mathbb{R}^n \backslash \{0\}.$$

(ii) *Let* $f, g : U \to \mathbb{R}$ *be partially differentiable. Then*

$$\nabla(f \cdot g) = f \cdot \nabla g + g \cdot \nabla f.$$

Proof. This follows from the product rule

$$\frac{\partial}{\partial x_i}(fg) = f\frac{\partial g}{\partial x_i} + g\frac{\partial f}{\partial x_i}.$$

□

2.2 Directional Derivatives

Fix a directional vector $v \in \mathbb{R}^n$ with $|v| = 1$ (of **unit length**!).

Definition 2.2.1. *The* ***directional derivative*** *of $f : U \to \mathbb{R}$ at a point $x \in U$ along the unit vector $v \in \mathbb{R}^n$ (i.e., with $|v| = 1$) is given by*

$$\partial_v f(x) := D_v f(x) := \lim_{h \to 0} \frac{f(x + hv) - f(x)}{h}.$$

Remark 2.2.2.

(i) Define a new function

$$h \to g_v(h) := f(x + hv).$$

If $g_v(h)$ is differentiable at $h = 0$, then $f(x)$ is differentiable at point $x \in U$ along direction v and

$$D_v f(x) = g_v'(0).$$

(ii) From the above definitions it is clear that the partial derivatives = directional derivatives along the basis vectors $e_i, 1 \leq i \leq n$,

$$\frac{\partial f}{\partial x_i}(x) = D_{e_i} f(x),\ 1 \leq i \leq n.$$

Example 2.2.3. *Consider the "saddle" function in $\mathbb{R}^2$*

$$f(x_1, x_2) := -x_1^2 + x_2^2,$$

and find $D_v f(x)$ along the direction $v := \left(\sqrt{2}/2, \sqrt{2}/2\right), |v| = 1$. Define

$$\begin{aligned} g_v(h) :&= -\left(x_1 + h\sqrt{2}/2\right)^2 + \left(x_2 + h\sqrt{2}/2\right)^2 \\ &= -x_1^2 + x_2^2 + \sqrt{2}h(x_2 - x_1). \end{aligned}$$

Then $D_v f(x) = g_v'(0) = \sqrt{2}(x_2 - x_1)$. Note that $D_v f(x) = 0$ if $x_1 = x_2$. The function f has its minimum at the diagonal $x_1 = x_2$.

Relation between $\nabla f(x)$ and $D_v f(x)$:

$$D_v f(x) = \langle \nabla f(x), v \rangle_{\mathbb{R}^n} = \sum_{i=1}^{n} \partial_i f(x) \cdot v_i. \qquad (*)$$

Proof. will be done later, as soon as we prove the chain rule for ∇f. □

2.3 Higher Order Partials

Let $f : U \to \mathbb{R}$ be partially differentiable, i.e.,

$$\exists \frac{\partial}{\partial x_i} f : U \to \mathbb{R} \ 1 \leq i \leq n.$$

Analogously, for $1 \leq j \leq n$ we can define (if it exists)

$$\frac{\partial}{\partial x_j}\left(\frac{\partial}{\partial x_i} f\right) : U \to \mathbb{R}.$$

Notation:

$$\frac{\partial^2 f}{\partial x_j \partial x_i} \text{ or } \frac{\partial^2 f}{\partial x_i^2} \quad \text{if } i = j.$$

Warning: In general,

$$\frac{\partial^2 f}{\partial x_j \partial x_i} \neq \frac{\partial^2 f}{\partial x_j \partial x_i} \quad \text{if } i \neq j.$$

Theorem 2.3.1 ((A. Schwarz); also known as Young's theorem)**.** *Let $U \subseteq \mathbb{R}^n$ be open and $f : U \to \mathbb{R}$ be **twice continuously differentiable**, $f \in C^2(U)$, (i.e., all derivatives $\frac{\partial^2 f}{\partial x_j \partial x_i}$, $1 \leq i, j \leq n$, are continuous). Then for all $x \in U$ and $1 \leq i, j \leq n$*

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

i.e., for cross-partial derivatives, the order of differentiation in their computing is irrelevant.

Example: (i) The above theorem works:

$$f(x_1, x_2) := x_1^2 + bx_1x_2 + x_2^2, \quad (x_1, x_2) \in \mathbb{R}^2.$$

Counterexample: **(ii)** The above theorem does not work:

$$f(x_1, x_2) := \begin{cases} x_1x_2\dfrac{x_1^2 - x_2^2}{x_1^2 + x_2^2}, & (x_1, x_2) \in \mathbb{R}^2 \setminus \{(0,0)\} \\ 0, & (x_1, x_2) = (0,0). \end{cases}$$

We calculate

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}(0,0) = -1 \neq 1 = \frac{\partial^2 f}{\partial x_2 \partial x_1}(0,0).$$

Reason: $f \notin C^2(U)$.

Notation:

$$D_{i_k i_1} f, \ \frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_1}},$$
for any $i_1, \dots, i_k \in \{1, \dots n\}$.

In general, for any $v \in \mathbb{R}^n$ with $|v| = 1$, we have by $(*)$

$$|D_v f(x)| \leq |\nabla f(x)|_{\mathbb{R}^n}.$$

Geometrical interpretation of ∇f: Define the normalized vector

$$v := \frac{\nabla f(x)}{|\nabla f(x)|_{\mathbb{R}^n}} \in \mathbb{R}^n.$$

Then, for this v

$$D_v f(x) = \langle \nabla f(x), v \rangle_{\mathbb{R}^n} = |\nabla f(x)|_{\mathbb{R}^n}.$$

In other words, the gradient $\nabla f(x)$ of f at point x is the direction in which the slope of f is the **largest** in absolute value.

2.4 Total Differentiability

Intuition: Repetition of the **1-dim case**

Definition 2.4.1. *A function $g : \mathbb{R} \to \mathbb{R}$ is differentiable at point $x \in \mathbb{R}$ if the following limit exists*

$$\lim_{h \to 0} \frac{g(x+h) - g(x)}{h} =: g'(x) \in \mathbb{R}. \tag{$*$}$$

Geometrical picture: Locally, i.e., for small $|h| \to 0$, we can approximate the values of $g(x+h)$ by the linear function $g(x) + ah$ with $a := g'(x) \in \mathbb{R}$. Indeed, the limit $(*)$ can be rewritten as

$$\lim_{h \to 0} \frac{g(x+h) - [g(x) + ah]}{h} = 0.$$

The approximation error $E_g(h)$ equals

$$E_g(h) := g(x+h) - [g(x) + ah] \in \mathbb{R}$$

and it goes to zero with h:

$$\lim_{h \to 0} \frac{E_g(h)}{h} = 0 \quad \text{i.e.,} \quad \lim_{h \to 0} \frac{|E_g(h)|}{|h|} = 0.$$

The latter can be written as

$$E_g(h) = o(h) \quad \text{as} \quad h \to 0,$$
$$g(x+h) \sim g(x) + ah \quad \text{as} \quad h \to 0.$$

Summary: $g : \mathbb{R} \to \mathbb{R}$ is **differentiable** at $x \in \mathbb{R}$ if, for points $x+h$ sufficiently close to x, the values $g(x+h)$ admit a "nice" approximation by a linear function $g(x) + ah$, with an error

$$E_g(h) := g(x+h) - g(x) - ah$$

that goes to zero "faster" than h itself, i.e.,

$$\lim_{h \to 0} \frac{|E_g(h)|}{|h|} = 0.$$

Now we extend the notion of differentiability to functions $f : \mathbb{R}^n \to \mathbb{R}^m$, for arbitrary $n, m \geq 1$:

$$\mathbb{R}^n \ni x = (x_1, \dots, x_n) \to f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix} \in \mathbb{R}^m.$$

Definition 2.4.2. *Let $U \subset \mathbb{R}^n$ be open, and let $f : U \to \mathbb{R}^m$. The function f is **(totally) differentiable** at a point $x \in U$ if there exists a **linear mapping***

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

such that in some neighbourhood of x, (i.e., for small enough $h \in \mathbb{R}^n$ with $|h| < \varepsilon$), there is a presentation

$$f(x+h) = f(x) + Ah + E_f(h), \tag{$**$}$$

where the error term

$$E_f(h) := f(x+h) - f(x) - Ah \in \mathbb{R}^m$$

obeys

$$\lim_{h \to 0} \frac{\|E_f(h)\|_{\mathbb{R}^m}}{\|h\|_{\mathbb{R}^n}} = 0.$$

*The **derivative** $Df(x)$ of f at point x is the matrix A.*

Remark 2.4.3.

*(i) Each linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by the $m \times n$**matrix** (with m rows and n columns)*

$$(a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

which describes the action of the linear map A on the ***canonical basis*** $(e_j)_{1\le j\le n}$ *in* $\mathbb{R}^n$, $e_j = (\underbrace{0,\ldots,0,1}_{j},0,\ldots,0)^{\mathrm{t}}$ *(vertical column or $n\times 1$matrix),*

$$Ae_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{R}^m, \quad 1 \le j \le n.$$

Below we always ***identify*** *the linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ with this matrix, which acts as*

$$Ah := \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} a_{11}h_1 + \ldots + a_{1n}h_n \\ a_{21}h_1 + \ldots + a_{2n}h_n \\ \vdots \\ a_{m1}h_1 + \ldots + a_{mn}h_n \end{pmatrix} \in \mathbb{R}^m,$$

whereby the vector $h = (h_1,\ldots,h_n) \in \mathbb{R}^n$ is considered as an $n\times 1$ matrix.

The identity (**) can be rewritten in ***coordinate*** *form as*

$$\left\{ \begin{array}{c} f_i(x+h) = f_i(x) + \sum_{j=1}^n a_{ij}h_j + E_i(h), \\ i = 1,\ldots,m, \end{array} \right.$$

with

$$\lim_{h\to 0} \frac{|E_i(h)|}{\|h\|_{\mathbb{R}^n}} = 0.$$

It is obvious that the vector-valued function $f : U \to \mathbb{R}^m$ is differentiable at a point $x \in U$ if and only if all coordinate mappings $f_i : U \to \mathbb{R}$, $1 \le i \le m$, are differentiable.

(ii) Symbolically we write

$$E_f(h) = o\left(\|h\|_{\mathbb{R}^n}\right), \text{ as } h \to 0.$$

(iii) Let $f : \mathbb{R}^n \to \mathbb{R}$ (i.e., $m = 1$). Then

$$A = \begin{pmatrix} a_1, & a_2, & \ldots, & a_n \end{pmatrix} = (a_j)_{j=1}^n \ (1\times n\text{-matrix})$$

and

$$f(x+h) = f(x) + \sum_{j=1}^n a_jh_j + E_f(h),$$

where $E_f(h) \in \mathbb{R}$ is such that

$$\lim_{h\to 0} \frac{|E_f(h)|}{\|h\|} = 0.$$

Theorem 2.4.4. *Let $f : U \to \mathbb{R}^m$ be **differentiable** at a point $x \in U$, i.e.,*

$$f(x+h) = f(x) + Ah + o\left(\|h\|_{\mathbb{R}^n}\right)$$

with a matrix

$$A = (a_{ij})_{\substack{1\le i\le m \\ 1\le j\le n}}.$$

Then:

*(i) f is **continuous** at x*

*(ii) All components $f_i : U \to \mathbb{R}, 1 \le i \le m$, are **partially differentiable** at the point x and*

$$\frac{\partial f_i(x)}{\partial x_j} = a_{ij}, \quad 1 \le j \le n.$$

In other words, the derivative $Df(x)$ of f at x is the **matrix of first partial derivatives** $\frac{\partial f_i(x)}{\partial x_j}$ of the component functions f_i:

$$Df(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{pmatrix}.$$

Such a matrix is called the **Jacobian matrix** of the function f. **Notation:**

$$Df(x) = \frac{\partial(f_1,\dots,f_m)}{\partial(x_1,\dots,x_n)}(x) = \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{\substack{1\le i\le m \\ 1\le j\le n}}.$$

Proof of Theorem 2.4.4.

(i) We have

$$f(x+h) = f(x) + Ah + o\left(\|h\|\right), \text{ as } h \to 0.$$

Since $\lim_{h\to 0} Ah = 0$ and $\lim_{h\to 0} o\left(\|h\|\right) = 0$, finally

$$\lim_{h\to 0} f(x+h) = f(x).$$

(ii) For each $1 \le i \le m$

$$f_i(x+h) = f_i(x) + \sum_{j=1}^{n} a_{ij}h_j + E_i(h), \qquad \text{with } E_i(h) = o\left(\|h\|\right) \text{ as } h \to 0.$$

Hence for

$$h := te_j \in \mathbb{R}^n, \quad \|h\| = |t|, \quad t \in \mathbb{R}, 1 \le j \le n,$$

with $e_j = (\underbrace{0, \dots, 0, 1}_{j}, 0, \dots, 0)$ being the canonical basis vector in $\mathbb{R}^n$, it holds

$$\begin{aligned} f_i(x + te_j) &= f_i(x) + ta_{ij} + E_i(te_j), \\ \frac{\partial f_i}{\partial x_j}(x) : &= \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t} = a_{ij} + \lim_{t \to 0} \frac{E_i(te_j)}{t} = a_{ij}. \end{aligned}$$

□

Warning: The inverse statement is not true! Partial differentiability alone does **not** imply total differentiability. However, the continuity of all $x \mapsto \frac{\partial f_i}{\partial x_j}(x)$ would be sufficient to guarantee total differentiability (cf. Theorem 2.4.5 below).

For functions of *real* variables $f : U \to \mathbb{R}^m, x \in U \subseteq \mathbb{R}$ with $n = 1$, the notions of partial and total differentiability *coincide*. So, the total differentiability is a new concept only in the multidimensional case $n > 1$.

Theorem 2.4.5 (without proof here)**.** *Let $U \subset \mathbb{R}^n$ be open, and let $f : U \to \mathbb{R}^m$ be partially differentiable. If all partial derivatives*

$$\frac{\partial f_i}{\partial x_j}, \ 1 \le i \le m, \ 1 \le j \le n,$$

*are **continuous** at the point $x \in U$, then f is (totally) differentiable at x.*

We **summarize:** For $f : U \to \mathbb{R}^m$ the following implications hold:

$$\begin{gathered} \text{continuously partially differentiable} \\ \Downarrow \\ \text{totally differentiable} \\ \Downarrow \\ \text{partially differentiable.} \end{gathered}$$

Example 2.4.6. *Let $C := (c_{ij})_{1 \le i,j \le n}$ be a **symmetric** $n \times n$ matrix, i.e.,*

$$c_{ij} = c_{ij}, \quad \textit{for all } i, j,$$

and let

$$f(x) := \langle Cx, x \rangle_{\mathbb{R}^n} := \sum_{i,j=1}^{n} c_{ij} x_i x_j, \quad f : \mathbb{R}^n \to \mathbb{R},$$

*be the corresponding **quadratic form**. Then*

$$\begin{aligned} f(x+h) &= \langle C(x+h), x+h\rangle_{\mathbb{R}^n} \\ &= \langle Cx, x\rangle + \langle Cx, h\rangle + \langle Ch, x\rangle + \langle Ch, h\rangle \\ &= \langle Cx, x\rangle + 2\langle Cx, h\rangle + \langle Ch, h\rangle \\ &= f(x) + \langle a, h\rangle + E(h), \end{aligned}$$

with

$$\begin{aligned} a &= 2Cx, \quad E(h) = \langle Ch, h\rangle_{\mathbb{R}^n}, \quad |E(h)| \le \|C\| \cdot \|h\|_{\mathbb{R}^n}{}^2, \\ \|C\| : &= \|C\|_{\mathbb{R}^n \to \mathbb{R}^n} := \max_{1\le i\le n} \left(\sum_{1\le j\le n} c_{ij}^2 \right)^{1/2}. \end{aligned}$$

Since

$$\lim_{h\to 0} \frac{|E_f(h)|}{\|h\|} = 0,$$

we conclude that

$$\exists Df(x) = 2Cx \in \mathbb{R}^n.$$

Alternatively, we can calculate the partial derivatives

$$\frac{\partial f}{\partial x_j}(x) = 2\sum_{i=1}^n c_{ij}x_i = 2\sum_{i=1}^n c_{ji}x_i = 2(Cx)_j \in \mathbb{R},$$

which are continuous functions of x. So, by Theorem 2.4.5

$$\exists Df(x) = 2Cx = 2\left((Cx)_j\right)_{j=1}^n \in \mathbb{R}^n\, (1\times n - \mathit{matrix}).$$

Remark 2.4.7 (Remark to Theorem 2.4.5)**.** *Partially differentiable functions need not be continuous! The reason is that we consider limits along the axes, but not arbitrary sequences $(x_k)_{k\ge 1} \subset U$ converging to a given point $x \in U$.*

Exercise 2.4.8. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by*

$$f(x,y) = \begin{cases} \frac{y}{x^2} e^{-\frac{y}{x^2}}, & x \ne 0, \\ 0, & x = 0. \end{cases}$$

Show that:

(i) f is continuous on every line drawn through $(0,0)$*;*

(ii) f is not continuous at $(0,0)$*.*

*(**Hint:** Consider $y_k := cx_k^2$ with $x_k \to 0$ as $k \to \infty$.)*

2.5 Chain Rule

Theorem 2.5.1 (Chain Rule, without proof)**.** *Let us be given two functions,*

$$f : U \to \mathbb{R}^m \text{ and } g : V \to \mathbb{R}^p,$$

where $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ are open and $f(U) \subseteq V$. Suppose that f is differentiable at some $x \in U$ and g respectively at $y := f(x)$. Then the composite function

$$h := g \circ f : U \to \mathbb{R}^p$$

is differentiable at x, and its derivative is given by (via matrix multiplication)

$$Dh(x) = \underbrace{Dg(f(x))}_{p\times m}\underbrace{Df(x)}_{m\times n}. \quad (p \times n\text{-matrix}) \tag{$*$}$$

Idea of the proof. For any $x, \tilde{x} \in U$

$$\begin{aligned} h(x) - h(\tilde{x}) \quad &= \quad g(f(x)) - g(f(\tilde{x})); \\ g \text{ diff.} \quad &\Rightarrow \quad h(x) - h(\tilde{x}) \sim Dg(f(x))\,(f(x) - f(\tilde{x})), \text{ as } f(\tilde{x}) \to f(x), \\ f \text{ diff.} \quad &\Rightarrow \quad h(x) - h(\tilde{x}) \sim Dg(f(x))\,Df(x)\,(x - \tilde{x}), \text{ as } \tilde{x} \to x. \end{aligned}$$

A rigorous proof should take into account the error terms. □

In $(*)$ we have the product of two matrices: Let B be a $p \times m$ matrix and A be an $m \times n$ matrix,

$$A = (a_{ij})_{\substack{1\le i\le m\\1\le j\le n}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \ (m \times n),$$

$$B = (b_{ki})_{\substack{1\le k\le p\\1\le i\le m}} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pm} \end{pmatrix} \ (p \times m).$$

Then their product $C := BA$ is a $p \times n$ matrix defined as follows:

$$BA =: C = (c_{kj})_{\substack{1\le k\le p\\1\le j\le n}} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{p1} & c_{p2} & \dots & c_{pn} \end{pmatrix},$$

with the entries

$$c_{kj} := \sum_{i=1}^{m} b_{ki} \cdot a_{ij}, \ 1 \le k \le p, \quad 1 \le j \le n.$$

Typical applications of the Chain Rule

(i) Let $f : \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$, we define

$$h := g \circ f : \mathbb{R} \to \mathbb{R}.$$

$$\mathbb{R} \ni t \to \begin{pmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{pmatrix} =: x \in \mathbb{R}^n,$$
$$\mathbb{R}^n \ni x = (x_1, \ldots, x_n) \to g(x_1, \ldots, x_n) \in \mathbb{R},$$
$$\mathbb{R} \ni t \to h(t) := g(f_1(t), \ldots, f_n(t)).$$

Then

$$Df(t) = \begin{pmatrix} f_1'(t) \\ \vdots \\ f_n'(t) \end{pmatrix} \in \mathbb{R}^n,$$
$$Dg(x) = \nabla g(x) = \left(\frac{\partial g}{\partial x_1}(x), \ldots, \frac{\partial g}{\partial x_n}(x) \right) \in \mathbb{R}^n.$$

By Theorem 2.5.1

$$\begin{aligned} h'(t) &= Dg[f(t)]Df(t) \\ &= \left(\frac{\partial g}{\partial x_1}(f(t)), \ldots, \frac{\partial g}{\partial x_n}(f(t)) \right) \times \begin{pmatrix} f_1'(t) \\ \vdots \\ f_n'(t) \end{pmatrix} \\ &= \langle \nabla_x g(f(t)), \nabla f(t) \rangle_{\mathbb{R}^n} = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(f(t)) \cdot f_i'(t) \in \mathbb{R}. \end{aligned}$$

Example 2.5.2 (Numerical Example)**.** *Let*

$$f(t) = \begin{pmatrix} t \\ t^2 \end{pmatrix} =: \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x \in \mathbb{R}^2, \quad g(x) = g(x_1, x_2) := x_1 - x_2^2.$$

Then

$$h(t) = g(f(t)) = t - t^4, \quad h'(t) = 1 - 4t^3, \; t \in \mathbb{R}.$$

On the other hand

$$f'(t) = \begin{pmatrix} 1 \\ 2t \end{pmatrix}, \quad \nabla g(x_1, x_2) = (1, -2x_2),$$

and hence (substituting x_2 by t^2)

$$h'(t) = (1, -2t^2) \times \begin{pmatrix} 1 \\ 2t \end{pmatrix} = 1 - 4t^3.$$

(ii) **Applications to directional derivatives (Section 2.2 revisited)**

Let $U \subset \mathbb{R}^n$ be open, and let $f : U \to \mathbb{R}$ be differentiable. Choose some unit vector $v \in \mathbb{R}^n$ with $|v| = 1$. Then the directional derivative along v is defined by

$$\begin{aligned}\partial_v f(x) :&= \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t} \\ &= \frac{df(x+tv)}{dt}\Big|_{t=0}.\end{aligned}$$

Theorem 2.5.3. *Let $f : U \to \mathbb{R}$ be totally differentiable and let $v \in \mathbb{R}^n$ with $|v| = 1$. Then, for any $x \in U$*

$$\partial_v f(x) = \langle \nabla f(x), v\rangle_{\mathbb{R}^n} = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) \cdot v_i.$$

Proof. By the above definition

$$\partial_v f(x) = g_v'(t)\big|_{t=0},$$

with a scalar function

$$\begin{gathered} g_v : \mathcal{I} \to \mathbb{R} \quad \mathcal{I} := (-\varepsilon, \varepsilon) \subset \mathbb{R} \ \ (\text{i.e.} n = m = 1)\,, \\ \mathcal{I} \ni t \to g_v(t) = f(x+tv) \in \mathbb{R}, \end{gathered}$$

where $\varepsilon > 0$ is small enough such that $B_\varepsilon(x) \subset U$. But

$$g_v(t) = f(\varphi(t)),$$

where we set

$$\mathcal{I} \ni t \to \varphi(t) := x + tv \in \mathbb{R}^n, \ \ \varphi(0) := x.$$

Obviously, φ is differentiable and $\varphi'(t) = v \in \mathbb{R}^n$ for all $t \in \mathcal{I}$. By the chain rule (Theorem 2.5.1)

$$g_v'(t) = Df(\varphi(t)) \cdot \varphi'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\varphi(t)) \cdot v_i = \langle \nabla f(\varphi(t)), v\rangle_{\mathbb{R}^n},$$

and for $t = 0$

$$\partial_v f(x) = g_v'(0) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) \cdot v_i = \langle \nabla f(x), v\rangle.$$

□

(iii) Further rules: Linearity, i.e., for any $f, g : U \to \mathbb{R}^m$

$$D(f+g) = Df + Dg,$$
$$D(\alpha f) = \alpha Df, \ \alpha \in \mathbb{R}.$$

Example: *Polar coordinates*

$$x = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}, \quad r > 0, \quad \varphi \in \mathbb{R}.$$

Let us be given a differentiable function $f : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \to f(x_1, x_2) \in \mathbb{R}$. Then,

$$g(r,\varphi) := f\begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}, \ r > 0, \ \varphi \in \mathbb{R},$$

defines a differential function $g : (0, +\infty) \times \mathbb{R} \to \mathbb{R}$ with partial derivatives

$$\frac{\partial g(r,\varphi)}{\partial r} = \frac{\partial f(r,\varphi)}{\partial x_1}\cos\varphi + \frac{\partial f(r,\varphi)}{\partial x_2}\sin\varphi,$$
$$\frac{\partial g(r,\varphi)}{\partial \varphi} = -r\frac{\partial f(r,\varphi)}{\partial x_1}\sin\varphi + r\frac{\partial f(r,\varphi)}{\partial x_2}\cos\varphi.$$

2.6 Taylor's Formula

Intuition: Review of **1-dim**
Let us recall the following:

Theorem 2.6.1 (Mean value theorem). *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function (i.e.,$f \in C^1(\mathbb{R})$). Then for each $a, b \in \mathbb{R}$, $a < b$, there exists $\theta \in (a, b)$ such that*

$$f(b) - f(a) = f'(\theta) \cdot (b - a). \tag{$*$}$$

Taylor's formula is a generalization of $(*)$ to $(k+1)-$times differentiable functions $(k = 0, 1, 2, \ldots)$. As a result we get a (finite) series expansion of a function f about a fixed point, up to the $(k+1)-$th **Taylor remainder**. The following is well known from Calculus:

Definition 2.6.2 (Taylor's Formula). *Let $f : \mathbb{R} \to \mathbb{R}$ be a $(k+1)-$times continuously differentiable function on an open interval $\mathcal{I} \subset \mathbb{R}$. Then for all $x, x+h \in \mathcal{I}$, we have the Taylor approximation of f*

$$f(x+h) = f(x) + \sum_{l=1}^{k} \frac{f^{(l)}(x)}{l!} h^l + E_{k+1}, \tag{$**$}$$

where the $(k+1)-th$ error term E_{k+1} can be represented by

$$E_{k+1}(x,h) = \frac{f^{(k+1)}(x+\lambda h)}{(k+1)!}h^{k+1}$$

for some $\lambda = \lambda(x,h) \in (0,1)$. Recall that $l! := l \cdot (l-1) \cdot \ldots \cdot 2 \cdot 1$ and $0! := 1$.

This is the so-called **Lagrange form** of the remainder term E_{k+1}. Of course, E_{k+1} and λ depend on the point x, around which we write the expansion, as well as on the increment h. Since $\lambda \in (0,1)$, we see that $x + \lambda h$ is some intermediate point between x and $x + \lambda h$. Obviously, $\lim_{h\to 0} E_{k+1}(x,h)/h^k = 0$ and hence $E_{k+1}(x,h) = o(h^k), h \to 0$.

Sometimes, Taylor's formula is written in the equivalent form

$$f(x+h) = f(x) + \sum_{l=1}^{k} \frac{f^{(l)}(x)}{l!}h^l + o(h^k),\ \ h \to 0.$$

If $k = 0$, we just get the mean value theorem $(*)$

$$f(x+h) - f(x) = f'(x+\lambda h)h,\ \ \lambda \in (0,1).$$

Generalization to several variables

Theorem 2.6.3 (Multi-dimensional Taylor's Formula)**.** *Let $U \subseteq \mathbb{R}^n$ be open; let $x \in U$ and hence $B_\delta(x) \subset U$ for some $\delta > 0$. Let*

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

be $(k+1)-$times continuously differentiable (i.e.,$f \in C^{k+1}(U)$)$\dot{}$Then for any $h \in \mathbb{R}^n$ with $\|h\|_{\mathbb{R}^n} < \delta$ there exists $\theta = \theta(x,h) \in (0,1)$ such that

$$f(x+h) = \sum_{0\le|\alpha|\le k} \frac{D^\alpha f(x)}{\alpha!}h^\alpha + E_{k+1} \qquad (***)$$

with $E_{k+1}(x,h) = \sum_{|\alpha|=k+1} \frac{D^\alpha f(x+\theta h)}{|\alpha|!}h^\alpha$, where the summation is over all (i.e., with all possible permutations) multi-indices

$$\alpha = (\alpha_1, \ldots, \alpha_n) \in (\mathbb{Z}_+)^n \text{ with order (degree)} |\alpha| \le k.$$

Multi-index notation:

$$|\alpha| := \alpha_1 + \ldots + \alpha_n,\ \ \alpha_i \in \mathbb{Z}_+ := \{0,1,2,\ldots\},$$
$$k! := k \cdot (k-1) \cdot \ldots \cdot 2 \cdot 1,\ \ 0! := 1,$$
$$h^\alpha = h_1^{\alpha_1} h_2^{\alpha_2} \cdots h_n^{\alpha_n},\ \ h = (h_1, h_2, \cdots, h_n) \in \mathbb{R}^n,$$
$$D^\alpha f(x) = D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} \cdots D_{x_n}^{\alpha_n} f(x) := \frac{\partial^{|\alpha|} f(x)}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}.$$

Proof will be done below.

Corollary 2.6.4. *Under the above conditions*

$$f(x+h) = \sum_{0\le|\alpha|\le k} \frac{D^\alpha f(x)}{|\alpha|!} h^\alpha + o(\|h\|^k), \;\; h \to 0.$$

Remark 2.6.5.

(i) Actually, the later formula with $o(\|h\|^k)$ is true if we just know that f is k-times differentiable at the point x. But for the Lagrange representation of the error term $E_{k+1}(x,h)$ in Theorem 2.6.3, we have to assume that $f \in C^{k+1}(U)$.

(ii) If we do not allow permutations of indexes, then in Taylor's formula instead of $|\alpha|!$ we should take $\alpha_1! \dots \alpha_n!$.

Example 2.6.6 (Particular Cases)**.**

(i) ***Taylor approximation of order*** $k = 2$ *for* $f \in C^2(U)$

$$\begin{aligned} f(x+h) &= f(x) + \sum_{i=1}^n \partial_i f(x) \cdot h_i + \frac{1}{2} \sum_{i,j=1}^n \partial^2_{i,j} f(x) \cdot h_i h_j + o(\|h\|^2) \\ &= f(x) + \langle \mathrm{grad} f(x), h \rangle_{\mathbb{R}^n} + \frac{1}{2} \langle h, \mathrm{Hess} f(x) \cdot h \rangle_{\mathbb{R}^n} + o(\|h\|^2), \quad h \to 0. \end{aligned}$$

We here use the ***gradient*** *of f*

$$\mathrm{grad} f(x) := \nabla f(x) := Df(x) := (\partial_1 f(x), \dots, \partial_n f(x)) \in \mathbb{R}^n$$

and the ***Hessian*** *of f (i.e., its* ***matrix of second derivatives****)*

$$\mathrm{Hess} f(x) := D^2 f(x) := \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}.$$

For shorthand,

$$\mathrm{Hess} f(x) := \big(\partial_{x_i} \partial_{x_j} f(x)\big)_{\substack{1\le i\le n,\\ 1\le j\le n}},$$

which is a symmetric $n \times n$ matrix by Theorem 2.3.1.

(ii) For $n = k = 2$ we have in the ***coordinate form***

$$\begin{aligned} f(x_1+h_1, x_2+h_2) \;\; = \;\; & f(x_1,x_2) + \partial_{x_1} f(x_1,x_2) h_1 + \partial_{x_2} f(x_1,x_2) h_2 \\ & + \frac{1}{2} \partial^2_{x_1} f(x_1,x_2) h_1^2 + \partial^2_{x_1 x_2} f(x_1,x_2) h_1 h_2 \\ & + \frac{1}{2} \partial^2_{x_2} f(x_1,x_2) h_2^2 + o(h_1^2 + h_2^2). \end{aligned}$$

Proof of Theorem 2.6.3. Set

$$g(t) = f(x + th),\ t \in \mathcal{I} \supseteq [0, 1].$$

Applying the one-dimensional Taylor formula to the function g on an open interval $\mathcal{I} \subset \mathbb{R}$, we get with some $\lambda = \lambda(t) \in (0, 1)$

$$\begin{aligned} g(t) &= g(0) + g'(0)t + \frac{1}{2}g''(0)t^2 + \frac{1}{6}g'''(0)t^3 + \dots \\ &\quad + \frac{1}{k!}g^{(k)}(0)t^k + \frac{1}{(k+1)!}g^{(k+1)}(\lambda)t^{k+1}. \end{aligned}$$

By the chain rule

$$\begin{aligned} g'(t) = \langle Df(x+th), h\rangle_{\mathbb{R}^n},\ g''(t) = \langle D^2 f(x+th)h, h\rangle_{\mathbb{R}^n},\ \dots \\ \text{and hence } g'(0) = \langle Df(x), h\rangle_{\mathbb{R}^n},\ g''(0) = \langle D^2 f(x)h, h\rangle_{\mathbb{R}^n},\ \dots \end{aligned}$$

Finally we put $t = 1$ and get the required expression for $g(1) = f(x + h)$.

□

Example 2.6.7. *Compute the Taylor approximation of order two ($k = n = 2$) of the Cobb-Douglas function*

$$f(x, y) = x^{1/4}y^{3/4} \text{ at point } (1, 1).$$

Solution 2.6.8. *In the open domain $U = \{x > 0,\ y > 0\} \subset \mathbb{R}^2$*

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{1}{4}x^{-3/4}y^{3/4}, \quad \frac{\partial f}{\partial y} = \frac{3}{4}x^{1/4}y^{-1/4}, \\ \frac{\partial^2 f}{\partial x^2} &= -\frac{3}{16}x^{-7/4}y^{3/4}, \quad \frac{\partial^2 f}{\partial y^2} = -\frac{3}{16}x^{1/4}y^{-5/4}, \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial^2 f}{\partial y \partial x} = \frac{3}{16}x^{-3/4}y^{-1/4}. \end{aligned}$$

Evaluating these derivatives at $x = y = 1$ gives

$$\frac{\partial f}{\partial x} = \frac{1}{4}, \qquad \frac{\partial f}{\partial y} = \frac{3}{4}, \qquad \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial y^2} = -\frac{3}{16}, \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{3}{16}.$$

Therefore,

$$f(x + th, y + tg) = 1 + \frac{1}{4}h + \frac{3}{4}g - \frac{3}{32}(h^2 + g^2) + \frac{3}{16}hg + o(h^2 + g^2),\ \text{as } h, g \to 0.$$

2.7 Implicit Functions

Before we studied **explicit** functions

$$f : U \to \mathbb{R}^m,$$
$$\mathbb{R}^n \supset \underbrace{U}_{open} \ni x \to f(x) =: y \in \mathbb{R}^m,$$
$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix} \in \mathbb{R}^m.$$

This ideal situation does not always occur in economic models. Frequently, such models are described by *"mixed"* equations like

$$F(x, y) = 0, \quad F : \underset{\subset \mathbb{R}^n}{U_1} \times \underset{\subset \mathbb{R}^m}{U_2} \to \mathbb{R}^m, \tag{$*$}$$

i.e., in coordinate form

$$\begin{cases} F_1(x_1, \dots, x_n; y_1, \dots, y_m) = 0, \\ \vdots \\ F_m(x_1, \dots, x_n; y_1, \dots, y_m) = 0, \end{cases}$$

where $x_1, \dots, x_n \in \mathbb{R}$ are called *exogenous* variables and $y_1, \dots, y_m \in \mathbb{R}$ resp. *endogenous* variables.

In particular, for $m = n = 1$ we have

$$F(x, y) = 0, \quad x, y \in \mathbb{R}.$$

As a rule, we cannot solve $(*)$ by some explicit formula separating the independent variables $x_1, \dots, x_n$ on one side and $y_1, \dots, y_m$ on the other.

Interpretation: $x = (x_1, \dots, x_n)$ is a *vector of parameters* and $y = (y_1, \dots, y_m)$ is the *output vector* we seek to describe the model. If for each $(x_1, \dots, x_n) \in U$ the equation $(*)$ determines a unique value $(y_1, \dots, y_m) \in \mathbb{R}^m$, we say that we have an ***implicit*** function

$$y = g(x) \in \mathbb{R}^m, \quad x \in U.$$

Below we study existence and differentiability properties of implicit functions.

Intuition: 1-dim case:

Let $U \subset \mathbb{R}^2$ be *open*, and consider a *differentiable* function

$$F : U \to \mathbb{R}, \quad (x, y) \to F(x, y).$$

Fix some point $(x_0, y_0) \in U$ such that $F(x_0, y_0) = 0$, and **suppose (!!!) that there exists a differentiable function**

$$g : \mathcal{I} \to \mathbb{R}, \quad x \to g(x), \quad g(x_0) = y_0,$$

defined on some **open interval** $\mathcal{I} \ni x_0$ such that

$$(x, g(x)) \in U \text{ and } F(x, g(x)) = 0 \text{ for all } x \in \mathcal{I}.$$

Differentiating the equation $F(x, g(x)) = 0$, we get by the Chain Rule that

$$\frac{\partial}{\partial x} F(x, g(x)) + \frac{\partial}{\partial y} F(x, g(x)) \cdot g'(x) = 0, \quad x \in \mathcal{I}.$$

Assuming that

$$\frac{\partial}{\partial y} F(x_0, y_0) \neq 0,$$

we conclude that

$$g'(x_0) = -\frac{\frac{\partial}{\partial x} F(x_0, y_0)}{\frac{\partial}{\partial y} F(x_0, y_0)}.$$

Indeed we have the following classical theorem from Calculus.

Theorem 2.7.1 (1-dim Implicit Function Theorem, IFT)**.** *Suppose that $F(x, y)$ is a **continuously differentiable** function on an open domain $U \subset \mathbb{R}^2$, (i.e., $F \in C^1(U)$, which means that $\partial_x F, \partial_y F : U \to \mathbb{R}$ are continuous). Let a point $(x_0, y_0) \in U$ be such that $F(x_0, y_0) = 0$. If*

$$\frac{\partial}{\partial y} F(x_0, y_0) \neq 0,$$

then there exist open intervals

$$\mathcal{I} \ni x_0, \quad \mathcal{J} \ni y_0, \quad \mathcal{I} \times \mathcal{J} \subset U,$$

*and a **continuously differentiable** function*

$$g : \mathcal{I} \to \mathcal{J}, \quad g(x_0) = y_0,$$

such that $F(x, g(x)) = 0$ for all $x \in \mathcal{I}$ and

$$g'(x_0) = -\frac{\frac{\partial}{\partial x} F(x_0, y_0)}{\frac{\partial}{\partial y} F(x_0, y_0)}.$$

Furthermore, such g is unique: if $(x, y) \in \mathcal{I} \times \mathcal{J}$ and $F(x, y) = 0$, then surely $y = g(x)$.

Remark 2.7.2. *The proof of the existence of g in Theorem 2.7.1 is based on the Banach Contraction Theorem (Th. 1.12) and is highly non-trivial. This is a* ***local result*** *since it is stated on some (probably very small) open intervals $\mathcal{I} \ni x_0$, $\mathcal{J} \ni y_0$.*

Interpretation in economics: Comparative Statistics:

The IFT allows to study in what direction does the equilibrium $y(x)$ change in a control variable x. The equilibrium is typically described by some equation $F(x, y) = 0$.

Example 2.7.3.

(i) Let $\mathcal{I} = (-a, a)$, consider the function describing an ***upper half-circle***

$$y := g(x) = \sqrt{a^2 - x^2},\ x \in \mathcal{I}.$$

By direct calculations

$$\exists g'(x) = -\frac{x}{\sqrt{a^2 - x^2}},\ x \in \mathcal{I}.$$

Let us check that IFT gives the same result. We have

$$y^2 := g^2(x) = a^2 - x^2 \iff$$
$$F(x, y) := x^2 + y^2 - a^2 = 0$$

on the open domain $U := \{(x, y) |\ x \in \mathcal{I},\ y > 0\} \subset \mathbb{R}^2$. So, for any $(x, y) \in U$

$$\frac{\partial F}{\partial x} = 2x, \quad \frac{\partial F}{\partial y} = 2y \neq 0,\ \textit{and}$$

$$\exists g'(x) = -\frac{2x}{2\sqrt{a^2 - x^2}} = -\frac{x}{\sqrt{a^2 - x^2}}.$$

(ii) A ***cubic*** *implicit function*

$$F(x, y) = x^2 - 3xy + y^3 - 7 = 0,\ (x, y) \in \mathbb{R}^2,$$

with

$$(x_0, y_0) = (4, 3) \quad \textit{and} \quad F(x_0, y_0) = 0.$$

Indeed,

$$\frac{\partial F}{\partial x} = 2x - 3y = -1\ \textit{at}\ (x_0, y_0),$$
$$\frac{\partial F}{\partial y} = -3x + 3y^2 = 15\ \textit{at}\ (x_0, y_0).$$

Theorem 2.7.1 tells us that $F(x, y) = 0$ indeed defines $y = g(x)$ as a C^1 function of x around the point with coordinates $x_0 = 4$ and $y_0 = 3$. Furthermore,

$$y'(x_0) = g'(x_0) = -\frac{\frac{\partial F}{\partial x}(x_0, y_0)}{\frac{\partial F}{\partial y}(x_0, y_0)} = \frac{1}{15}.$$

*(iii) A **unit circle** is described by*

$$F(x,y) = x^2 + y^2 - 1 = 0$$
$$\text{with } \frac{\partial F}{\partial x} = 2x, \quad \frac{\partial F}{\partial y} = 2y.$$

(a) *Let first $(x_0, y_0) = (0, 1)$, so that $\frac{\partial F}{\partial x}(x_0, y_0) = 0$, $\frac{\partial F}{\partial y}(x_0, y_0) = 2 \neq 0$. By Theorem 2.7.1 the implicit function $y = g(x)$ exists around $x_0 = 0$ and $y_0 = 1$, with $g'(x_0) = -0/2 = 0$. In this case we have an explicit formula*

$$\begin{aligned} y^2(x) &= 1 - x^2 \Rightarrow \\ y(x) &= \sqrt{1 - x^2} > 0. \end{aligned}$$

We also can compute directly

$$y'(x) = \frac{1}{2}\frac{-2x}{\sqrt{1-x^2}}, \quad y'(x_0) = 0.$$

(b) *On the other hand, no nice function $y = g(x)$ exists around the initial point $(x_0, y_0) = (1, 0)$. Actually, Theorem 2.7.1 does not apply since $\frac{\partial F}{\partial y}(x_0, y_0) = 0$. On the picture we can see two branches tending to the point $(1, 0)$:*

$$y(x) = \pm\sqrt{1 - x^2}.$$

IFT, Multidimensional Case

Theorem 2.7.4 (Multidimensional IFT)**.** *Let $U_1 \subset \mathbb{R}^n$ and $U_2 \subset \mathbb{R}^m$ be open domains and let*

$$F : U_1 \times U_2 \to \mathbb{R}^m, \quad (x, y) \to F(x, y),$$

*be **continuously differentiable**, i.e., $F \in C^1(U_1 \times U_2)$, which means that all $\frac{\partial F_i}{\partial x_j}, \frac{\partial F_i}{\partial y_k} : U_1 \times U_2 \to \mathbb{R}$ are continuous, $1 \leq j \leq n$, $1 \leq i, k \leq m$. Let a point $(x_0, y_0) \in U_1 \times U_2$ be such that $F(x_0, y_0) = 0$. Suppose that the $m \times m$-matrix of partial derivatives w.r.t. $y = (y_1, \dots, y_m)$*

$$\frac{\partial F}{\partial y} = \frac{\partial(F_1, \dots, F_m)}{\partial(y_1, \dots, y_m)} = \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \frac{\partial F_2}{\partial y_1} & \frac{\partial F_2}{\partial y_2} & \cdots & \frac{\partial F_2}{\partial y_m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial F_m}{\partial y_1} & \frac{\partial F_m}{\partial y_2} & \cdots & \frac{\partial F_m}{\partial y_m} \end{pmatrix}$$

*is **invertible** at the point (x_0, y_0), i.e., its **determinant***

$$\det \frac{\partial F}{\partial y}(x_0, y_0) \neq 0.$$

*Then, there **exist**:*

(i) ***open neighbourhoods*** $V_1 \subseteq U_1$ *of* x_0 *resp.* $V_2 \subseteq U_2$ *of* y_0 *(in general, they can be smaller than* U_1 *resp.* U_2*),*

(ii) a ***continuously differentiable*** *function*

$$g : V_1 \to V_2, \text{ with } g(x_0) = y_0,$$

such that

$$F(x, g(x)) = 0 \text{ for all } x \in V_1.$$

Such function is ***unique*** *in the following sense: if* $(x, y) \in V_1 \times V_2$ *obey* $F(x, y) = 0$*, then* $y = g(x)$*. Furthermore, the* ***derivative*** *at point* x_0 ***equals***

$$\underbrace{Dg(x_0)}_{m\times n} = -\underbrace{\left[\frac{\partial F}{\partial y}(x_0, y_0)\right]^{-1}}_{m\times m} \cdot \underbrace{\frac{\partial F}{\partial x}(x_0, y_0)}_{m\times n}.$$

Example 2.7.5 (Special Cases)**.**

(i) $m = 1$*, i.e.,* $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$*,*

$$F(x_1, \ldots, x_n; y_1) = 0.$$

The implicit function

$$y = g(x_1, \ldots, x_n) \in \mathbb{R}$$

exists under the sufficient condition

$$\frac{\partial F}{\partial y}(x_0, y_0) \neq 0.$$

Then $Dg(x_0) = \nabla g(x_0) = (\partial_i g(x_0))_{i=1}^n$*, whereby the partial derivatives* $\partial_j g(x_0)$ *w.r.t.* x_j *are given by*

$$\partial_j g(x_0) = -\frac{\partial_j F(x_0, y_0)}{\frac{\partial F}{\partial y}(x_0, y_0)}, \quad 1 \leq j \leq n.$$

(ii) $n = 1$*,* $m = 2$*, i.e.,* $F : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$*,*

$$\begin{cases} F_1(x, y_1, y_2) = 0, \\ F_2(x, y_1, y_2) = 0. \end{cases}$$

The sufficient condition is stated in terms of

$$\frac{\partial F}{\partial y} = \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} \\ \frac{\partial F_2}{\partial y_1} & \frac{\partial F_2}{\partial y_2} \end{pmatrix},$$

namely,

$$\det \frac{\partial F}{\partial y}(x_0, y_0) = \left(\frac{\partial F_1}{\partial y_1} \cdot \frac{\partial F_2}{\partial y_2} - \frac{\partial F_1}{\partial y_2} \cdot \frac{\partial F_2}{\partial y_1} \right)(x_0, y_0) \neq 0.$$

Then there exists $g(x) = (g_1(x), g_2(x)) \in \mathbb{R}^2$ *around* x_0 *and*

$$Dg(x_0) = \begin{pmatrix} g_1'(x_0) \\ g_2'(x_0) \end{pmatrix} = -\left(\frac{\partial F}{\partial y}(x_0, y_0) \right)^{-1} \cdot \begin{pmatrix} \frac{\partial F_1}{\partial x}(x_0, y_0) \\ \frac{\partial F_2}{\partial x}(x_0, y_0) \end{pmatrix}.$$

Numerical Example: $n = 1$, $m = 2$, *i.e.,* $F : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$,

$$F(x, y_1, y_2) = \begin{cases} -2x^2 + y_1^2 + y_2^2 = 0, \\ x^2 + e^{y_1 - 1} - 2y_2 = 0 \end{cases}$$

at point $x_0 = 1, y_0 = (1, 1)$. *After calculations*

$$D_y F(x, y_1, y_2) = \begin{pmatrix} 2y_1 & 2y_2 \\ e^{y_1 - 1} & -2 \end{pmatrix},$$

and at the point $(x_0, y_0) \in \mathbb{R} \times \mathbb{R}^2$

$$D_y F(x_0, y_0) = \begin{pmatrix} 2 & 2 \\ 1 & -2 \end{pmatrix},$$

$$\det D_y F(x_0, y_0) = 2 \cdot (-2) - 1 \cdot 2 = -6 \neq 0.$$

The inverse matrix

$$\left(\frac{\partial F}{\partial y}(x_0, y_0) \right)^{-1} = \frac{1}{-6} \cdot \begin{pmatrix} -2 & -2 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1/3 & 1/3 \\ 1/6 & -1/3 \end{pmatrix}.$$

Also, by direct calculations

$$D_x F(x_0, y_0) = \begin{pmatrix} -4 \\ 2 \end{pmatrix}.$$

Thus,

$$\frac{dg}{dx}(x_0) = \begin{pmatrix} 1/3 & 1/3 \\ 1/6 & -1/3 \end{pmatrix} \cdot \begin{pmatrix} -4 \\ 2 \end{pmatrix} = \begin{pmatrix} -2/3 \\ -2 \end{pmatrix}.$$

Reminder: *Let*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ with } \det A := ad - bc \neq 0.$$

Then, the inverse matrix is calculated by

$$A^{-1} = \frac{1}{\det A} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

2.8 Inverse Functions

Let $f : U \to \mathbb{R}^n$, $U \subset \mathbb{R}^n$ – open set (now $m = n$!).
Problem: Does there exist an inverse mapping

$$g := f^{-1} : f(U) \to U?$$

Theorem 2.8.1. *Let $U \subset \mathbb{R}^n$ be open domains and let $f : U \to \mathbb{R}^n$ be continuously differentiable, i.e., $f \in C^1(U)$. Let $x_0 \in U$ and $y_0 := f(x_0)$. Suppose that the* ***Jacobi*** *matrix of partial derivatives*

$$Df = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

is ***invertible*** *at point x_0, i.e., its* ***determinant*** $\neq 0$. *Then, there exist* ***open neighbourhoods*** *$U_0 \subseteq U$ of x_0 resp. $V_0 \subseteq \mathbb{R}^n$ of y_0 such that the mapping*

$$f : U_0 \to V_0$$

is ***one-to-one*** *(bijection) and the* ***inverse function***

$$g := f^{-1} : V_0 \to U_0, \quad \textit{acting by}$$
$$(f^{-1} \circ f)(x) = x, \ (f \circ f^{-1})y = y, \ \forall x \in U_0, \ \forall y \in V_0,$$

is ***continuously differentiable*** *on V_0. Furthermore, the following holds:*

$$Dg(y_0) = [Df(x_0)]^{-1}.$$

Proof. Define the function

$$F : U \times \mathbb{R}^n \to \mathbb{R}^n,$$
$$F(x, y) := y - f(x).$$

Then $F(x_0, y_0) = 0$ and

$$\frac{\partial F}{\partial x}(x, y) = -Df(x), \quad \frac{\partial F}{\partial x}(x_0, y_0) = -Df(x_0) \neq 0, \quad \frac{\partial F}{\partial y}(x, y) = \mathrm{Id}_{\mathbb{R}^n},$$

where $\mathrm{Id}_{\mathbb{R}^n}$ is the identity $n \times n$-matrix. We claim that the equation

$$F(x, y) := y - f(x) = 0$$

locally defines the implicit function $x := g(y) = f^{-1}(y)$. Indeed, by Theorem 2.8.1 there exist $V_0 \subseteq \mathbb{R}^n$ and a function $g : V_0 \to \mathbb{R}^n, g \in C^1(V_0)$, such that

$$x = g(y),\ y = f(g(y)),\ y \in V_0.$$

So,

$$g = f^{-1} \text{ on } V_0$$

and

$$Dg(y_0) = -\left[\frac{\partial F}{\partial x}(x_0, y_0)\right]^{-1} \cdot \mathrm{Id}_{\mathbb{R}^n} = -\left[Df(x_0)\right]^{-1}.$$

□

Special case: $n = 1$ and $f : U \to \mathbb{R}$. The sufficient condition is

$$f'(x_0) \neq 0.$$

$$\text{Then, } g'(y_0) = \frac{1}{f'(x_0)}.$$

Example 2.8.2. *Let*

$$f\begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix} \in \mathbb{R}^2, \quad x, y \in \mathbb{R}. \qquad \textit{Then,}$$

$$Df(x,y) = \frac{\partial f(x,y)}{\partial(x,y)} = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}, \quad \det Df(x,y) = 4(x^2 + y^2).$$

By IFT, f is (locally) invertible at every point $(x, y) \in \mathbb{R}^2$ except $(0, 0)$. But globally f is not one-to-one, since for all $(x, y) \in \mathbb{R}^2$

$$f\begin{pmatrix} x \\ y \end{pmatrix} = f\begin{pmatrix} -x \\ -y \end{pmatrix}.$$

2.9 Unconstrained Optimization

We now turn to study of optimization theory under *assumptions of differentiability.*

Definition 2.9.1. *Let $U \subset \mathbb{R}^n$ be an* ***open*** *domain and let*

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

be an ***objective function*** *whose extrema we would like to analyse.*

(i) A point $x^* \in U$ is a ***local maximum*** *(resp.* ***minimum****) of f if there exists a ball $B_\varepsilon(x^*) \subset U$ such that for all $x \in B_\varepsilon(x^*)$*

$$f(x^*) \geq f(x)(\text{ resp. } f(x^*) \leq f(x)).$$

Local max or min are called ***local extrema****.*

*(ii) A point $x^* \in U$ is a **global** (or **absolute**) **maximum** (resp.**minimum**) of f if for all $x \in U$*

$$f(x^*) \geq f(x)(\text{ resp. } f(x^*) \leq f(x)).$$

*(iii) A point $x^* \in U$ is a **strict local maximum** (resp. **minimum**) of f if there exists a ball $B_\varepsilon(x^*) \subset U$ such that for all $x \neq x^*$ in $B_\varepsilon(x^*)$*

$$f(x^*) > f(x)(\text{ resp. } f(x^*) < f(x)).$$

Remark 2.9.2. *In the definition of the global extrema, the function $f : U \to \mathbb{R}^n$ can be defined on any domain U, which is **not necessarily open**.*

We want to use methods of **Calculus** to find local extrema. So, we need smoothness (i.e., differentiability) of f.

2.10 First-Order Conditions

Aim: To find **necessary** conditions for local extrema.

Theorem 2.10.1 (Necessary Condition for Local Extrema)**.** *Let $U \subset \mathbb{R}^n$ be an open domain and $f : U \to \mathbb{R}$ be **partially differentiable** on U (i.e., all its partial derivatives $\partial f/\partial x_i : U \to \mathbb{R}$, $1 \leq i \leq n$, exist). Then,*

$$x^* \in U \text{ is a } \textbf{local extremum} \text{ for } f$$
$$\implies \text{grad} f(x^*) := \nabla f(x^*) = \left(\frac{\partial f}{\partial x_1}(x^*), \ldots, \frac{\partial f}{\partial x_n}(x^*)\right) = 0.$$

Proof. For $i = 1, \ldots, n$ define a function

$$t \to g_i(t) := f(x^* + te_i), \text{ where}$$
$$e_i = (0, \ldots, 0, \underbrace{1}_{i}, 0, \ldots, 0) \in \mathbb{R}^n \text{ is a unit basis vector in } \mathbb{R}^n.$$

Here $t \in (-\varepsilon, \varepsilon)$ with a sufficiently small $\varepsilon > 0$ such that

$$\{x^* + te_i \mid -\varepsilon < t < \varepsilon\} \subset B_\varepsilon(x^*) \subset U \text{ for all } 1 \leq i \leq n.$$

If x^* is a local extremum for $f(x_1, \ldots, x_n)$, then clearly each real function $g_i(t) : (-\varepsilon, \varepsilon) \to \mathbb{R}$ has a local extremum at $t = 0$. Applying the one-dimensional necessary condition for extrema (well known from Calculus), we conclude that

$$\frac{\partial f}{\partial x_i}(x^*) = g_i'(0) = 0.$$

□

2.11 Second-Order Conditions

Aim: To find **sufficient** conditions for local extrema.

Definition 2.11.1. *Any point $x^* \in U$ satisfying the 1st condition $\nabla f(x^*) = 0$ is called a **critical point** of f on U.*

The 1st order conditions for local optima do **not** distinguish between maxima and minima. To determine whether some critical point $x^* \in U$ is a local max or min, we need to examine the behaviour of the second derivative $D^2 f(x^*)$. To this end, we assume that f is **twice continuously differentiable** on U, i.e., $f \in C^2(U)$, which means that all $\frac{\partial^2 f(x)}{\partial x_i \partial x_j} : U \to \mathbb{R}$ are continuous, $1 \le i, j \le n$. To formulate the sufficient conditions we need to use the **Hessian** of f, which is the $n \times n$ matrix of 2nd partial derivatives:

$$\mathrm{Hess} f(x) := D^2 f(x) := \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}.$$

Since $f \in C^2(U)$, by Theorem 2.3.1

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}, \quad 1 \le i, j \le n,$$

so that $D^2 f(x)$ is a *symmetric* matrix. By *Taylor's approximation* of the 2nd order

$$f(x^* + h) = f(x^*) + \langle \mathrm{grad} f(x^*), h \rangle_{\mathbb{R}^n} + \frac{1}{2} \langle h, \mathrm{Hess} f(x^*) \cdot h \rangle_{\mathbb{R}^n} + o(\|h\|^2), \;\; h \to 0.$$

Since $\nabla f(x^*) = 0$,

$$f(x^* + h) \sim f(x^*) + \frac{1}{2} \langle h, \mathrm{Hess} f(x^*) \cdot h \rangle_{\mathbb{R}^n}, \;\; h \to 0.$$

If $\mathrm{Hess} f(x^*)$ is a **negative definite** matrix, i.e.,

$$\langle y, \mathrm{Hess} f(x^*)\, y \rangle_{\mathbb{R}^n} < 0 \text{ for all } 0 \ne y \in \mathbb{R}^n,$$

then $f(x^* + h) < f(x^*)$, i.e., x^* is a **strict local max**.
If $\mathrm{Hess} f(x^*)$ is a **positive definite** matrix, i.e.,

$$\langle y, \mathrm{Hess} f(x^*)\, y \rangle_{\mathbb{R}^n} > 0 \text{ for all } 0 \ne y \in \mathbb{R}^n,$$

then $f(x^* + h) > f(x^*)$, i.e., x^* is a **strict local min**.

We summarize the above analysis in the following theorem:

Theorem 2.11.2 (Sufficient Conditions for Local Extrema)**.** *Let $U \subset \mathbb{R}^n$ be open, the function $f : U \to \mathbb{R}$ be* ***twice continuously differentiable*** *on U, and let $x^* \in U$ obey $\nabla f(x^*) = 0$. Then:*

(i) $\mathrm{Hess} f(x^*)$ *is* ***positive definite*** *(i.e.,* $\mathrm{Hess} f(x^*) > 0$ *as a symmetric $n \times n$ matrix)* $\Longrightarrow x^*$ *is a* ***strict local min****.*

The positive definiteness of $\mathrm{Hess} f(x^*)$ *is* ***equivalent*** *to the* ***positivity*** *of all n* ***leading principal minors*** *of $D^2 f(x^*)$:*

$$\partial^2_{1,1} f(x^*) > 0, \quad \begin{vmatrix} \partial^2_{1,1} f(x^*) & \partial^2_{1,2} f(x^*) \\ \partial^2_{2,1} f(x^*) & \partial^2_{2,2} f(x^*) \end{vmatrix} > 0,$$

$$\begin{vmatrix} \partial^2_{1,1} f(x^*) & \partial^2_{1,2} f(x^*) & \partial^2_{1,3} f(x^*) \\ \partial^2_{2,1} f(x^*) & \partial^2_{2,2} f(x^*) & \partial^2_{2,3} f(x^*) \\ \partial^2_{3,1} f(x^*) & \partial^2_{3,2} f(x^*) & \partial^2_{3,3} f(x^*) \end{vmatrix} > 0, \; \ldots \;, \; |D^2 f(x^*)| = \det D^2 f(x^*) > 0.$$

(ii) $\mathrm{Hess} f(x^*)$ *is* ***negative definite*** *(i.e.,* $\mathrm{Hess} f(x^*) > 0$ *as a symmetric $n \times n$ matrix)* $\Longrightarrow x^*$ *is a* ***strict local max****.*

The negative definiteness of $\mathrm{Hess} f(x^*)$ *means that the* ***leading principal minors alternate*** *in sign:*

$$\partial^2_{1,1} f(x^*) < 0, \quad \begin{vmatrix} \partial^2_{1,1} f(x^*) & \partial^2_{1,2} f(x^*) \\ \partial^2_{2,1} f(x^*) & \partial^2_{2,2} f(x^*) \end{vmatrix} > 0,$$

$$\begin{vmatrix} \partial^2_{1,1} f(x^*) & \partial^2_{1,2} f(x^*) & \partial^2_{1,3} f(x^*) \\ \partial^2_{2,1} f(x^*) & \partial^2_{2,2} f(x^*) & \partial^2_{2,3} f(x^*) \\ \partial^2_{3,1} f(x^*) & \partial^2_{3,2} f(x^*) & \partial^2_{3,3} f(x^*) \end{vmatrix} < 0, \; \ldots \;, \; (-1)^n \left|D^2 f(x^*)\right| > 0.$$

(iii) $\mathrm{Hess} f(x^*)$ *is* ***indefinite****, i.e., for some vectors $y_1 \neq 0$, $y_2 \neq 0$*

$$\langle y_1, \mathrm{Hess} f(x^*) y_1 \rangle_{\mathbb{R}^n} > 0 \quad \textit{but} \quad \langle y_2, \mathrm{Hess} f(x^*) y_2 \rangle_{\mathbb{R}^n} < 0,$$

$\Longrightarrow x^*$ *is* ***not*** *a local extremum (i.e., x^* is a* ***saddle*** *point)*

Remark 2.11.3. *A saddle point x^* is a min of f in some direction $h_1 \neq 0$ and a max of f in other direction $h_2 \neq 0$ (such that $\langle h_1, \mathrm{Hess} f(x^*) h_1 \rangle_{\mathbb{R}^n} > 0, \langle h_2, \mathrm{Hess} f(x^*) h_2 \rangle_{\mathbb{R}^n} < 0$).*

Warning: The **positive semidefiniteness** $\mathrm{Hess} f(x^*) \geq 0$, i.e.,

$$\langle y, \mathrm{Hess} f(x^*)\, y \rangle_{\mathbb{R}^n} \geq 0 \quad \text{for all} \quad y \in \mathbb{R}^n,$$

or the **negative semidefiniteness** $\mathrm{Hess} f(x^*) \leq 0$, i.e.,

$$\langle y, \mathrm{Hess} f(x^*)\, y \rangle_{\mathbb{R}^n} \leq 0 \quad \text{for all} \quad y \in \mathbb{R}^n,$$

does **not** imply in general that x^* is a local (*non-strict*) minimum, or respectively, maximum. Now we cannot ignore the terms $o(\|h\|^2)$ in Taylor's formula.

Unlike Theorem 2.10.1, the conditions of Theorem 2.11.2 are **not necessary** conditions! Remember a standard **Counterexample:**

$$f_1(x) = x^4, \quad f_2(x) = -x^4,$$

$$f_1'(0) = f_1''(0) = 0, \quad f_2'(0) = f_2''(0) = 0.$$

But f_1 (resp. f_2) has a **strict** global min (rep. max) at $x = 0$.

Numerical Examples: $f : \mathbb{R}^2 \to \mathbb{R}, (x, y) \to f(x, y)$

$$(\mathbf{i}) \ \ f(x, y) := x^2 + y^2,$$

$$\nabla f(x) = (2x, 2y) = 0 \iff x = y = 0.$$

$$D^2 f(0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \text{ the same for all } (x, y),$$

$$\det D^2 f(0) = 4 > 0$$

Answer: $(0, 0)$ is a **strict local min**.

$$(\mathbf{ii}) \ f(x, y) := x^2 - y^2,$$

$$\nabla f(x) = (2x, -2y) = 0 \iff \quad x = y = 0.$$

$$D^2 f(0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix},$$

$$\det D^2 f(0) = -4 < 0.$$

Answer: $(0, 0)$ is a **saddle point**.

(iii) Hess$f(x^*)$ is **semidefinite**, but we cannot say something about critical points. Consider functions

$$f_1(x, y) := x^2 + y^4, \ f_2(x, y) := x^2,$$
$$f_3(x, y) := x^2 + y^3.$$

For each $i = 1, 2, 3$, we have $f_i(0) = 0, \nabla f(0) = 0$,

$$\text{Hess} f(0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \text{ is positive semidefinite,}$$

i.e., $\langle h, \text{Hess} f(x^*) h \rangle_{\mathbb{R}^n} \geq 0$ for any $h \in \mathbb{R}^2$.

But, the point $(0, 0)$ is:

(1) *strict local* min for f_1;

(2) *a non-strict local* min for f_2 (since $f_2(0,y) = 0, \forall y \in \mathbb{R}$);

(3) *not a local extremum* for f_3 ($f_3(t,0) = t^2 > 0, f_3(0,t) = t^3 < 0$ if $t < 0$).

Reminder from Linear Algebra:

Proposition 2.11.4. *A symmetric* 2×2 *matrix*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \ a_{12} = a_{21},$$

is ***positive definite*** *if and only if*

$$a_{11} > 0 \text{ and } \det A := a_{11}a_{22} - a_{12}^2 > 0.$$

The matrix A is ***negative definite*** *if and only if*

$$a_{11} < 0 \text{ and } \det A = a_{11}a_{22} - a_{12}^2 > 0.$$

If $\det A < 0$*, the matrix A is surely* ***indefinite****.*

Indeed, for any vector $y = (y_1, y_2) \in \mathbb{R}^2$:

$$\langle Ay, y \rangle = a_{11}y_1^2 + a_{22}y_2^2 + 2a_{12}y_1y_2.$$

Let us assume that $y_2 \neq 0$ and set $z = y_1/y_2$, then the quadratic polynomial

$$\frac{\langle Ay, y \rangle}{y_2^2} = P(z) = a_{11}z^2 + 2a_{12}z + a_{22}, \ z \in \mathbb{R},$$

takes only positive (resp. negative) values for all $z \in \mathbb{R}$ **iff** its discriminant $\Delta := a_{12}^2 - a_{11}a_{22} = -\det A < 0$.

2.12 A Rough Guide: How to Find the Global Maxima/Minima

Problem: to find **global maxima** (minima) for

$$f : D \to \mathbb{R}, \ D \subseteq \mathbb{R}^n \text{ (arbitrary set, not necessary open).}$$

(i) Find and compare the local maxima (minima) in int$D$ – interior of D – and choose the **best**.

(ii) Compare with the **boundary values** $f(x)$, $x \in D \setminus \text{int}D$.

Numerical Example: Find the max/min of

$$f(x) = 4x^3 - 5x^2 + 2x \quad \text{over } x \in [0,1].$$

Since $I := [0,1]$ is **compact** and f is **continuous** on I, the **Weierstrass theorem** guarantees that f has a **global max** on this interval. There are 2 possibilities: either the maximum is a local maximum attained on the open interval $(0,1)$, or it occurs at one of the boundary points $x = 0, 1$. In the first case we should meet the 1st order condition:

$$\begin{aligned} & f'(x) = 12x^2 - 10x + 2 = 0 \\ \implies & x_1 = 1/2 \text{ or } x_2 = 1/3. \end{aligned}$$

So, we have two critical points x_1 and x_2. The 2nd order condition says that

$$\begin{aligned} & f''(x) = 24x - 10 \\ \implies \quad & f''(x_1) = 2 > 0 \text{ and } f''(x_2) = -2 < 0. \end{aligned}$$

Thus, $x = 1/2$ is local min and $x = 1/3$ is local max. Evaluating f at the four points 0, 1/3, 1/2, and 1 shows that

$$f(0) = 0,\ f(1/3) = 7/27,\ f(1/2) = 1/4,\ f(1) = 1;$$

so $x = 1$ is the global max resp. $x = 1/2$ is the global min for $f(x)$, $x \in [0,1]$.

Literature: Chapters 16, 17 of C. Simon, L. Blume *"Mathematics for Economists"*.

Example 2.12.1 (Economical Example: Cobb–Douglas Function)**.**
Cobb–Douglas production function: $f(x,y) = x^a y^b, x, y > 0$.
Find the maximum of the ***profit*** $V(x,y) = px^a y^b - k_x x - k_y y$. *1st order conditions:*

$$\begin{cases} pax^{a-1}y^b = k_x, \\ pbx^a y^{b-1} = k_y. \end{cases} \tag{$*$}$$

After dividing the 1st line by the 2nd one, we get

$$\frac{a}{b} \cdot \frac{y}{x} = \frac{k_x}{k_y} \implies y = \frac{bk_x}{ak_y} x.$$

Putting back in $(*)$*, we have*

$$k_x = pax^{a-1} \left(\frac{bk_x}{ak_y} x \right)^b = pa^{1-b} b^b \left(\frac{k_x}{k_y} \right)^b x^{a+b-1}$$

which allows us to find a unique critical point (x^*, y^*)

$$\begin{aligned} x^* &= \left(\frac{k_x^{1-b} k_y^b}{p a^{1-b} b^b} \right)^{\frac{1}{a+b-1}} = \frac{p^{\frac{1}{1-(a+b)}} a^{1-\frac{a}{1-(a+b)}} b^{\frac{b}{1-(a+b)}}}{k_x^{1-\frac{a}{1-(a+b)}b} k_y^{1-\frac{b}{1-(a+b)}}}, \\ y^* &= \frac{b k_x}{a k_y} x^*. \end{aligned}$$

Is it a maximum? Calculate

$$\text{Hess} V(x,y) = \text{Hess} f(x,y) = p \begin{pmatrix} a(a-1)x^{a-2}y^b & abx^{a-1}y^{b-1} \\ abx^{a-1}y^{b-1} & b(b-1)x^a y^{b-2} \end{pmatrix},$$
$$\det \text{Hess} V(x,y) = \left[a(a-1)b(b-1) - a^2b^2 \right] x^{2a-2} y^{2b-2} > 0$$

if $(a-1)(b-1) > ab$ *or* $a + b < 1$*. We also have that*

$$\frac{\partial^2 f}{\partial x^2}(x,y) < 0 \textit{ if } a < 1.$$

So, a sufficient condition for max is $a + b < 1$*.*

2.13 Envelope Theorems

The **Envelope Theorem** (Umhüllenden-Theorem) is a general principle describing how the optimal value of the objective function in a **parametrized** optimization problem changes as the parameters of the problem change. In economics, such parameters can be prices, tax rates, income levels, etc. Such problems constitute the subject of **Comparative Statistics**.

In microeconomic theory, the envelope theorem is used, e.g., to prove *Hotelling's lemma* (1932), *Shepard's lemma* (1953) and *Roy's identity* (1947).

In applications, it is usually stated *non-rigorously*, i.e., without the suitable assumptions which guarantee the *differentiability* of the so-called optimal value function.

Let

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

be a **continuously differentiable** function. We call it the **objective function** $f(x, \alpha)$, it depends on the choice variable $x \in \mathbb{R}^n$ and the parameter $\alpha \in \mathbb{R}^m$. We consider the unconstrained maximization problem for f, i.e.,

$$\textbf{maximize } f(x; \alpha) \text{ w.r.t. } x \in \mathbb{R}^n.$$

Let $x^*(\alpha) \in \mathbb{R}^n$ be a solution of the above problem, i.e.,

$$f(x^*(\alpha); \alpha) \geq f(x; \alpha) \text{ for all } x \in \mathbb{R}^n.$$

Here we *assume* that, at each $\alpha \in \mathbb{R}^m$, such a solution $x^*(\alpha) \in \mathbb{R}^n$ *exists*;
in the case of non-uniqueness we take for $x^*(\alpha)$ *any one* of the maximum points x for $f(x;\alpha)$. Then,

$$V(\alpha) := \max_{x \in \mathbb{R}^n} f(x;\alpha) = f(x^*(\alpha);\alpha)$$

is the corresponding (**optimal**) **value function**.

We are interested in how $V(\alpha)$ depends on $\alpha \in \mathbb{R}^m$.

Note that $V(\alpha) = f(x^*(\alpha);\alpha)$ changes for 2 reasons:

(i) **directly** w.r.t. α, because α is the 2nd variable in $f(x;\alpha)$;

(ii) **indirectly**, since $x^*(\alpha)$ itself nontrivially depends on α.

Theorem 2.13.1 (Envelope Theorem)**.** *Suppose that $f(x;\alpha)$ is **continuously differentiable** w.r.t. $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}^m$. Suppose **additionally** that $x^*(\alpha)$ is a **continuously differentiable** function of $\alpha \in \mathbb{R}^m$. Then $V(\alpha)$ is also **continuously differentiable** and for any $\alpha \in \mathbb{R}^m$ and $1 \le i \le m$*

$$\frac{\partial V}{\partial \alpha_i}(\alpha) = \frac{\partial f}{\partial \alpha_i}(x^*(\alpha);\alpha).$$

Proof. By our assumption we have

$$V(\alpha) = f(x^*(\alpha);\alpha), \ \forall \alpha \in \mathbb{R}^m.$$

Therefore, by the chain rule

$$\frac{\partial V}{\partial \alpha_i}(\alpha) = \frac{\partial f}{\partial \alpha_i}(x^*(\alpha);\alpha) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x^*(\alpha);\alpha)\frac{\partial x_j}{\partial \alpha_i}(\alpha), \quad 1 \le i \le m.$$

The second sum vanishes since by the 1st order condition for extrema (cf. Theorem 2.10.1)

$$\frac{\partial f}{\partial x_j}(x^*(\alpha);\alpha) = 0, \quad \text{for all } 1 \le j \le n.$$

Thus we get

$$\frac{\partial V}{\partial \alpha_i}(\alpha) = \frac{\partial f}{\partial \alpha_i}(x^*(\alpha);\alpha).$$

□

Remark 2.13.2. *The same inequality holds if we minimize $f(x;\alpha)$.*

Simplified rule: When calculating $\partial V/\partial \alpha_i$, just forget the $\max_{x\in\mathbb{R}^n}$ and take the derivatives of $f(x;\alpha)$ w.r.t. α_i, and then plug in the optimal solution $x^*(\alpha)$. So, we need to **consider only the direct effect** of α on $V(\alpha)$, **ignoring the indirect effect** of $x^*(\alpha)$.

At this point it would be useful to know when $x^*(\alpha)$ **exists** and is **continuously differentiable** w.r.t. α. To answer this question we can use the **Implicit Function Theorem (IFT)**.

Assume that $f \in C^2(\mathbb{R}^n \times \mathbb{R}^m)$. We know that $x^*(\alpha)$ is a solution to

$$\nabla_x f(x,\alpha) = 0$$

(the necessary condition for extrema), i.e.,

$$\begin{cases} \frac{\partial f}{\partial x_1}(x,\alpha) & = 0, \\ \quad \cdots & \cdots \\ \frac{\partial f}{\partial x_n}(x,\alpha) & = 0. \end{cases}$$

Consider a function

$$\begin{aligned} g \ &: \ \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n, \\ &(x;\alpha) \to \left(\frac{\partial f}{\partial x_j}(x;\alpha)\right)_{1\le j\le n}. \end{aligned}$$

The **IFT** (cf. Theorem 2.8.1) tells us that $x^*(\alpha)$ exists as an implicit function and is continuously differentiable w.r.t. α if the $n \times n$-matrix of partial derivatives of g w.r.t. $x = (x_1, \ldots, x_n)$ is **invertible**, i.e., $\det D_x g(x,\alpha) \neq 0$, where

$$D_x g = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_n}{\partial x_1} & & \frac{\partial g_n}{\partial x_1} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix},$$

$$D_x g = \mathrm{Hess}_x f = D_x^2 f.$$

Assume that $\mathrm{Hess}_x f$ at point $(x^*(\alpha);\alpha)$ is a negative definite matrix (which is the sufficient condition for a strict local maximum w.r.t. x). Hence $\det D_x g(x^*(\alpha);\alpha) > 0$ if $n = 2,4,6,\ldots$ (or < 0 if respectively, $n = 1,3,5,\ldots$). These arguments lead to the following result.

Theorem 2.13.3 (Deep Envelope Theorem, Sammuelson (1947), Auspitz–Lieben (1889))**.** *Let $U_1 \subset \mathbb{R}^n$ and $U_2 \subset \mathbb{R}^m$ be open domains and let*

$$f : U_1 \times U_2 \to \mathbb{R}, \ (x,\alpha) \to f(x;\alpha),$$

*be **twice continuously differentiable** (i.e., $f \in C^2(U_1 \times U_2)$). Suppose that $\operatorname{Hess}_x f(x;\alpha)$ is **negative definite** for all $x \in U_1$, $\alpha \in U_2$. Fix some $\alpha \in U_2$, and let $x^*(\alpha) \in U_1$ be a maximum of $f(x;\alpha)$ on U_1, i.e.,*

$$f(x^*(\alpha);\alpha) = \max_{x \in U_1} f(x;\alpha)$$

$$\Downarrow \quad \textit{which, by Theorem 2.11.2, implies} \quad \Downarrow$$

$$\nabla_x f((x^*(\alpha);\alpha) = 0.$$

*Then there exists a **continuously differentiable** function $x^* : V_2 \to \mathbb{R}^n$ defined on some open set $V_2 \subseteq U_2$ such that*

$$V(\alpha) := \max_{x \in U_1} f(x;\alpha) = f(x^*(\alpha);\alpha)$$

$$\textit{and} \quad \frac{\partial V}{\partial \alpha_i}(\alpha) = \frac{\partial f}{\partial \alpha_i}(x^*(\alpha);\alpha).$$

Geometrical picture: The curve $\mathbb{R}^m \ni \alpha \mapsto y = V(\alpha) := f(x^*(\alpha);\alpha)$ is the **envelope** of the family of curves $\mathbb{R}^m \ni \alpha \mapsto y = V_x(\alpha) := f(x;\alpha)$, indexed by the parameter $x \in \mathbb{R}^n$. Indeed, for each x and α we have

$$f(x;\alpha) \leq V(\alpha).$$

None of the $V_x(\alpha)-$curves can lie above the curve $y = V(\alpha)$. On the other hand, for each value of α there exists at least one value $x^*(\alpha)$ of x such that $f(x^*(\alpha);\alpha) = V(\alpha)$. The curve $\alpha \mapsto V_{x^*(\alpha)}(\alpha)$ will just **touch** the curve $\alpha \mapsto y = V(\alpha)$ at the point $(x^*(\alpha), V(\alpha))$, and so must have exactly the **same tangent** as the graph of V at this point, i.e.,

$$\frac{\partial V}{\partial \alpha_i}(\alpha) = \frac{\partial f}{\partial \alpha_i}(x^*(\alpha);\alpha).$$

So, the graph of $V(\alpha)$ is like an **envelope** that is used to "wrap" or cover all the curves $y = V_x(\alpha)$.

Example 2.13.4 (Hotelling's Lemma)**.** *A competitive firm cannot change:*

*(i) **output prices** p (if you increase p, you lose customers);*

*(ii) **wages** w (workers will go to other firms).*

*But the firm can chose x – the **number of workers** it uses. Let $f(x)$ is the corresponding **production function**. The **profit** of the firm at given x, p, w is given by*

$$\pi(x;p,w) = pf(x) - wx.$$

*The **maximum profit function** (also called the firm's profit function)*

$$V(p, w) = \max_{x \geq 0} \{pf(x) - wx\}.$$

It is important to know how the profit of the firm changes if p, w change:

$$\frac{\partial V}{\partial p}, \quad \frac{\partial V}{\partial w} \quad ?$$

By the Envelope Theorem, if the model is "nice" (i.e., we have a continuously differentiable function $x^*(p, w)$), then formally

$$\frac{\partial V}{\partial p} = f(x^*(p, w)),$$
$$\frac{\partial V}{\partial w} = -x^*(p, w),$$

where $x^*(p, w)$ is the optimal number of workers.

***Conclusion:** when wages are increasing, the maximum profit will be decreasing proportionally to the number of workers.*
Formally x^* obeys

$$g(x, w, p) = pf'(x^*) - w = 0$$

By the IFT, a "nice" solution exists if $f''(x^*) < 0$.

2.14 Gâteaux and Fréchet Differentials

The notions of directional and total differentiability can be naturally extended to infinite dimensional spaces.

Let $(X, \| \cdot \|)$ be a normed space, $U \subset X$ – open set and $f : U \to \mathbb{R}$.

Definition 2.14.1 (Gâteaux differentiability)**.** *The function $f : U \to \mathbb{R}$ is Gâteaux differentiable at a point $x \in U$ along direction $v \in X, \|v\| = 1$, if the following limit exists:*

$$\lim_{t \to 0} \frac{1}{t} \left[f(x + tv) - f(x) \right] =: D_v f(x).$$

*$D_v f(x) \in X$ is called the **Gâteaux derivative**.*

Definition 2.14.2 (Fréchet differentiability)**.** *The function $f : U \to \mathbb{R}$ is Fréchet differentiable at a point $x \in U$ if there exists a linear continuous mapping $Df(x) : X \to X$ such that*

$$\lim_{\|h\| \to 0} \frac{1}{\|h\|} \left[f(x + h) - f(x) - Df(x)h \right] = 0;$$

*$Df(x) \in \mathcal{L}(X, X)$ is called the **Fréchet derivative**.*

Fréchet differentiability $\Longrightarrow$ Gâteaux differentiability along all directions $v \in X, \|v\| = 1$.

Proposition 2.14.3 (Sufficient condition for Fréchet differentiability)**.** *If all directional derivatives*

$$D_v f(x), \quad \forall v \in X, \quad \|v\| = 1,$$

exist in all points $x \in U$ and can be represented as

$$D_v f(x) = L(x)v$$

with a linear bounded operator $L(x) : X \to X$ and the mapping

$$U \ni x \to L(x) \in \mathcal{L}(X, X)$$

*is **continuous** (in the operator norm), then $f : U \to \mathbb{R}$ is also Fréchet differentiable at all points $x \in U$ and*

$$Df(x) = L(x).$$

Proposition 2.14.4 (Necessary condition for extrema)**.** *If f has a local extrema in U, then each $D_v f(x) = 0$ for $v \in X, \|v\| = 1$, (provided this directional derivative exists).*