Lecture Notes on Multivariable Calculus

Notes written by Barbara Niethammer and Andrew Dancer

Lecturer Jan Kristensen

Trinity Term 2018

1 Introduction

In this course we shall extend notions of differential calculus from functions of one variable to more general functions

$$f\colon \mathbb{R}^n \to \mathbb{R}^m. \tag{1.1}$$

In other words functions $f = (f_1, \ldots, f_m)$ with m components, all depending on n variables $x_1, \ldots, x_n$. In fact, we shall take it one step further and consider maps

$$f\colon X \to Y$$

where X and Y are normed vector spaces over $\mathbb{R}$. In this course we restrict attention to the case where X and Y are *finite* dimensional, and so by a choice of bases in X and in Y we reduce this more general looking case to (**??**). It is however often convenient to have this *invariant* definition.

The questions we will be interested in studying include the following.

- If $n = m$, under what circumstances is f invertible?
- If $n \geq m$, when does the equation $f(x, y) = 0$, with $x \in \mathbb{R}^{n-m}$ and $y \in \mathbb{R}^m$, implicitly determine y as a function of x ?
- Is the zero locus $f^{-1}(0)$ a smooth subset of $\mathbb{R}^n$ in a suitable sense, for example a smooth surface in $\mathbb{R}^3$?

The first major new idea is to define the derivative at a point as a linear map, which we can think of as giving a first-order approximation to the behaviour of the function near that point. A key theme will be that, subject to suitable nondegeneracy assumptions, the derivative at a point will give qualitative information about the function on a neighbourhood of the point. In particular, the Inverse Function Theorem will tell us that invertibility of the derivative at a point (as a linear map) will actually guarantee local invertibility of the function in a neighbourhood.

The results of this course are foundational for much of mathematics and the notion of a *smooth manifold* is in particular central to mathematical physics and geometry. A smooth submanifold in $\mathbb{R}^n$ is, intuitively, a generalisation to higher dimensions of the notion of a smooth surface in $\mathbb{R}^3$. Among many other things we shall use our theorems to obtain a criterion for when the locus defined by a system of nonlinear equations is a manifold. Manifolds are the setting for much of higher-dimensional geometry and mathematical physics and in fact the concepts of differential (and integral) calculus that we study in this course can be developed on general manifolds. The Part B course Geometry of Surfaces and the Part C course Differentiable Manifolds develop these ideas further.

2 Differentiation of functions of several variables

2.1 Introduction

In this chapter we will extend the concept of differentiability of a function of one variable to the case of a function of several variables. We first recall the definitions for a function of one variable.

Differentiability of a function of one variable

Let $I \subseteq \mathbb{R}$ be an open interval. A function $f\colon I \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable at $x \in I$ if

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

exists. Equivalently we can say that f is differentiable in $x \in I$ if there exists a linear map* $L\colon \mathbb{R} \to \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{f(x+h)-f(x)-Lh}{h} = 0. \tag{2.1}$$

In this case, L is given by $L : h \mapsto f'(x)\cdot h$.

Another way of writing (**??**) is

$$f(x+h)-f(x)-Lh = R_f(h) \qquad \text{with } R_f(h) = o(|h|), \text{ i.e. } \lim_{h\to 0}\frac{R_f(h)}{|h|} = 0. \tag{2.2}$$

This definition is more suitable for the multivariable case, where h is now a vector, so it does not make sense to divide by h.

Differentiability of a vector-valued function of one variable

Completely analogously we define the derivative of a vector-valued function of one variable. More precisely, if $f\colon I \subseteq \mathbb{R} \to \mathbb{R}^m, m > 1$, with components $f_1, \dots, f_m$, we say that f is differentiable at $x \in I$ if

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

exists.

*Here and in what follows, we will often write Lh instead of $L(h)$ if L is a linear map.

It is easily seen that f is differentiable at $x \in I$ if and only if $f_i \colon I \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable in $x \in I$ for all $i = 1, \dots, m$. Also, f is differentiable in $x \in I$ if and only if there exists a linear map $L \colon \mathbb{R} \to \mathbb{R}^m$ such that

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

How can we now generalize the concept of differentiability to functions of several variables, say for a function $f \colon \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$, $f = f(x, y)$? A natural idea is to freeze one variable, say y, define $g(x) = f(x, y)$ and check whether g is differentiable at x. This will lead to the notion of partial derivatives and most of you have seen this already in lectures in the first year, e.g. in Calculus. However, we will see that the concept of partial derivatives alone is not completely satisfactory. For example we will see that the existence of partial derivatives does not guarantee that the function itself is continuous (as it is the case for a function of one variable).

The notion of the (total) derivative for functions of several variables will not have this deficiency. It is based on a generalisation of the formulation in (**??**). In order to do that we will need a suitable norm (length function) on $\mathbb{R}^n$. You may have learned already, e.g. in Topology, that all norms on $\mathbb{R}^n$ are equivalent, and hence properties of sets, such as openness or boundedness, and of functions, such as continuity, do not depend on the choice of the norm.

In the sequel we will always use the *Euclidean norm* on $\mathbb{R}^n$ and denote it by $|\cdot|$. More precisely, for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ we denote

$$|x| = \sqrt{x_1^2 + \dots x_n^2}\,.$$

You may check yourself that this defines a norm (and hence a metric). For the proof of the triangle inequality you will need to use the Cauchy-Schwarz inequality.

We shall also use the matrix (Hilbert-Schmidt) norm

$$\|C\| = \left(\sum_{i,j=1}^{n} C_{ij}^2 \right)^{\frac{1}{2}}$$

on the space of $n \times n$ real matrices. We have the following useful inequality:

$$|Ch| = \Big(\sum_i \Big(\sum_j C_{ji} h_j \Big)^2 \Big)^{\frac{1}{2}} \leq \Big(\sum_i \Big(\sum_j C_{ij}^2 \Big) \Big(\sum_j h_j^2 \Big) \Big)^{\frac{1}{2}} = |h|\, \|C\|.$$

We shall also occasionally use the fact that

$$\|AB\| \leq \|A\| \|B\|.$$

2.2 Partial derivatives

We are going to consider[†] functions $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Here and in what follows we always assume that Ω is an open and connected subset of $\mathbb{R}^n$ (a domain).

We first consider a few examples of such functions.

Example 2.2.1. a) The function $f\colon \mathbb{R}^n \to \mathbb{R}$ given by

$$x \mapsto |x| = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}.$$

b) The function $g : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$g\colon (x,y,z) \mapsto \begin{pmatrix} x^3 + y^3 + z^3 - 7 \\ xy + yz + zx + 2 \end{pmatrix}.$$

c) The electric field in a vacuum induced by a point charge q in a point $x_0 \in \mathbb{R}^3$ is given by

$$f\colon \mathbb{R}^3 \setminus \{x_0\} \to \mathbb{R}^3, \qquad f(x) = q\frac{x - x_0}{|x - x_0|^3}.$$

d) $f\colon \mathbb{C} \to \mathbb{C}, f(z) = e^z$ has – with the usual identifications – the real representation $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$f(x,y) = \begin{pmatrix} e^x \cos y \\ e^x \sin y \end{pmatrix}.$$

e) $f\colon \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ where we identify $\mathbb{R}^{n^2}$ with $M_{n\times n}(\mathbb{R})$, the vector space of $n \times n$ matrices, and define f by $f : A \mapsto A^2$.

We shall sometimes use the concepts of *graphs* and *level sets*. Let $f\colon \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function. Then the graph of f, given by

$$\Gamma_f = \{(x,y) \in \Omega \times \mathbb{R} \mid y = f(x)\},$$

is usually a surface in $\mathbb{R}^3$. Its level set at level $c \in \mathbb{R}$ is

$$N_f(c) = \{x \in \Omega \mid f(x) = c\},$$

which is usually a curve in Ω.

The concepts of graphs and level sets generalise in an obvious way to functions $f\colon \Omega \to \mathbb{R}$ defined on domains Ω in $\mathbb{R}^n$ and beyond.

[†]We will use the shorthand $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$' to mean that "$\Omega$ is a domain in $\mathbb{R}^n$ and $f\colon \Omega \to \mathbb{R}^m$ is a function".

Definition 2.2.2. (Partial derivative) Let $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$. We say that the i-th partial derivative of f in $x \in \Omega$ exists, if

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t\to 0} \frac{f(x+te_i) - f(x)}{t}$$

exists, where e_i is the i-th unit vector. In other words, the i-th partial derivative is the derivative of $g(t) = f(x+te_i)$ at $t = 0$.

Other common notations for the i-th partial derivative of f at x are

$$\partial_i f(x), \quad D_i f(x), \quad \partial_{x_i} f(x), \quad f_{x_i}(x) \quad \text{or} \quad f_{,i}(x).$$

We will mostly use $\partial_i f(x)$. If $f\colon \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$ we often write $f = f(x,y)$ and denote the partial derivatives by $\partial_x f$ and $\partial_y f$ respectively.

Example 2.2.3. a) Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = |x|$. Then for $x \neq 0$ we have

$$\frac{1}{t}\big(|x+te_i| - |x|\big) = \frac{1}{t}\frac{|x+te_i|^2 - |x|^2}{|x+te_i| + |x|} = \frac{1}{t}\frac{2tx_i + t^2}{|x+te_i| + |x|} = \frac{2x_i + t}{|x+te_i| + |x|} \to \frac{x_i}{|x|}$$

as $t \to 0$. Hence, for $x \neq 0$, the function f has partial derivatives, given by $\partial_i f(x) = \frac{x_i}{|x|}$. Notice that no partial derivative of f exists at $x = 0$.

b) Let $f(x) = g(r)$ with $r(x) = |x|$ and differentiable $g\colon [0,\infty) \to \mathbb{R}$. Then, for $x \neq 0$, by the Chain Rule from Prelims Analysis 2, we find

$$\partial_i f(x) = g'(r)\partial_i r(x) = g'(r)\frac{x_i}{|x|} = \frac{g'(r)}{r}x_i\,.$$

The following example shows that, surprisingly, functions whose partial derivatives all exist are in general not continuous.

Example 2.2.4. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$

$$f(x,y) = \begin{cases} \dfrac{xy}{(x^2+y^2)^2} & \text{for } (x,y) \neq (0,0), \\[2ex] 0 & \text{for } (x,y) = (0,0). \end{cases}$$

- f has partial derivatives on $\mathbb{R}^2 \setminus \{(0,0)\}$ with

$$\partial_x f(x,y) = \frac{y}{(x^2+y^2)^2} - \frac{4x^2y}{(x^2+y^2)^3}$$

  and a similar expression for $\partial_y f(x,y)$.

- f has partial derivatives at 0, since f is identically zero on the x and y axes. Explicitly, for all $t \neq 0$ we have

$$\frac{f((0,0)+t(1,0)) - f(0,0)}{t} = \frac{f(t,0) - f(0,0)}{t} = \frac{0-0}{t} = 0$$

  and thus $\partial_x f(0) = 0$; similarly $\partial_y f(0) = 0$.

- **But:** f is not continuous at $(x, y) = (0, 0)$. To see this, consider the behaviour of the function on the line $\{(t, t) : t \in \mathbb{R}\}$. On this line, $f(t, t) = \frac{1}{4t^2}$ which tends to ∞ as $t \to 0$.

This shows that existence of partial derivatives is *not* the correct notion of differentiability in higher dimensions. We shall see later that the correct higher dimensional version of differentiability, using the ideas concerning linear maps from the beginning of this chapter, *will* imply continuity of the function. We shall also see that functions with *continuous* partial derivatives are differentiable in this correct sense and hence are also continuous. In our example above, the partial derivatives $\partial_i f(x)$ are not continuous at $x = 0$.

Before we define differentiability, we shall make some more remarks about partial derivatives.

The partial derivative is a special case of the *directional derivative* which we will now define.

Definition 2.2.5. (Directional derivative) Suppose that $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $v \in \mathbb{R}^n \backslash \{0\}$. If

$$\partial_v f(x) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t}$$

exists, we call it the *directional derivative* of f in direction v at the point $x \in \Omega$.

Observe that if v is one of the unit coordinate vectors e_i, then we recover the notion of partial derivative.

Example 2.2.6. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = |x|$ and let $x, v \in \mathbb{R}^n \setminus \{0\}$. Then

$$\begin{aligned}
\partial_v f(x) &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} |x + tv| = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \Big(\sum_{i=1}^n |x_i + tv_i|^2 \Big)^{\frac{1}{2}} \\
&= \frac{1}{2|x|} \sum_{i=1}^n 2x_i v_i = \sum_{i=1}^n \frac{x_i}{|x|} v_i = \Big\langle \frac{x}{|x|}, v \Big\rangle ,
\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual scalar product in $\mathbb{R}^n$. If $v = e_i$ we recover the formula $\partial_i |x| = \frac{x_i}{|x|}$ from Example **??**.

Definition 2.2.7. (Gradient) Let $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ and assume that all partial derivatives exist at $x \in \Omega$. We call the vector field $\nabla f(x) \in \mathbb{R}^n$ given by

$$\nabla f(x) = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix}$$

the *gradient* of f at x.

Note that the directional derivative is related to the gradient via the formula:

$$\partial_v f(x) = \langle \nabla f(x), v \rangle.$$

Example 2.2.8. If $f(x) = |x|$ and $x \neq 0$ then $\nabla f(x) = \begin{pmatrix} \frac{x_1}{|x|} \\ \vdots \\ \frac{x_n}{|x|} \end{pmatrix} = \frac{x}{|x|}$

and $\partial_v f(x) = \langle \nabla f(x), v \rangle$.

Definition 2.2.9. (Partial derivatives for vector-valued functions) Consider a map $f \colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$. We say that the i-th partial derivative of f at $x \in \Omega$ exists, if it exists for all components $f_1, \dots, f_m$. In that case we write

$$\partial_i f(x) = \begin{pmatrix} \partial_i f_1(x) \\ \vdots \\ \partial_i f_m(x) \end{pmatrix}.$$

The following definition will prove useful when we study differentiable functions.

Definition 2.2.10. (Jacobian matrix) Suppose that $f \colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and that all partial derivatives exist at $x \in \Omega$. Then the $(m \times n)$-matrix

$$Df(x) = \begin{pmatrix} \partial_1 f_1(x) & \dots & \partial_n f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \dots & \partial_n f_m(x) \end{pmatrix}$$

is called the *Jacobian matrix* of f in the point x.
If $n = m$ we call $J_f(x) = \det Df(x)$ the *Jacobian determinant* or the *functional determinant* of f at x.

We can write the Jacobian matrix in terms of the gradients of the components:

$$Df(x) = \begin{pmatrix} (\nabla f_1(x))^T \\ \vdots \\ (\nabla f_m(x))^T \end{pmatrix},$$

where the superscript T denotes transposition.

Note for future reference that the Jacobian matrix $Df(x)$ can be consider as a *linear* map from $\mathbb{R}^n$ to $\mathbb{R}^m$: $v \mapsto Df(x)v$.

2.3 Higher partial derivatives

Definition 2.3.1. (Partial derivatives of order k, $C^k(\Omega, \mathbb{R}^n)$) Suppose that we are given a domain $\Omega \subseteq \mathbb{R}^n$.

a) Let $f \colon \Omega \to \mathbb{R}^m$. The *partial derivatives* of order k are inductively defined via

$$\partial_{j_1} \dots \partial_{j_k} f(x) = \partial_{j_1}(\partial_{j_2} \dots \partial_{j_k}) f(x) \quad \text{where} \quad j_1, \dots, j_k \in \{1, \dots, n\}.$$

Notice that $j_1, \dots, j_k$ are not necessarily distinct, and that (a priori) their order is important. A common notation is $\frac{\partial^k f(x)}{\partial_{j_1} \dots \partial_{j_k}}$ or even $\partial_{j_1 j_2 \dots j_k} f(x)$.

b) Let $C^k(\Omega, \mathbb{R}^m)$ be the set of continuous functions $f\colon \Omega \to \mathbb{R}^m$ whose partial derivatives exist up to order k for all $x \in \Omega$ and are continuous in Ω. If $m = 1$, we write $C^k(\Omega)$. Given a domain $\Theta \subset \mathbb{R}^m$, we will sometimes write $C^k(\Omega, \Theta)$ to be the set of functions $f \in C^k(\Omega, \mathbb{R}^m)$ with $f(\Omega) \subseteq \Theta$.

Proposition 2.3.2. (Exchangeability of partial derivatives, Theorem of Schwarz) *Suppose that $f \in C^2(\Omega)$. Then we have for $1 \leq i, j \leq n$ and for any $x \in \Omega$ that*

$$\partial_i \partial_j f(x) = \partial_j \partial_i f(x)$$

Proof. (Not examinable) Let $\partial_j^t f(x) = \frac{f(x+te_j)-f(x)}{t}$ be the difference quotient of f in x_j. By definition

$$\partial_i \partial_j f(x) = \lim_{s\to 0}(\lim_{t\to 0} \partial_i^s \partial_i^t f(x))$$

We need to show that both limits can be interchanged. By the Intermediate Value Theorem we have for all functions $g\colon \Omega \to \mathbb{R}$, for which $\partial_i g(x)$ exists, that

$$\partial_i^s g(x) = \partial_i g(x + \alpha s e_i) \quad \text{for some } \alpha \in (0,1).$$

If we apply this for $g = \partial_j^t f$ and $g = \partial_i^s f$, we get

$$\begin{aligned} \partial_i^s \partial_j^t f(x) &= \partial_i \partial_j^t f(x + \alpha s e_i) \quad \text{for some } \alpha \in (0,1) \\ &= \partial_j^t(\partial_i f(x + \alpha s e_i)) \\ &= \partial_j(\partial_i f(x + \alpha s e_i + \beta t e_j)) \quad \text{for some } \beta \in (0,1) \end{aligned}$$

Since $\partial_j \partial_i f$ is continuous, it follows that

$$\partial_j \partial_i f(x + \alpha s e_i + \beta t e_j) \xrightarrow{s\to 0} \partial_j \partial_i f(x + \beta t e_j) \xrightarrow{t\to 0} \partial_j \partial_i f(x).$$

□

Corollary 2.3.3. *Suppose that $f \in C^k(\Omega)$. Then all partial derivatives up to order k can be interchanged.*

The following example shows that the condition in Proposition **??** that the second partial derivatives must be continuous is indeed necessary.

Example 2.3.4. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} xy\dfrac{x^2 - y^2}{x^2 + y^2} & \text{for } (x,y) \neq (0,0); \\ 0 & \text{for } (x,y) = (0,0). \end{cases}$$

One can show that $f \in C^1(\mathbb{R}^2)$, but $\partial_x \partial_y f(0,0) = 1$ and $\partial_y \partial_x f(0,0) = -1$.

2.4 Differentiability

We will now introduce the notion of the (total) derivative which is based on the idea that the function can be approximated well near a point by a linear map.

Definition 2.4.1. (Differentiable map)

a) We say that a function $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, is *differentiable at* $x \in \Omega$ if there exists a linear map $L\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - Lh}{|h|} = 0$$

We call L the *(total) derivative* of f at x and denote it by $df(x)$.

b) We say that f is *differentiable in* Ω, if f is differentiable at every $x \in \Omega$.

Remark 2.4.2.

a) You might wonder if the total derivative $df(x)$ as defined above is really well-defined: check that if two linear maps L_1, $L_2\colon \mathbb{R}^n \to \mathbb{R}^m$ exist satisfying the condition for differentiability of f at x, then necessarily $L_1 = L_2$.

b) Alternatively we can say that f is differentiable at $x \in \Omega$ if there exists a linear map $L\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(x+h) - f(x) - Lh = R_f(h) \qquad \text{with} \qquad R_f(h) = o(|h|)$$

c) $f\colon \Omega \to \mathbb{R}^m$ is differentiable at $x \in \Omega$ if and only if every component $f_i\colon \Omega \to \mathbb{R}, i = 1, \ldots, m$, is differentiable at $x \in \Omega$.

Intuitively, $f(x)$ is the zero-order approximation to $f(x+h)$, while $f(x) + Lh$ is the first order or linear approximation to $f(x+h)$. The term $R_f(h)$ tends to zero faster than h and can be viewed as a remainder or error term for the linear approximation. This kind of thinking is familiar from the theory of Taylor approximations.

By inspection of Definition **??** you will notice that we do not really have to confine attention to functions between $\mathbb{R}^n$ and $\mathbb{R}^m$. The definition makes perfect sense also for functions between normed vector spaces. This is often useful and we highlight it with a definition:

Definition 2.4.3. (Differentiable map between normed spaces) Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be finite dimensional and normed vector spaces over $\mathbb{R}$. Assume that $f\colon \Omega \to Y$ is a map where Ω is an open subset of X.

A) We say that f is *differentiable at* $x \in \Omega$ if there exists a linear map $L\colon X \to Y$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - Lh\|_Y}{\|h\|_X} = 0.$$

We call L the *(total) derivative* of f at x and denote it by $df(x)$.

B) We say that f is *differentiable in* Ω (or simply differentiable when Ω is clear from the context), if f is differentiable at every $x \in \Omega$.

Since all norms on a finite dimensional vector space are equivalent it is not difficult to check that the notion of differentiability is independent of the chosen norms. This is different if we allowed X, Y to be infinite-dimensional, and is one of the reasons why we do not consider that case here.

We shall now give some examples of computing the derivative. The strategy is always the same–expand out $f(x+h)-f(x)$, and (provided f is differentiable at x) we are left with the terms linear in h, which give the derivative, and terms of higher order which are collected together to form the remainder term.

Example 2.4.4.

1. Let $A = (a_{ij})$ be a $m \times n$ matrix and let $f : \mathbb{R}^n \to \mathbb{R}^m$ be the linear map $f : x \mapsto Ax$. Now

$$\begin{aligned} f(x+h) - f(x) &= A(x+h) - Ax \\ &= Ax + Ah - Ax \\ &= Ah \end{aligned}$$

So in this case $f(x+h) - f(x)$ is exactly given by the linear term Ah and the remainder term $R_f(h)$ is zero. So f is differentiable and the linear map $L = df(x)$ is given by $df(x) : h \mapsto Ah$.

2. Let $C = (c_{ij}) \in M_{n\times n}(\mathbb{R})$ be symmetric and let $f \colon \mathbb{R}^n \to \mathbb{R}$ be the quadratic form corresponding to C, that is $f(x) = x^T Cx = \langle x, Cx\rangle$. Letting $h \in \mathbb{R}^n$, we see:

$$\begin{aligned} f(x+h) - f(x) &= \langle x+h, C(x+h)\rangle - \langle x, Cx\rangle \\ &= \langle x, Cx\rangle + \langle h, Cx\rangle + \langle x, Ch\rangle + \langle h, Ch\rangle - \langle x, Cx\rangle \\ &= 2\,\langle Cx, h\rangle + \langle h, Ch\rangle \end{aligned}$$

where we use the fact that $\langle x, Ch\rangle$ is a scalar so

$$\langle x, Ch\rangle = x^T Ch = (x^T Ch)^T = h^T C^T x = h^T Cx = \langle h, Cx\rangle$$

as C is symmetric. Hence a candidate for $df(x)$ is $(2Cx)^T$, as $2(Cx)^T h = 2\,\langle Cx, h\rangle$. Indeed,

$$\begin{aligned} \left| \frac{f(x+h) - f(x) - 2(Cx)^T h}{|h|} \right| &= \left| \frac{\langle h, Ch\rangle}{|h|} \right| \le \frac{|h||Ch|}{|h|} \\ &\le \|C\|\,|h| \to 0 \text{ for } h \to 0\,, \end{aligned}$$

where $\|C\| = \left(\sum_{i,j=1}^n c_{ij}^2\right)^{\frac{1}{2}}$. Thus f is differentiable at every $x \in \mathbb{R}^n$ and $df(x) = (2Cx)^T$, that is $df(x)h = (2Cx)^T h = \langle 2Cx, h\rangle$.

3. $f\colon M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R}), f(A) = A^2$.
Let $H \in M_{n\times n}(\mathbb{R})$. Then

$$\begin{aligned} f(A+H) - f(A) &= (A+H)(A+H) - A^2 \\ &= AH + HA + H^2 \end{aligned}$$

The linear term $AH + HA$ is a candidate for the derivative:

$$\frac{f(A+H) - f(A) - (AH + HA)}{|H|} = \frac{H^2}{|H|} \to 0 \qquad \text{as } H \to 0.$$

Hence f is differentiable in every $A \in M_{n\times n}(\mathbb{R})$ with $df(A)H = AH + HA$.

There is a nice general formula for the differential $df(x)$ in terms of the Jacobian matrix $Df(x)$.

Proposition 2.4.5. *If $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \Omega$, then f is continuous, the partial derivatives $\partial_1 f(x), \ldots, \partial_n f(x)$ exist and*

$$df(x)h = Df(x)h$$

for all $h \in \mathbb{R}^n$. That is, with $h = \sum_{i=1}^n h_i e_i$ we have

$$\begin{pmatrix} df_1(x) \\ \vdots \\ df_m(x) \end{pmatrix} = \begin{pmatrix} \partial_1 f_1(x) & \ldots & \partial_n f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \ldots & \partial_n f_m(x) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}.$$

In other words, the Jacobian matrix $Df(x)$ is the representation of $df(x)$ with respect to the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$.

Proof. It suffices to prove the statement for $m = 1$. Continuity of f at x follows from

$$\lim_{h\to 0} \big(f(x+h) - f(x)\big) = \lim_{h\to 0} \big(Lh - R_f(h)\big) = 0.$$

To show that the partial derivatives exist, choose $h = te_i$. Then differentiability of f at x implies

$$\left|\frac{1}{t}\big(f(x+te_i) - f(x)\big) - Le_i\right| \to 0 \qquad \text{as } t \to 0.$$

Hence $\partial_i f(x) = Le_i$. Since $h = \sum_{i=1}^n h_i e_i$ we find $Lh = \sum_{i=1}^n h_i Le_i = \sum_{i=1}^n h_i \partial_i f(x)$. $\square$

Remark 2.4.6. a) Proposition **??** in particular implies that if $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at $x \in \Omega$, then

$$df(x)h = \sum_{i=1}^n \partial_i f(x) h_i = \langle \nabla f(x), h \rangle.$$

This can also be seen as the definition of the gradient. The gradient of f is the vector field ∇f such that $Lh = \langle \nabla f, h \rangle$ for all $h \in \mathbb{R}^n$.

b) It is important to realise that Proposition **??** just says that *if* f is differentiable, *then* the derivative is given by the Jacobian matrix of partial derivatives. As we have seen before in Example **??**, existence of the Jacobian matrix is *not* sufficient to guarantee differentiability.

2.5 A sufficient condition for differentiability

It is often not so easy to use the definition of differentiability to decide whether a function is differentiable or not. The following result gives a useful criterion.

Proposition 2.5.1. (Continuous partial derivatives imply differentiability) *Suppose that the function $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ has continuous partial derivatives. Then f is differentiable in Ω.*

Proof. It suffices to consider the case $m = 1$. Let $h \in \mathbb{R}^n$; a candidate for the derivative is $Lh = \sum_{k=1}^n \partial_k f(x) h_k$.
Let $x_0 = x, x_k = x + \sum_{j=1}^k h_j e_j$, such that $x_n = x + h$. Then $f(x_k) - f(x_{k-1}) = f(x_{k-1} + h_k e_k) - f(x_{k-1})$ and the Mean Value Theorem (for functions of one variable) implies $f(x_k) - f(x_{k-1}) = \partial_k f(x_{k-1} + \theta_k h_k e_k) h_k$ with $\theta_k \in [0, 1]$. Hence

$$
\begin{aligned}
\frac{|f(x+h) - f(x) - Lh|}{|h|} &= \frac{1}{|h|}\left|\sum_{k=1}^n (f(x_k) - f(x_{k-1})) - \sum_{k=1}^n \partial_k f(x) h_k\right| \\
&= \frac{1}{|h|}\left|\sum_{k=1}^n \partial_k f(x_{k-1} + \theta_k h_k e_k) h_k - \sum_{k=1}^n \partial_k f(x) h_k\right| \\
&\le \frac{|h|}{|h|}\left(\sum_{k=1}^n \Big(\partial_k f(x_{k-1} + \theta_k h_k e_k) - \sum_{k=1}^n \partial_k f(x)\Big)^2\right)^{\frac{1}{2}} \\
&\to 0 \quad \text{for } h \to 0,
\end{aligned}
$$

since $\partial_k f$ is continuous in x. □

Corollary 2.5.2. *If f has continuous partial derivatives, then f is continuous.*

2.6 The Chain Rule

Proposition 2.6.1. (Chain Rule) *Let $\Omega \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ be open and connected sets, let $g\colon \Omega \to V$ and $f\colon V \to \mathbb{R}^k$. Suppose that g is differentiable at $x \in \Omega$ and f is differentiable at $y = g(x) \in V$. Then the map $f \circ g\colon \Omega \to \mathbb{R}^k$ is differentiable at x and*

$$d(f \circ g)(x) = df(g(x))dg(x)\,.$$

In coordinates this reads

$$\frac{\partial}{\partial x_j}(f_i \circ g)(x) = \sum_{l=1}^m \frac{\partial}{\partial y_l} f_i(g(x)) \frac{\partial}{\partial x_j} g_l(x)\,, \qquad i = 1, \ldots, k\,, \quad j = 1, \ldots, n\,.$$

Proof. We define $A = dg(x)$ and $B = df(g(x))$. We need to show that $d(f \circ g)(x) = BA$. Since g and h are differentiable, we have

$$g(x+h) = g(x) + Ah + R_g(h) \quad \text{with } h \in \mathbb{R}^n, R_g(h) = o(|h|)$$

and

$$f(y+\eta) = f(y) + B\eta + R_f(\eta) \qquad \text{with } \eta \in \mathbb{R}^m, R_f(\eta) = o(|\eta|).$$

We choose now $\eta = g(x+h) - g(x) = Ah + R_g(h)$. Then

$$f(g(x+h)) = f(g(x)+\eta) = f(g(x)) + B\eta + R_f(\eta),$$

so

$$f(g(x+h)) = f(g(x)) + B(Ah + R_g(h)) + R_f(Ah + R_g(h))\,.$$

It remains to show $g(h) = BR_g(h) + R_f(Ah + R_g(h)) = o(|h|)$ To that aim notice that

$$\frac{|BR_g(h)|}{|h|} \leq \frac{|B||R_g(h)|}{|h|} \to 0 \text{ for } |h| \to 0$$

and, for sufficiently small $|h|$,

$$|Ah + R_g(h)| \leq |A||h| + |R_g(h)| \leq (|A|+1)|h|.$$

Hence

$$\frac{R_f(Ah + R_g(h))}{|h|} \to 0 \text{ for } |h| \to 0$$

□

Corollary 2.6.2. (Derivative of the Inverse) *Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is invertible with inverse $g\colon \mathbb{R}^n \to \mathbb{R}^n$. Suppose further that f is differentiable at x and that g is differentiable at $g = f(x)$.*
Then the Chain Rule implies for $g(f(x)) = x$ that

$$dg(f(x))df(x) = Id \qquad \text{and hence} \qquad dg(f(x)) = (df(x))^{-1}\,.$$

Example 2.6.3. (Polar coordinates in $\mathbb{R}^2$) Let $f\colon \mathbb{R}_+ \times (0, 2\pi) \to \mathbb{R}^2$ be given by $f(r,\varphi) = (r\cos\varphi, r\sin\varphi) =: (x,y)$. Let g be the inverse function to f. From

$$Df(r,\varphi) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$$

we deduce that $\det Df(r,\varphi) = r > 0$. Then

$$Dg(x,y) = Df(r,\varphi)^{-1} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix}$$

Corollary 2.6.4. (The gradient is perpendicular to level sets) *Let $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable and let γ be a regular curve parametrized by $\gamma\colon (\alpha, \beta) \to \Omega$, which lies in a level set of f, that is, $f(\gamma(t)) = c$ for all $t \in (\alpha, \beta)$. Then we have for all $t \in (\alpha, \beta)$ that*

$$0 = df(\gamma(t))\gamma'(t) = \big\langle \nabla f(\gamma(t)), \gamma'(t) \big\rangle$$

Remark 2.6.5. The direction of $\nabla f(x)$ is the direction of steepest ascent at x, $(-\nabla f(x))$ is the direction of steepest descent. Indeed, consider any $v \in \mathbb{R}^n$ with $|v| = 1$. Then

$$|df(x)v| = |\langle \nabla f(x), v \rangle| \leq |\nabla f(x)||v| = |\nabla f(x)|1 = |\nabla f(x)|$$

and equality holds if $v = \frac{\nabla f(x)}{|\nabla f(x)|}$.

2.7 Mean Value Theorems

Our goal in this section is to use information about the derivative of a function to obtain information about the function itself.

Remark 2.7.1. In the case $n = 1$ we know the following Mean Value Theorem for a differentiable function f: $f(x) - f(y) = f'(\xi)(x - y)$ for some $\xi \in (x, y)$. We cannot generalize this, however, for vector-valued functions, since in general we get a different ξ for every component. The Fundamental Theorem of Calculus does not have this disadvantage: $f(y) - f(x) = \int_x^y f'(\xi)\,\mathrm{d}\xi$ is also true for vector-valued functions, but of course requires f' to be continuous.

We are now going to prove some versions of the Mean Value Theorem for functions of several variables.

Proposition 2.7.2. *Suppose that $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable and let $x, y \in \Omega$ be such that the line segment $[x; y] = \{tx + (1-t)y \,|\, t \in [0, 1]\}$ is also contained in Ω. Then there exists $\xi \in [x; y]$, such that*

$$f(x) - f(y) = df(\xi)(x - y) = \langle \nabla f(\xi), x - y \rangle$$

Proof. Let $\gamma(t) = tx + (1-t)y, t \in [0, 1]$, and $F(t) = f(\gamma(t))$. Then $f(x) = F(1)$ and $f(y) = F(0)$. The Chain Rule implies that f is differentiable and

$$\frac{\mathrm{d}}{\mathrm{d}t} F(t) = df(\gamma(t))\gamma'(t)\,.$$

By the Mean Value Theorem for $n = 1$ there exists $\tau \in (0, 1)$, such that $F(1) - F(0) = F'(\tau)$. Hence

$$f(x) - f(y) = df(\gamma(\tau))(x - y) = df(\xi)(x - y) \qquad \text{with } \xi = \gamma(\tau)\,.$$

□

Corollary 2.7.3. *Let $\Omega \subseteq \mathbb{R}^n$ be path connected. If $f\colon \Omega \to \mathbb{R}$ satisfies $df(x) = 0$ for all $x \in \Omega$, then f is constant in Ω.*

Proof. Connect two arbitrary points by a polygon and apply the Mean Value Theorem to each part. □

Proposition 2.7.4. *Let $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m, f \in C^1(\Omega, \mathbb{R}^m)$. Suppose that for $x, y \in \Omega$ there exists a regular curve $\gamma : [\alpha, \beta] \to \Omega$ which connects x and y, i.e $\gamma(\alpha) = y, \gamma(\beta) = x$. Then*

$$f(x) - f(y) = \int_\alpha^\beta df(\gamma(t))\gamma'(t)\,\mathrm{d}t\,.$$

Proof. Use the Fundamental Theorem of Calculus and the Chain Rule for for $F(t) = f(\gamma(t))$. □

Remark 2.7.5. Another version: let $x \in \Omega, \xi \in \mathbb{R}^n$ und $\forall t \in [0, 1] : x + t\xi \in \Omega$. Then

$$f(x + \xi) - f(x) = \int_0^1 df(x + t\xi)\xi\,\mathrm{d}t\,.$$

Proposition 2.7.6. *Let $\Omega \subseteq \mathbb{R}^n$ be open and convex, i.e. for all points $x, y \in \Omega$ we also have that the line segment $[x; y] \subset \Omega$. Suppose that $f \in C^1(\Omega, \mathbb{R}^m)$ and*

$$\sup_{x\in\Omega} |Df(x)| \leq K.$$

Then we have for all $x, y \in \Omega$ that

$$|f(x) - f(y)| \leq K|x - y|\,,$$

that is, f is Lipschitz continuous in Ω with Lipschitz constant K.

Proof. Exercise. □

3 The Inverse Function Theorem and the Implicit Function Theorem

The Inverse Function Theorem and the Implicit Function Theorem are two of the most important theorems in Analysis. The Inverse Function Theorem tells us when we can locally invert a function; the Implicit Function Theorem tells us when a function is given implicitly as a function of other variables. We will discuss both theorems in $\mathbb{R}^n$ here, but they are also valid in basically the same form in infinite-dimensional spaces (more precisely, in Banach spaces). The range of their applications is vast and we can only get a small glimpse of their significance in this course.

The flavour of both results is similar: we linearise the problem at a point by taking the derivative df. Now, subject to a suitable nondegeneracy condition on df, we obtain a result that works on a neighbourhood of the point. In this way we go from an infinitesimal statement to a local (but not a global) result.

The theorems are equivalent; the classical approach, however, is to prove first the Inverse Function Theorem via the Contraction Mapping Fixed Point Principle and then deduce the Implicit Function Theorem from it. The proof of the Inverse Function Theorem is however lengthy and technical and we do not have the time to go through it in this lecture course. We recommend the books by Spivak (*Calculus on Manifolds*, W.A. Benjamin) [?] and Krantz and Parks (*The Implicit Function Theorem, History, Theory and Applications*, Birkhäuser), where you can also find an elementary (but still not short) proof of the Implicit Function Theorem which does not use the Inverse Function Theorem. The latter then follows directly as a corollary from the Implicit Function Theorem.

In these lecture notes we first prove the Implicit Function Theorem in the simplest setting, which is the case of two variables. We then state carefully the Implicit Function Theorem and the Inverse Function Theorem in higher dimensions, deduce the Implicit Function Theorem from the Inverse Function Theorem, and give some examples of applications.

3.1 The Implicit Function Theorem in $\mathbb{R}^2$

We start with a simple example. Consider $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$, which is just the unit circle in the plane. Can we find a function $y = y(x)$ such that $x^2 + y(x)^2 = 1$? Obviously, in this example, we cannot find one function to describe the whole unit circle in this way. However, we can do it *locally*, that is in a neighbourhood of a point $(x_0, y_0) \in S^1$, as long as $y_0 \neq 0$. In this example we can find y explicitly: it is $y(x) = \sqrt{1 - x^2}$ if

$y_0 > 0$ and $y(x) = -\sqrt{1-x^2}$ if $y_0 < 0$ both for $|x| < 1$. Notice also, that if $y_0 = 0$, we cannot find such a function y, but we can instead write x as a function of y.

The Implicit Function Theorem describes conditions under which certain variables can be written as functions of the others. In $\mathbb{R}^2$ it can be stated as follows.

Theorem 3.1.1. (Implicit Function Theorem in $\mathbb{R}^2$) *Let $\Omega \subseteq \mathbb{R}^2$ be open and $F \in C^1(\Omega)$. Let $(x_0, y_0) \in \Omega$ and assume that*

$$f(x_0, y_0) = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(x_0, y_0) \neq 0\,.$$

Then there exist open intervals $I, J \subseteq \mathbb{R}$ with $x_0 \in I$, $y_0 \in J$ and a unique function $g\colon I \to J$ such that $y_0 = g(x_0)$ and

$$f(x, y) = 0 \qquad \text{if and only if} \qquad y = g(x) \quad \text{for all } (x, y) \in I \times J.$$

Furthermore, $g \in C^1(I)$ with

$$g'(x_0) = -\frac{\frac{\partial f}{\partial x}(x_0, y_0)}{\frac{\partial f}{\partial y}(x_0, y_0)}\,. \tag{3.1}$$

Remark 3.1.2. Obviously, an analogous result is true if $\frac{\partial f}{\partial x}(x_0, y_0) \neq 0$.

Proof. (Not examinable) Without loss of generality we can assume that $\frac{\partial f}{\partial y}(x_0, y_0) > 0$. Due to the continuity of $\frac{\partial f}{\partial y}$ we can also assume – by making Ω smaller if necessary – that

$$\frac{\partial f}{\partial y}(x, y) \geq \delta > 0 \qquad \text{for all } (x, y) \in \Omega\,. \tag{3.2}$$

As a consequence we can find $y_1 < y_0 < y_2$ such that $f(x_0, y_1) < 0 < f(x_0, y_2)$ and due to the continuity of f we can find an open interval I containing x_0 such that

$$f(x, y_1) < 0 < f(x, y_2) \qquad \text{for all } x \in I\,. \tag{3.3}$$

The Intermediate Value Theorem and (**??**) imply that for each $x \in I$ there exists a unique $y \in (y_1, y_2) =: J$ such that $f(x, y) = 0$. Denote this y by $g(x)$. The continuity of f and the uniqueness of y also imply that g is continuous.

To complete the proof of the theorem, we need to show that g is continuously differentiable in I and that (**??**) holds. With the notation $y = g(x)$ we find

$$f(x+s, y+t) - f(x, y) = s\frac{\partial f}{\partial x}(x, y) + t\frac{\partial f}{\partial y}(x, y) + \varepsilon(s, t)\sqrt{s^2 + t^2}\,, \tag{3.4}$$

with $\varepsilon(s, t) \to 0$ as $(s, t) \to 0$. We now choose $t = g(x+s) - g(x)$ such that the left hand side in (**??**) vanishes and obtain

$$t\frac{\partial f}{\partial y}(x, y) = -s\frac{\partial f}{\partial x}(x, y) - \varepsilon(s, t)\sqrt{s^2 + t^2}\,. \tag{3.5}$$

We rearrange to obtain

$$\Big|\frac{t}{s}+\frac{\partial f}{\partial x}(x,y)/\frac{\partial f}{\partial y}(x,y)\Big| \le \frac{|\varepsilon|}{|\frac{\partial f}{\partial y}(x,y)|}\Big(1+\frac{|t|}{|s|}\Big)\,.$$

Thus, if we can show that $\frac{|t|}{|s|} \le C$ as $s \to 0$, then we can let $s \to 0$ in the above inequality to find that indeed $g'(x)$ exists for all $x \in I$. For $(x,y) = (x_0, y_0)$ we find the formula in (**??**) and the properties of f and g also imply that g' is continuous.

We still need to show that $\frac{|t|}{|s|} \le C$. We obtain from (**??**) that

$$\frac{|t|}{|s|} \le \Big|\frac{\partial f}{\partial x}(x,y)\Big|/\Big|\frac{\partial f}{\partial y}(x,y)\Big| + \frac{|\varepsilon|}{\big|\frac{\partial f}{\partial y}(x,y)\big|}\Big(1+\frac{|t|}{|s|}\Big) \le \Big|\frac{\partial f}{\partial x}(x,y)\Big|/\Big|\frac{\partial f}{\partial y}(x,y)\Big| + \frac{|\varepsilon|}{\delta}\Big(1+\frac{|t|}{|s|}\Big)\,.$$

We can choose now $|s|$ so small such that $\frac{|\varepsilon|}{\delta} \le \frac{1}{2}$ and then

$$\frac{|t|}{|s|} \le 2\Big|\frac{\partial f}{\partial x}(x,y)\Big|/\Big|\frac{\partial f}{\partial y}(x,y)\Big| + 1.$$

This finishes the proof of the theorem. □

Example 3.1.3. In the example at the beginning of this section we have $f(x,y) = x^2 + y^2 - 1$. The theorem tells us that we this relation defines y as a function of x in a neighbourhood of a point where $\frac{\partial f}{\partial y}$ is nonzero, that is, in a neighbourhood of points other than $(\pm 1, 0)$.

Example 3.1.4. We show that for sufficiently small $a > 0$ there exists a function $g \in C^1(-a,a)$ with $g(0) = 0$ such that

$$g^2(x)x + 2x^2 e^{g(x)} = g(x)\,.$$

Indeed, define $f\colon \mathbb{R}^2 \to \mathbb{R}$ via $f(x,y) = y^2x + 2x^2e^y - y$. Then $f(0,0) = 0$ and $\partial_y f(0,0) = -1$. Hence the Implicit Function Theorem implies the existence of the function g as claimed. Furthermore we can compute $g'(0) = -\partial_x f(0,0)/\partial_y f(0,0) = 0$. Of course, we cannot hope for an *explicit* expression for g, but the Implicit Function Theorem tells us quite easily that such a g exists.

3.2 The Implicit Function Theorem in $\mathbb{R}^n$

To formulate the Implicit Function Theorem in $\mathbb{R}^n$ we first introduce some convenient notation. We write

- $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^m \ni (x_1, \dots, x_k, y_1, \dots y_m) =: (x,y)$
- $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m, (x_0, y_0) \in \Omega \subseteq \mathbb{R}^n, f(x_0,y_0) =: z_0$

We are looking for open neighbourhoods U of x_0 and V of y_0 as well as a function $g\colon U \to V$ such that

$$\forall (x,y) \in U \times V : f(x,y) = z_0 = f(x_0,y_0) \Leftrightarrow y = g(x)$$

It is instructive to consider first the linear case. Let $f(x,y) = Ax + By$ with $A \in M_{m\times k}(\mathbb{R})$ and $B \in M_{m\times m}(\mathbb{R})$. If B is invertible then the equation $f(x,y) = Ax_0 + By_0 =: z_0$ can be solved for y via

$$y = B^{-1}\big(z_0 - Ax\big).$$

Notice that B is just the $m \times m$ submatrix of Df consisting of the partial derivatives with respect to the y-variables.

Now consider the nonlinear case. Let $f \in C^1(\Omega)$ and write

$$Df(x,y)) = (D_x f(x,y), D_y f(x,y))\,,$$

where

$$D_x f(x,y) = \left(\frac{\partial f_j}{\partial x_i}\right) \in M_{m\times k}(\mathbb{R}) \qquad (j = 1,\dots,m; i = 1,\dots,k)$$

and

$$D_y f(x,y) = \left(\frac{\partial f_j}{\partial y_i}\right) \in M_{m\times m}(\mathbb{R}) \qquad (j = 1,\dots,m; i = 1,\dots,m)\,.$$

Then

$$f(x,y) = f(x_0,y_0) + D_x f(x_0,y_0)(x-x_0) + D_y f(x_0,y_0)(y-y_0) + o(|(x,y)-(x_0,y_0)|).$$

If the remainder term were zero, then we would have $f(x,y) = f(x_0,y_0) = z_0$ iff

$$D_x f(x_0,y_0)(x-x_0) = -D_y f(x_0,y_0)(y-y_0)\,.$$

which, if $D_y f(x_0,y_0)$ is invertible, is equivalent to

$$y = y_0 - \Big(D_y f(x_0,y_0)\Big)^{-1} D_x f(x_0,y_0)(x-x_0).$$

Hence there exists a function $g(x)$ auch that $F(x,y) = z_0$ iff $y = g(x)$, as desired.

In the nonlinear case, of course the remainder term is nonzero. The Implicit Function Theorem is the statement that we can still conclude the existence of such a function g, subject to the nondegeneracy condition that $D_y f$ is invertible.

Theorem 3.2.1. (The Implicit Function Theorem) *Let $f\colon \Omega \subseteq \mathbb{R}^{k+m} \to \mathbb{R}^m$, where $n = k+m$, $f \in C^1(\Omega,\mathbb{R}^m)$ and let $(x_0,y_0) \in \Omega$ with $z_0 = f(x_0,y_0)$. If $D_y f(x_0,y_0)$ is invertible then there exist open neighbourhoods U of x_0 and V of y_0, and a function $g \in C^1(U,V)$ such that*

$$\{(x,y) \in U \times V \mid f(x,y) = z_0\} = \{(x,y) \mid x \in U, y = g(x)\}.$$

Furthermore

$$Dg(x_0) = -\Big(D_y f(x_0,y_0)\Big)^{-1} D_x f(x_0,y_0)\,.$$

So $y \in V \subset \mathbb{R}^m$ is defined implicitly as a function of $x \in U \subset \mathbb{R}^k$ via the relation $f(x, y) = 0$.

Example 3.2.2.

a) (Nonlinear system of equations)

Consider the system of equations

$$f(x, y_1, y_2) = \begin{pmatrix} x^3 + y_1^3 + y_2^3 - 7 \\ xy_1 + y_1y_2 + y_2x + 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The function f is zero at the point $(2, -1, 0)$ and

$$Df(x, y_1, y_2) = \begin{pmatrix} 3x^2 & 3y_1^2 & 3y_2^2 \\ y_1 + y_2 & x + y_2 & x + y_1 \end{pmatrix},$$

hence

$$D_y f(2, -1, 0) = \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix} \qquad \text{with} \qquad \det D_y f(2, -1, 0) = 3 \neq 0.$$

The Implicit Function Theorem implies that there exist open neighbourhoods I of 2 and $V \subseteq \mathbb{R}^2$ of $(-1, 0)$ and a continuously differentiable function $g\colon I \to V$, with $g(2) = (-1, 0)$, such that

$$f(x, y_1, y_2) = 0 \quad \Leftrightarrow \quad y = (y_1, y_2) = g(x) = (g_1(x), g_2(x))$$

for all $x \in I, y \in V$. Furthermore, the derivative of g at $x_0 = 2$ is given by

$$Dg(2) = -\begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 12 \\ -1 \end{pmatrix} = -\frac{1}{3}\begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix}\begin{pmatrix} 12 \\ -1 \end{pmatrix} = \begin{pmatrix} -4 \\ 9 \end{pmatrix}.$$

b) The function $f\colon \mathbb{R}^4 \to \mathbb{R}^2$ is given by

$$f(x, y, u, v) = \begin{pmatrix} x^2 + uy + e^v \\ 2x + u^2 - uv \end{pmatrix}.$$

Consider the point $(2, 5, -1, 0)$ such that $f(2, 5, -1, 0) = (0, 5)^T$. The Jacobian matrix of f is

$$Df(x, y, u, v) = \begin{pmatrix} 2x & u & y & e^v \\ 2 & 0 & 2u - v & -u \end{pmatrix}.$$

Hence

$$D_{(u,v)} f(x, y, u, v) = \begin{pmatrix} y & e^v \\ 2u - v & -u \end{pmatrix} \quad \text{and} \quad D_{(u,v)} f(2, 5, -1, 0) = \begin{pmatrix} 5 & 1 \\ -2 & 1 \end{pmatrix}.$$

Since $\det Df(2, 5, -1, 0) = 7 \neq 0$, the Implicit Function Theorem implies that there exist open neighbourhoods $U \subset \mathbb{R}^2$ of $(2, 5)$ and $V \subset \mathbb{R}^2$ of $(-1, 0)$ and a function $g \in C^1(U,V)$ with $g(2,5) = (-1,0)$ and $f(x,y,g(x,y)) = (0,5)^T$ for all $(x,y) \in U$. We can also compute that

$$Dg(2,5) = -\begin{pmatrix} 5 & 1 \\ -2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 4 & -1 \\ 2 & 0 \end{pmatrix} = -\frac{1}{7}\begin{pmatrix} 2 & -1 \\ 18 & -2 \end{pmatrix}.$$

c) (Writing a surface locally as a graph)

Let $h\colon \mathbb{R}^3 \to \mathbb{R}$ with $h(x,y,z) = xy - z\log y + e^{xz} - 1$. Can we represent the 'surface' given by $h(x,y,z) = 0$ locally in a neighbourhood of $(0,1,1)$ either in the form $x = f(y,z)$, $y = g(x,z)$ or $z = p(x,y)$? The Jacobian matrix of h is $Dh(x,y,z) = (y + ze^{xz}, x - \frac{z}{y}, -\log y + xe^{xz})$ and thus $Dh(0,1,1) = (2,-1,0)$. Hence, the Implicit Function Theorem tells us that we can represent the surface locally as $x = f(y,z)$ or $y = g(x,z)$, but it does not tell us whether we can do it in the form $z = p(x,y)$. In fact, one can show that the latter is not possible.

3.3 The Inverse Function Theorem

In this section we consider the following problem. Given a function $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$, does there exist locally around a point x_0 an inverse function $g = f^{-1}$?

The idea is, as usual, to linearise around a point. Let $x_0 \in \mathbb{R}^n, y_0 = f(x_0)$ and assume that the Jacobian matrix $Df(x_0)$ is invertible. Then we find for general x that

$$f(x) = y_0 + Df(x_0)(x - x_0) + o(|x - x_0|).$$

Now, if the remainder term were not present, then we could just invert the function by

$$y = f(x) \quad \text{if} \quad x = x_0 + Df(x_0)^{-1}(y - y_0)$$

Of course, this will only be true if f is itself linear! The content of the Inverse Function Theorem is that for general differentiable f there will still be a *local* inverse, that is, an inverse defined on a neighbourhood of x_0, provided the Jacobian $Df(x_0)$ is invertible.

Example 3.3.1. Let $f\colon \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$.

- For $x_0 > 0$ or $x_0 < 0$ we have that f is invertible in a neighbourhood of x_0
- For $x_0 = 0$ there is no neighbourhood of x_0 where f has an inverse. Indeed, $f'(0) = 0$ is not invertible

We first make a definition that is relevant for our discussion of the local behaviour of f.

Definition 3.3.2. (Diffeomorphism) Let $f\colon U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^n$ and suppose that U and V are open in $\mathbb{R}^n$. We say that f is a diffeomorphism if f is bijective, that is there exists $f^{-1}\colon V \to U$, and if $f \in C^1(U,V)$ and $f^{-1} \in C^1(V,U)$.

Example 3.3.3. Here is a simple example for a function $f\colon (-1,1) \to (-1,1)$ which is bijective, $f \in C^1$, but f^{-1} is not differentiable on $(-1,1)$.

Let $f\colon (-1,1) \to (-1,1)$ be given by $f(x) = x^3$. Obviously f is bijective with inverse $f^{-1}\colon (-1,1) \to (-1,1)$ given by $f^{-1}(y) = y^{\frac{1}{3}}$. Furthermore, $f \in C^\infty(-1,1)$, but f^{-1} is not differentiable in any neighbourhood of 0. Hence, f is not a diffeomorphism.

We sometimes informally think of a diffeomorphism as a 'smooth change of coordinates'.

We can now state our theorem.

Theorem 3.3.4. (The Inverse Function Theorem in $\mathbb{R}^n$) *Let $\Omega \subseteq \mathbb{R}^n$ be open, let $f \in C^1(\Omega, \mathbb{R}^n)$ and let $x_0 \in \Omega$. If $Df(x_0)$ is invertible, then there exists an open neighbourhood U of x_0 such that $f(U)$ is open and $f\colon U \to f(U)$ is a diffeomorphism. Furthermore*

$$Df^{-1}(f(x_0)) = \big(Df(x_0)\big)^{-1}.$$

Remark 3.3.5. Notice that it follows that $f^{-1} \in C^1$ in a neighbourhood of $f(x_0)$, we do not need to assume it.

This is an example of our philosophy that the linearisation of a function at a point will give us *local* qualitative information about the function's behaviour on a neighbourhood of the point, provided we have suitable nondegeneracy conditions (in this case the assumption that the derivative is invertible at the point). It is important to realise, however, that we do not necessarily get a *global* result. Even if the derivative is everywhere invertible the function need not be a global diffeomorphism from Ω to $f(\Omega)$, or even a global bijection. We shall see examples of this in (c) and (d) below.

Example 3.3.6.

a) $f\colon \mathbb{R} \to \mathbb{R}, f(x) = x^2$. If $x_0 > 0$ we can choose $U = (0,\infty)$, if $x_0 < 0$ we can choose $U = (-\infty, 0)$.

b) Let $f\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}^2$, be given by

$$(r,\varphi) \mapsto (r\cos\varphi, r\sin\varphi)$$

$$Df(r,\varphi) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} \qquad \text{such that} \qquad \det Df(r,\varphi) = r > 0.$$

We have already seen that

$$(Df(r,\varphi))^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\varphi & r\sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}.$$

Hence f is locally invertible everywhere, but not globally (in fact f is 2π-periodic in the φ variable). The local inverse can be computed: Let $f(r,\varphi) =: (x,y) \in \mathbb{R}^2$ and let

$$U = \left\{(r,\varphi) \mid \varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)\right\} \quad \text{and} \quad V = \{(x,y) \in \mathbb{R}^2 \mid x > 0\}.$$

$f\colon U \to V$ is a diffeomorphism, where $g = f^{-1}\colon V \to U$ is given by

$$g(x,y) = \left(\sqrt{x^2+y^2}, \arctan\frac{y}{x}\right).$$

c) The following important example is one we encountered in Example **??**. The exponential function $\exp\colon \mathbb{C} \to \mathbb{C}$ given by $z \mapsto e^z$ is, in real cordinates, the map

$$(x,y) \mapsto (e^x \cos y, e^x \sin y)$$

The Jacobian is

$$Df(x,y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}$$

and $\det Df(x,y) = e^{2x}(\cos^2 y + \sin^2 y) = e^{2x}$ which never vanishes. Hence Df is always invertible, and the Inverse Function Theorem tells us that the exponential map is a *local* diffeomorphism. However it is not a *global* diffeomorphism as it is not bijective. The map is periodic in y with period 2π–equivalently $\exp(z+2\pi i) = \exp(z)$. (For those of you who have seen the concept in topology, the exponential map is a covering map from $\mathbb{C}$ onto $\mathbb{C} - \{0\}$).

d) Let $f\colon (0,\infty) \times \mathbb{R} \to \mathbb{R}^2$ be given by $f(x,y) = (\cosh x \cos y, \sinh x \sin y) =: (u,v)$. Then

$$Df(x,y) = \begin{pmatrix} \sinh x \cos y & -\cosh x \sin y \\ \cosh x \sin y & \sinh x \cos y \end{pmatrix}.$$

Hence $\det Df(x,y) = \sinh^2 x + \sin^2 y$ and thus $\det Df(x,y) > 0$ for all $x > 0, y \in \mathbb{R}$. As a consequence of the Inverse Function Theorem we have that f is locally a diffeomorphism for all (x,y). (The function f is not a global diffeomorphism as it is periodic in y.)

Notice that for fixed $x > 0$ the image $f(x,y)$ describes an ellipse with axes of length $\cosh x > 1$ and $\sinh x$ respectively. Hence $f((0,\infty)\times\mathbb{R}) = \mathbb{R}^2 \setminus \{(u,0) \,|\, |u| \le 1\}$.

We conclude by deducing the Implicit Function Theorem from the Inverse Function Theorem.

Recall the setup of the latter theorem. We have a C^1 map $f\colon \mathbb{R}^{k+m} \to \mathbb{R}^m$. We write a point of $\mathbb{R}^{k+m}$ as (x,y) with $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^m$, and we assume the $m \times m$ submatrix $D_y f$ of Df at (x_0,y_0) is invertible. We want to locally find a function g such that $f(x,y) = 0$ iff $y = g(x)$.

In order to apply the Inverse Function Theorem we will expand f to a function $\mathbb{R}^{k+m} \to \mathbb{R}^{k+m}$: explicitly, we let

$$F(x,y) = (x, f(x,y)).$$

Now

$$DF = \begin{pmatrix} I & 0 \\ D_x f & D_y f \end{pmatrix}$$

so as we are assuming invertibility of $D_y f$, we have that DF is invertible at (x_0, y_0). The Inverse Function Theorem now tells us F has a local differentiable inverse $h\colon (x,y) \mapsto (h_1(x,y), h_2(x,y))$. We have

$$(x,y) = F \circ h(x,y) = (h_1(x,y), f \circ h(x,y))$$

so $h_1(x,y) = x$, and hence $h(x,y) = (x, h_2(x,y))$ with $f(x, h_2(x,y)) = f \circ h(x,y) = y$. In particular, $f(x, h_2(x,0)) = 0$, and we can take $g(x) = h_2(x,0)$.

In fact, this way of approaching the Implicit Function Theorem also yields the following useful theorem.

Theorem 3.3.7. *Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ (where $m \le n$) be a C^1 function such that $f(a) = 0$ and rank $Df(a) = m$. Then there is an open neighbourhood U of a and a differentiable function $h : U \to \mathbb{R}^n$ with differentiable inverse such that*

$$f \circ h(x_1, \ldots, x_n) = (x_{n-m+1}, \ldots, x_n)$$

Proof. This is basically contained in the proof of the Implicit Function Theorem above. After applying a permutation of coordinates (which is a diffeomorphism of $\mathbb{R}^n$) we can assume that the $m \times m$ matrix formed from the m last columns of $Df(a)$ is invertible. Now the proof we saw above shows the existence of h such that $f \circ h(x,y) = y$, as required. $\square$

The significance of this is that if $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and Df has rank m at a point (ie is of maximal rank at a), we can *locally* apply a diffeomorphism which makes f into the simplest possible rank m map, that is, a projection to $\mathbb{R}^m$. In particular, the local structure of level sets of f around points of maximum rank is very simple, up to a diffeomorphism ('change of coordinates').

4 Submanifolds in $\mathbb{R}^n$ and constrained minimisation problems

We are now going to introduce the notion of submanifolds of $\mathbb{R}^n$, which are generalisations to general dimensions of smooth surfaces in $\mathbb{R}^3$.

4.1 Submanifolds in $\mathbb{R}^n$

Let us begin by looking at hypersurfaces in $\mathbb{R}^n$, that is, zero loci of functions $f\colon \mathbb{R}^n \to \mathbb{R}$. These can be very complicated and singular in general, but we expect that for a generic choice of f we get a smooth set. We can make this precise using the Implicit Function Theorem as follows.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a C^1-function and $M = \{x \in \mathbb{R}^n \mid f(x) = 0\} = f^{-1}\{0\}$ its zero set. If $Df(a) \neq 0$ for some $a \in M$, then we know from the Implicit Function Theorem, that we can represent M in a neighbourhood of a as a graph of a function of $n-1$ variables (after a suitable reordering of coordinates $x_n = h(x_1, \dots, x_{n-1})$ In this way, a neighbourhood of a is seen to be diffeomorphic to an open set in $\mathbb{R}^{n-1}$. We could also see this via the result at the end of the previous section, because this tells us that after a diffeomorphism we can reduce to the case when the map is a projection.

If this kind of behaviour holds for all points $a \in M$, we say M is an $n-1$-dimensional *submanifold* of $\mathbb{R}^n$. So we know that if $Df(x)$ is nonzero for all $x \in M$, then M is an $n-1$-dimensional submanifold of $\mathbb{R}^n$.

We are now going to generalize this definition to k-dimensional submanifolds of $\mathbb{R}^n$, for general k.

Definition 4.1.1. (Submanifolds of $\mathbb{R}^n$) Let $0 < k < n$ be an integer. A set $M \subseteq \mathbb{R}^n$ is called a *k-dimensional submanifold* of $\mathbb{R}^n$, if for every $x_0 \in M$ there exists an open neighbourhood Ω of x_0 in $\mathbb{R}^n$ and $f \in C^1(\Omega, \mathbb{R}^{n-k})$, such that

$$M \cap \Omega = f^{-1}\{0\} \quad \text{and} \quad \operatorname{rank} Df(x) = n-k \quad \text{for all } x \in \Omega\,.$$

Remark 4.1.2. It suffices to require $\operatorname{rank} Df(x) = n-k$ for all $x \in M \cap \Omega$. Indeed, if $\operatorname{rank} Df(x) = n-k$ for an $x \in M \cap \Omega$, this means that the matrix $Df(x)$ has $n-k$ independent columns, or in other words, that the determinant of the matrix formed by these $n-k$ independent columns is nonzero. Since Df is continuous and the determinant is a continuous function, it follows that the determinant of this submatrix is also nonzero in an open neighbourhood of x.

The next proposition, which is a consequence of the Implicit Function Theorem, will tell us that a submanifold can be locally represented as a graph of a differentiable function.

Proposition 4.1.3. (Submanifolds can be locally represented as graphs) *For a set $M \subseteq \mathbb{R}^n$ the following properties are equivalent.*

(1) M is a k-dimensional submanifold of $\mathbb{R}^n$ (in the sense of definition **??***).*

(2) For each $x \in M$ we can, after suitably relabelling the coordinates, write $x = (z_0, y_0)$ with $z_0 \in \mathbb{R}^k$, $y_0 \in \mathbb{R}^{n-k}$ and find an open neighbourhood U of z_0 in $\mathbb{R}^k$, an open neighbourhood V of y_0 in $\mathbb{R}^{n-k}$, and a map $g \in C^1(U, V)$ with $g(z_0) = y_0$, such that

$$M \cap (U \times V) = \{(z, g(z)) \mid z \in U\}.$$

Remark 4.1.4. In (2) it is important that we remember that the statement is true only after relabelling the coordinates. For instance, consider again the unit circle $S^1 = \big\{(x_1, x_2) \,|\, x_1^2 + x_2^2 = 1\big\}$ in $\mathbb{R}^2$. If $x = (1, 0) \in S^1$ we have

$$S^1 \cap \big((0, 2) \times (-1, 1)\big) = \big\{(\sqrt{1 - z^2}, z) \,|\, |z| < 1\big\}\,.$$

Hence, to get the statement in (2) we have to relabel (x_1, x_2) as (x_2, x_1).

Proof. We first show that (1) implies (2): After possibly relabelling the coordinates we can write x as $x = (z_0, y_0)$ such that $D_y f(x)$ is invertible. Then property (2) follows from the Implicit Function Theorem.

Now assume that (2) is satisfied. Define $\Omega = U \times V$ and $f \in C^1(\Omega, \mathbb{R}^{n-k})$ via

$$f(z, y) = y - g(z)$$

Then $M \cap \Omega = f^{-1}\{0\}$ and $Df(z, y) = (-Dg(z), \mathrm{Id}_{n-k})$. It follows that $\mathrm{rank} Df(z, y) = n - k$. □

As in the hypersurface case, we see that a suitable neighbourhood of each point in a k-dimensional submanifold can be identified by a diffeomorphism with an open set in $\mathbb{R}^k$. We can think of submanifolds as subsets of $\mathbb{R}^n$ that locally look like open sets in Euclidean space $\mathbb{R}^k$.

Remark 4.1.5. In fact it is possible to use these ideas to define abstract k-dimensional manifolds without reference to an embedding in $\mathbb{R}^n$. Roughly speaking, such a manifold is a space covered by open sets ('charts'), each homeomorphic to an open set in $\mathbb{R}^k$, such that the charts fit together smoothly in a suitable sense. It is in fact possible to transfer the machinery of differential and integral calculus to such abstract manifolds. These ideas are explored further in the Part C course on Differentiable Manifolds.

Example 4.1.6.

a) (Curves in $\mathbb{R}^2$)

i) The unit circle in $\mathbb{R}^2$:

  * A definition as the level set of a function is given by $\{(x,y) \in \mathbb{R}^2 \mid x^2+y^2-1=0\}$. Note that this is $f^{-1}(0)$ where $f(x,y) = x^2+y^2-1$, and that $Df(x,y) = (2x, 2y)$ which has rank 1 at all points of the unit circle (in fact, at all points except the origin). So the circle is a 1-dimensional submanifold of $\mathbb{R}^2$.
  * A local representation as a graph of a function is for example $y(x) = \pm\sqrt{1-x^2}\,;\ x \in [-1,1]$.
  * A parametrisation is given by $\gamma\colon [0,2\pi) \to \mathbb{R}^2; \gamma(t) = (\cos t, \sin t)$.

ii)
$$M = \{(x,y) \in \mathbb{R}^2 \,|\, x^3 - y^2 - 1 = 0\}$$
defines a one-dimensional submanifold (a regular curve) in $\mathbb{R}^2$. Here, $Df(x,y) = (3x^2, -2y)$ which again has rank 1 on all points of M.

iii) Consider now
$$M = \{(x,y) \in \mathbb{R}^2 \mid x^2 - y^2 - 0\}$$
Now $M = f^{-1}(0)$ where $f(x,y) = x^2 - y^2$, and $Df(x,y) = (2x, -2y)$ has rank 1 except at the origin, which *is* on the curve. So the curve has the submanifold property away from the origin, but not at the origin itself. Geometrically, we can see that the curve is the union of the two lines $y = x$ and $y = -x$, which meet at $(0,0)$. So away from the origin the curve looks like a 1-dimensional submanifold, but this breaks down at the origin.

b) (Ellipsoids)

An ellipsoid is given by
$$M = \left\{ x \in \mathbb{R}^3 \mid f(x) = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} - 1 = 0 \right\}.$$
for some $a, b, c > 0$. We check that this defines a two-dimensional submanifold of $\mathbb{R}^3$. Indeed,
$$Df(x) = 2\left(\frac{x_1}{a^2}, \frac{x_2}{b^2}, \frac{x_3}{c^2}\right)$$
and thus $Df(x) = 0$ if and only if $x = 0$, but $x = 0 \notin M$.

$$f(x) = x_1^2 + x_2^2 - x_3^2 - c$$

c) (Torus)

Let $n = 3$ and $k = 2$. For $0 < r < R$ a torus is given by
$$T = \left\{(x,y,z) \in \mathbb{R}^3 \,|\, f(x,y,z) = \left(\sqrt{x^2+y^2} - R\right)^2 + z^2 - r^2 = 0\right\}.$$
That is, the torus consists of the points in $\mathbb{R}^3$ which have distance r to a circle with radius R. The defining function f is continuously differentiable away from

the z-axis. However, when $r < R$ the torus does not contain any point on the z-axis. We calculate

$$Df(x,y,z) = \big(2\big(\sqrt{x^2+y^2}-R\big)\frac{x}{\sqrt{x^2+y^2}}, 2\big(\sqrt{x^2+y^2}-R\big)\frac{y}{\sqrt{x^2+y^2}}, 2z\big)$$

and see that $Df(x,y,z) \neq 0$ when $(x,y,z) \in T$. Consequently, $\mathrm{rank} Df(x,y,z) = 1$ for $(x,y,z) \in T$ and we conclude that T is a two-dimensional submanifold of $\mathbb{R}^3$.

d) (Orthogonal group)

We claim that

$$O(n) = \Big\{X \in M_{n\times n}(\mathbb{R}) \mid X^T X = \mathrm{Id}\Big\}$$

is a submanifold of $\mathbb{R}^{n^2}$ of dimension $\frac{1}{2}n(n-1)$.

To see this let

$$S(n) = \Big\{X \in M_{n\times n}(\mathbb{R}) \mid X^T = X\Big\}$$

be the set of symmetric matrices. $S(n)$ is isomorphic to $\mathbb{R}^r$ with $r = n + (n-1) + (n-2) + \ldots + 1 = \frac{n(n+1)}{2}$.

Let $f\colon M_{n\times n}(\mathbb{R}) \to S(n)$ be defined via

$$f(X) = X^T X\,.$$

Then $O(n) = f^{-1}\{\mathrm{Id}\}$ and we need to identify the range of $df(x)$.

We have for all $H \in M_{n\times n}(\mathbb{R})$ that

$$df(X)H = H^T X + X^T H \in S(n).$$

It remains to show that for all $X \in O(n)$ the map $df(X)$ is surjective. Let $Z \in S(n)$ and define $H := \frac{1}{2}XZ$. Then

$$df(X)H = \frac{1}{2}Z^T X^T X + \frac{1}{2}X^T X Z = \frac{1}{2}(Z^T + Z) = Z\,.$$

Hence the range of $df(X)$ is $S(n)$, thus rank $df(X) = \dim S(n) = \frac{1}{2}n(n+1)$ and $O(n)$ is a submanifold of dimension $k = n^2 - \frac{1}{2}n(n+1) = \frac{1}{2}n(n-1)$.

This is an example of a *Lie group*, a manifold which is also a group, such that the group operations of multiplication and inversion are given by differentiable maps. Many symmetry groups in physical problems turn out to be Lie groups. This important topic linking geometry and algebra is the subject of the Part C course Lie Groups. There are also many excellent books on the subject [**?**, **?**, **?**].

We now define an important concept for manifolds, the tangent space at a point of the manifold.

Definition 4.1.7. (Tangent vector, tangent space, normal vector) Let $M \subseteq \mathbb{R}^n$ be a k-dimensional submanifold of $\mathbb{R}^n$.

1. We call $v \in \mathbb{R}^n$ a *tangent vector* to M at $x \in M$, if there exists a C^1-function $\gamma\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$, such that $\gamma(t) \in M$ for all $t \in (-\varepsilon, \varepsilon)$, $\gamma(0) = x$ and $\gamma'(0) = v$.
2. The set of all tangent vectors to M at x is called the *tangent space* to M at x, and we denote it by T_xM.
3. We call $w \in \mathbb{R}^n$ a *normal vector* to M at $x \in M$ if $\langle w, v\rangle = 0$ for all $v \in T_xM$. Thus the set of all normal vectors to M at x is precisely the orthogonal complement $T_xM^\perp$ of T_xM in $\mathbb{R}^n$.

Next we prove the generalisation of the property that 'the gradient is perpendicular to the level sets of a function' (see Corollary **??**). This result in particular also shows that T_xM is indeed a k-dimensional vector space and as a consequence that the space of normal vectors is an $(n-k)$-dimensional vector space.

Proposition 4.1.8. *Let M be a k-dimensional submanifold of $\mathbb{R}^n$. Let Ω be an open subset of $\mathbb{R}^n$ and let $f \in C^1(\Omega, \mathbb{R}^{n-k})$ be such that $M \cap \Omega = f^{-1}\{0\}$ and $\mathrm{rank} Df(x) = n-k$ for all $x \in \Omega$. Then we have*

$$T_xM = \ker Df(x),$$

for all $x \in M \cap \Omega$, that is the tangent space equals the kernel of $Df(x)$.

Proof. We first claim that $T_xM \subseteq \ker Df(x)$:

Indeed, let $v \in T_xM$, then there exists $\gamma : (-\varepsilon, \varepsilon) \to M$ such that

$$\gamma(0) = x \quad \text{and} \quad \gamma'(0) = v.$$

It follows for all $t \in (-\varepsilon, \varepsilon)$, that $f(\gamma(t)) = 0$. Hence

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t)) = Df(\gamma(t))\gamma'(t)$$

and for $t = 0$ we find $0 = Df(x)v$, hence $v \in \ker Df(x)$.

Now recall that, possibly after a suitable relabelling, we can assume in view of Proposition **??** that $x = (z_0, y_0) \in \mathbb{R}^k \times \mathbb{R}^{n-k}$ and that there exist open subsets $U \subseteq \mathbb{R}^k$ with $z_0 \in U$ and $V \subseteq \mathbb{R}^{n-k}$ with $y_0 \in V$ and a function $g \in C^1(U, V)$ with $g(z_0) = y_0$ such that

$$M \cap (U \times V) = \{(z, g(z)) \mid z \in U\}.$$

We define $G\colon U \to \mathbb{R}^n$ by $G(z) = (z, g(z))$ and for an arbitrary $\xi \in \mathbb{R}^k$ and sufficiently small ε we let $\gamma\colon (-\varepsilon, \varepsilon) \to M$ be given by

$$\gamma(t) = G(z_0 + t\xi)\,.$$

Then $\gamma'(t) = DG(z_0 + t\xi)\xi$ and

$$\gamma'(0) = DG(z_0)\xi \quad \text{with } DG(z_0) = \begin{pmatrix} \mathrm{Id}_k \\ Dg(z_0) \end{pmatrix}.$$

Hence $\mathrm{im}DG(z_0) \subseteq T_xM$ and thus we have shown so far that $\mathrm{im}DG(z_0) \subseteq T_xM \subseteq \ker Df(x)$. But $DG(z_0)$ is obviously injective, hence $\dim \mathrm{im}Dg(z_0) = k = n - \mathrm{rank}Df(x) = \dim \ker Df(x)$. Hence $\mathrm{im}Dg(z_0) = \ker Df(x) = T_xM$. $\square$

Example 4.1.9.

a) Let $A \in M_{n\times n}(\mathbb{R})$ be symmetric, $\det A \neq 0$, and let $M = \{x \in \mathbb{R}^n \mid f(x) = \langle x, Ax\rangle - 1 = 0\}$. We have $Df(x) = 2(Ax)^T$ and since A is regular and $x \neq 0$ for $x \in M$ we have $\text{rank} Df(x) = 1$ for all $x \in M$. Hence M is an $(n-1)$-dimensional submanifold of $\mathbb{R}^n$ and Proposition **??** implies that

$$T_x M = \{v \in \mathbb{R}^n \mid 2\langle v, Ax\rangle = 0\}$$

and the 1-dimensional space of normal vectors is spanned by Ax.

In particular, if $A = \text{Id}$, then M is the unit sphere in $\mathbb{R}^n$, the space of normal vectors at x is spanned by x and the tangent space is given by all vectors which are perpendicular to x.

b) We have seen that a (two-dimensional) torus (in $\mathbb{R}^3$) is given by

$$T = \left\{(x,y,z) \in \mathbb{R}^3 \,|\, f(x,y,z) = \left(\sqrt{x^2+y^2} - R\right)^2 + z^2 - r^2 = 0\right\},$$

where $0 < r < R$. Hence, the space of normal vectors is

$$(T_{(x,y,z)}M)^{\perp} = \left\{w \in \mathbb{R}^3 \,|\, w = \lambda \nabla f(x,y,z) \text{ for some } \lambda \in \mathbb{R}\right\}.$$

c) Consider the orthogonal group

$$O(n) = \left\{X \in M_{n\times n}(\mathbb{R}) \mid f(X) = X^T X - \text{Id} = 0\right\}.$$

We have seen that $O(n)$ is a submanifold of $\mathbb{R}^{n^2}$ of dimension $\frac{1}{2}n(n-1)$ and we also have $\text{Id} \in O(n)$. With $df(X)H = X^T H + H^T X$ and $df(\text{Id})H = H^T + H$ it follows

$$T_{\text{Id}} M = \{H \in M_{n\times n}(\mathbb{R}) \mid H^T + H = 0\},$$

that is the tangent space at Id is the skew-symmetric matrices.

In fact, the tangent space to a general Lie group at the identity element carries a very rich algebraic structure, beyond the basic vector space structure it has as a tangent space. This structure is that of a *Lie algebra*, a vector space V which also carries a skew-symmetric bilinear map $V \times V \to V$ satisfying a certain identity called the Jacobi identity. Writing the bilinear form, as traditional, using bracket notation $[X, Y]$, the Jacobi identity is

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$$

For matrix Lie groups such as $O(n)$, the bracket is actually given by $[X, Y] = XY - YX$ (check that this does indeed satisfy the Lie algebra axioms!). For further material on Lie algebras, we refer the reader to the books [**?**, **?**] and also the Part C course Lie Algebras.

4.2 Extremal problems with constraints

We now consider an application of these ideas to the study of constrained extremisation problems. We know from elementary vector calculus that the critical points of a function on $\mathbb{R}^n$ are the points where the gradient vanishes. We now want to consider the more subtle problem of extremising a function subject to a constraint–that is, finding the extrema of a function on some subset of Euclidean space defined by one or more equations.

Let us first consider the simplest case, where we have two functions $f, g \in C^1(\mathbb{R}^2)$, and our goal is to minimise (or maximise) g under the constraint that $f(x, y) = 0$.

We can often (under some assumptions on f) think of the set

$$\Gamma = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = 0\}$$

as a curve in $\mathbb{R}^2$. Let (x_0, y_0) be such that for some $\varepsilon > 0$ and all $(x, y) \in \Gamma \cap B_\varepsilon(x_0, y_0)$:

$$g(x_0, y_0) \leq g(x, y).$$

Suppose that $\nabla f(x_0, y_0) \neq 0$, and assume without loss of generality that $\partial_y f(x_0, y_0) \neq 0$. The Implicit Function Theorem guarantees that we can represent Γ in an open neighbourhood of (x_0, y_0) as $(x, \varphi(x))$ for $x \in I$, where I is an open interval with $x_0 \in I, \varphi \in C^1(I)$ and $\varphi(x_0) = y_0$. The tangent to Γ at $(x, \varphi(x))$ is given by the vector $(1, \varphi'(x))$ and since the gradient is perpendicular to the level sets we have

$$\begin{pmatrix} 1 \\ \varphi'(x_0) \end{pmatrix} \perp \nabla f(x, \varphi(x)).$$

Define $G(x) = g(x, \varphi(x))$ and consider the point (x_0, y_0) where g has a local minimiser on Γ. Then, by Fermat's Theorem and the Chain Rule,

$$0 = G'(x_0) = \partial_x g(x_0, y_0) + \partial_y g(x_0, y_0)\varphi'(x_0) = \left\langle \nabla g(x_0, y_0), \begin{pmatrix} 1 \\ \varphi'(x_0) \end{pmatrix} \right\rangle.$$

Hence there exists $\lambda \in \mathbb{R}$, such that

$$\nabla g(x_0, y_0) = \lambda \nabla f(x_0, y_0).$$

We can interpret this geometrically as follows. If the extremisation problem were unconstrained, the criterion would just be vanishing of ∇g. As we are just looking for extrema on the constraint curve $f = 0$, we actually want the component of ∇g tangent to the curve to vanish, ie ∇g should be normal to the curve $f = 0$. This is exactly the condition that ∇g is proportional to ∇f (cf Proposition **??**).

In the following, we want to generalize this procedure by extremising a function g on a submanifold M of $\mathbb{R}^n$ given by a system of equations $f = 0$. The following theorem will provide a necessary condition for an extremal point.

Theorem 4.2.1. (Theorem on Lagrange multipliers) *Let $\Omega \subseteq \mathbb{R}^n$ be open, $g \in C^1(\Omega)$ and $f \in C^1(\Omega, \mathbb{R}^{n-k})$. If $x_0 \in f^{-1}\{0\}$ is a local extremum of g on $f^{-1}\{0\}$, that is there exists an open neighbourhood V of x_0 such that for all $x \in V$ which satisfy $f(x) = 0$ we have*

$$g(x) \geq g(x_0) \qquad (\textit{or } g(x) \leq g(x_0)),$$

and if $\mathrm{rank} Df(x_0) = n - k$*, then there exist $\lambda_1, \ldots, \lambda_{n-k} \in \mathbb{R}$, such that*

$$\nabla g(x_0) = \sum_{i=1}^{n-k} \lambda_i \nabla f_i(x_0).$$

The numbers $\lambda_1, \ldots, \lambda_{n-k}$ are called Lagrange multipliers*.*

Proof. If V is sufficiently small then we have for all $x \in V$ that $\mathrm{rank} Df(x) = n - k$, hence $M = f^{-1}\{0\} \cap V$ is a k-dimensional submanifold of $\mathbb{R}^n$. For a $v \in T_{x_0}M$ let $\gamma : (-\varepsilon, \varepsilon) \to M$ be a C^1-function such that $\gamma(0) = x_0$ and $\gamma'(0) = v$. The function $g \circ \gamma$ has in $t = 0$ a local minimum. Thus

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} g(\gamma(t))|_{t=0} = \left\langle \nabla g(\gamma(t)), \gamma'(t) \right\rangle|_{t=0} = \langle \nabla g(x_0), v \rangle$$

and thus $\nabla g(x_0) \in (T_{x_0}M)^{\perp}$. Furthermore we have for all $x \in M$ and $i = 1, \ldots, n-k$ that

$$f_i(x) = 0, \qquad \text{and thus in particular } \nabla f_i(x_0) \perp T_{x_0}M.$$

Since $\mathrm{rank} Df(x_0) = n-k$ the vectors $\nabla f_i(x_0)$ are linearly independent and form a basis of $(T_{x_0}M)^{\perp}$. Hence there exist $\lambda_1, \ldots, \lambda_{n-k}$ such that $\nabla g(x_0) = \sum_{i=1}^{n-k} \lambda_i \nabla f_i(x_0)$. □

Again, we can interpret this as saying that ∇g is normal to the submanifold $f = 0$, and the normal space is spanned by $\nabla f_i : (i = 1, \ldots, n-k)$.

Example 4.2.2.

a) We want to determine the minimal and maximal value of $g(x, y, z) = 5x + y - 3z$ on the intersection of the plane $E = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}$ with the unit sphere $S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 - 1 = 0\}$.

   In other words we want to determine the extremal values of g on

$$M = \left\{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = \begin{pmatrix} x + y + z \\ x^2 + y^2 + z^2 - 1 \end{pmatrix} = 0\right\}.$$

   We compute

$$Df(x, y, z) = \begin{pmatrix} 1 & 1 & 1 \\ 2x & 2y & 2z \end{pmatrix}$$

   and conclude that $\mathrm{rank} Df(x, y, z) = 2$ for all $(x, y, z) \in M$ (as $(1, 1, 1) \notin M$). Hence M is a one-dimensional submanifold of $\mathbb{R}^3$. Furthermore M is compact, g

is continuous and hence g attains its infimum and supremum on M. Theorem **??** implies that there exist $\lambda_1, \lambda_2 \in \mathbb{R}$ such that in an extremal point (x, y, z) we have $\nabla g(x, y, z) = \lambda_1 \nabla f_1(x, y, z) + \lambda_2 \nabla f_2(x, y, z)$. Thus, we find for an extremal point (x, y, z) that

$$\begin{pmatrix} 5 \\ 1 \\ -3 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}$$

and $f(x, y, z) = 0$. This implies that

$$\lambda_1 = 5 - 2\lambda_2 x = 1 - 2\lambda_2 y = -3 - 2\lambda_2 z \tag{4.1}$$

and thus (since $\lambda_2 = 0$ is excluded)

$$x = y + 2/\lambda_2 \qquad \text{and} \qquad z = y - 2/\lambda_2\,.$$

Next we conclude from $f_1(x, y, z) = 0$ that $y = 0$ and hence $x = 2/\lambda_2$ and $z = -2/\lambda_2$. Then it follows from $f_2(x, y, z) = 0$ that $\lambda_2 = \pm 2\sqrt{2}$. Thus g attains its maximum, which is $4\sqrt{2}$, at $(2^{-1/2}, 0, -2^{-1/2})$ and its minimum, which is $-4\sqrt{2}$, at $(-2^{-1/2}, 0, 2^{-1/2})$.

b) (Inequality between geometric and arithmetic mean)

Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be given by $g(x) = \Pi_{i=1}^n x_i$. Consider the set

$$M = \left\{ x \in \mathbb{R}^n \mid f(x) = \big(\sum_{i=1}^n x_i\big) - 1 = 0, x_i > 0 \right\},$$

which obviously defines an $(n-1)$-dimensional submanifold of $\mathbb{R}^n$. Since $\overline{M}$ is compact and g is continuous, g attains its supremum at a point $z \in \overline{M}$. Since $g = 0$ on ∂M and positive in the interior we conclude that indeed $z \in M$.

Now we may use Lagrange multipliers to find z and hence the maximum value of g on $\overline{M}$. We can then deduce the arithmetic mean/geometric mean inequality

$$\left(\prod_{i=1}^n a_i\right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^n a_i$$

The details of the proof are an exercise on the second problem sheet.

c) (Eigenvalues of symmetric real matrices)

Let $A \in M_{n\times n}(\mathbb{R})$ be symmetric and let $g\colon \mathbb{R}^n \to \mathbb{R}$ be given by

$$g(x) = \langle x, Ax \rangle \quad \text{und} \quad M = \{x \in \mathbb{R}^n \,|\, f(x) = |x|^2 - 1 = 0\}.$$

M is an $(n-1)$-dimensional submanifold of $\mathbb{R}^n$ since for all $x \in M$ we have

$$\nabla f(x) = 2x \neq 0.$$

M is compact, so that g attains its supremum on M in a point x_0. By Theorem **??** there exists $\lambda \in \mathbb{R}$, such that $\nabla g(x_0) = \lambda \nabla f(x_0)$ that is $Ax_0 = \lambda x_0$. This implies that x_0 is an eigenvector of A with eigenvalue λ. Thus, we have shown that every real symmetric $n \times n$ matrix has a real eigenvalue. We also find that

$$\lambda \langle x_0, x_0 \rangle = \langle x_0, Ax_0 \rangle = g(x_0) \qquad \text{and thus } \ \lambda = g(x_0)\,.$$

Since $g(x_0)$ is the maximal value of g on M this also implies that λ is the largest eigenvalue of A.

Bibliography

[1] T. Bröcker and T. tom Dieck, *Representations of compact Lie groups*, Graduate Texts in Mathematics, Springer, 1985.

[2] R. Carter,. G. Segal and I. Macdonald, *Lectures on Lie groups and Lie algebras*, LMS Student Texts, CUP, 1995.

[3] K. Erdmann and M. Wildon, *Introduction to Lie algebras*, Springer, 2006.

[4] B. Hall, *Lie groups, Lie algebras and representations : an elementary introduction*, Graduate Texts in Mathematics, Springer, 2003.

[5] S.G. Krantz and H. Parks, *The implicit function theorem*, Birkhäuser, 2002.

[6] M. Spivak, *Calculus on Manifolds*, W.A. Benjamin, 1965.