ADDITIONAL NOTES ON DIFFERENTIATION

GERD SCHMALZ

1. The o notation

It is very convenient to use the notation $o(h)$ for any function that, after been divided by $h$, tends to zero as $h$ tends to zero. More precisely, if a function $f(h)$, which is defined in some interval around 0 has the property

$$
\lim_{h \to 0} \frac{f(h)}{h} = 0
$$

we may write briefly

$$
f(h) = o(h) \text{ as } h \to 0.
$$

Here $h$ can be any expression.

Intuitively, this means that $f(h)$ tends to zero faster than $h$ itself. E.g. $f(h) = h^2 = o(h)$ because $\frac{f(h)}{h} = \frac{h^2}{h} = h$ tends to zero as $h$ tends to zero. Other examples are $f(x) = x \sin x = o(x)$, $f(x) = x^2 \ln x = o(x)$. On the other hand $f(x) = \sqrt{x} \neq o(x)$.

We will use the $o$ notation when the particular form of the function does not matter and we are only interested in the particular limit behaviour of the function.

The $o$ arithmetic is rather simple as the following examples show:

$$
o(h) + o(h) = o(h)
$$

This means that sums of $o$’s can be combined into one $o$ which reflects the fact that

$$
\lim_{h \to 0} \frac{f(h)}{h} = 0 \text{ and } \lim_{h \to 0} \frac{g(h)}{h} = 0 \implies \lim_{h \to 0} \frac{f(h) + g(h)}{h} = 0.
$$

Another rule is

$$
g(h) o(h) = o(h) \text{ if } g(h) \text{ is a bounded function.}
$$

Using the $o$ notation we can express the fact that $A = f'(a)$ is the derivative of $f$ at $a$ by

$$
f(x) = f(a) + A(x - a) + o(x - a).
$$

Indeed,

$$
f(x) - f(a) - A(x - a) = o(x - a)
$$

means

$$
\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = A = 0,
$$

i.e.

$$
\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = A.
$$
2. Differentiation in several variables

The formula
\[ \Delta f = f(x) - f(a) = A\Delta x + o(\Delta x) \]
means that the increment of the function \( \Delta f \) at some point \( a \) is approximated by the linear function \( A\Delta x = A(x - a) \) up to some error term \( o(\Delta x) \) which tends to zero faster than \( \Delta x \), i.e. is very small when \( x \) is close enough to \( a \).

Approximation of non-linear functions by linear ones is the whole point of differential calculus and can be easily generalised to many variables. Let \( f \) be a \( \mathbb{R}^k \)-valued function defined on some ball around a point \( a \) in \( \mathbb{R}^n \). Then
\[
f(x) = f(a) + A(x - a) + o(\|x - a\|)\]
can be read as follows: the vector-valued increment of the function \( \Delta f = f(x) - f(a) \) is equal to a linear mapping
\[
A(x - a) = A\Delta x
\]
up to a (vector-valued) function that tends to zero faster than \( \|\Delta x\| \) as \( \Delta x \to 0 \). The linear mapping \( A \) takes arguments in \( \mathbb{R}^n \) (namely \( \Delta x \)) and values in \( \mathbb{R}^k \), hence it can be expressed as multiplication of a \( k \times n \) matrix \( A \) with the \( n \times 1 \) column \( \Delta x \).

It is important to understand that the multivariable analog of the derivative for a mapping from \( \mathbb{R}^n \) to \( \mathbb{R}^k \) is a \( k \times n \) matrix. The matrix \( A \) is called Jacobi matrix and consists of the partial derivatives
\[
A = \left( \begin{array}{cccc}
\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_k}{\partial x_1} & \cdots & \frac{\partial f_k}{\partial x_n}
\end{array} \right).
\]

The linear mapping \( \Delta x \mapsto A\Delta x \) is called the differential of \( f \) at \( a \). We write
\[
df = A\Delta x \text{ or } df = A\,dx.
\]
(We have used \( \Delta x = dx \).)

Now most of the statements and proofs carry over from one-variable calculus to multivariable calculus. In particular, the multivariable chain rule becomes very simple: Let \( f \) be a \( \mathbb{R}^k \)-valued function defined in some ball in \( \mathbb{R}^m \) around \( b \) and \( g \) be a \( \mathbb{R}^m \)-valued function defined in some ball in \( \mathbb{R}^n \) around \( a \) such that \( g(a) = b \) and \( g \) is differentiable at \( a \) and \( f \) is differentiable at \( b \), i.e.

\[
f(y) = f(b) + B(y - b) + o(\|y - b\|) \\
g(x) = g(a) + A(x - a) + o(\|x - a\|).
\]

Then \( f(g(x)) \) is differentiable at \( a \) and
\[
f(g(x)) = f(g(a)) + BA(x - a) + o(\|x - a\|),
\]
where \( BA \) is the matrix product of \( B \) and \( A \). Thus, the chain rule reduces to multiplying matrices. The proof is similar to the proof for one-variable functions.
3. The implicit mapping theorem

One of the most important tasks in calculus is to solve systems of non-linear equations
\begin{align*}
f_1(x_1, \ldots, x_n) &= 0 \\
\vdots \\
f_k(x_1, \ldots, x_n) &= 0
\end{align*}
where \(f_1, \ldots, f_k\) are functions of \(x_1, \ldots, x_n\). To solve such a system usually means to express \(k\) of the \(x\) variables (e.g. the first \(k\) variables \(x_1, \ldots, x_k\)) as functions of the remaining \(n-k\) variables
\begin{align*}
x_1 &= g_1(x_{k+1}, \ldots, x_n) \\
\vdots \\
x_k &= g_k(x_{k+1}, \ldots, x_n)
\end{align*}
in such a way that inserting \(g_1, \ldots, g_k\) instead of \(x_1, \ldots, x_k\) turns the system (1) into an identity, i.e.
\begin{align*}
f_1(g_1(x_{k+1}, \ldots, x_n), \ldots, g_k(x_{k+1}, \ldots, x_n), x_{k+1}, x_n) &\equiv 0 \\
\vdots \\
f_k(g_1(x_{k+1}, \ldots, x_n), \ldots, g_k(x_{k+1}, \ldots, x_n), x_{k+1}, x_n) &\equiv 0
\end{align*}
If (1) was a system of linear equations
\begin{align*}
f_1(x_1, \ldots, x_n) &= a_{11}x_1 + \cdots + a_{1n}x_n = 0 \\
\vdots \\
f_k(x_1, \ldots, x_n) &= a_{k1}x_1 + \cdots + a_{kn}x_n = 0
\end{align*}
we could solve by using the Gaussian algorithm and we would find a solution as desired if the determinant of the matrix of the first \(k\) columns of coefficients
\[
\det \begin{pmatrix}
a_{11} & \cdots & a_{1k} \\
\vdots & \vdots & \vdots \\
a_{k1} & \cdots & a_{kk}
\end{pmatrix} \neq 0.
\]
The implicit mapping theorem states

**Theorem 1.** If \(f\) is a \(\mathbb{R}^k\)-valued mapping defined on some ball around \(a \in \mathbb{R}^n\) such that \(f(a) = 0\) and all partial derivatives \(\frac{\partial f_i}{\partial x_j}\) are defined and continuous and the determinant
\[
\det \begin{pmatrix}
\frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_k}(a) \\
\vdots & \vdots & \vdots \\
\frac{\partial f_k}{\partial x_1}(a) & \cdots & \frac{\partial f_k}{\partial x_k}(a)
\end{pmatrix} \neq 0
\]
then the system $f(x) = 0$ has a solution of the form (2).

The following inverse mapping theorem is a consequence of Theorem 1.

**Theorem 2.** If $f$ is a $\mathbb{R}^n$-valued mapping defined on some ball around $a \in \mathbb{R}^n$ such that $f(a) - b$ and all partial derivatives $\frac{\partial f_i}{\partial x_j}$ are defined and continuous and the determinant

$$
\det \begin{pmatrix}
\frac{\partial f_1}{\partial x_1}(a) & \ldots & \frac{\partial f_1}{\partial x_n}(a) \\
\vdots & \ddots & \vdots \\
\frac{\partial f_n}{\partial x_1}(a) & \ldots & \frac{\partial f_n}{\partial x_n}(a)
\end{pmatrix} \neq 0
$$

then the mapping $f(x)$ has an inverse $g(y)$ defined on some ball around $b$, i.e. $g(f(x)) \equiv x$ and $f(g(y)) \equiv y$.

To prove this we apply the implicit mapping theorem to the $\mathbb{R}^n$-valued mapping $f(x) - y$ on $\mathbb{R}^{2n}$. 