Euclidean Norm

If \( x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \), then its **Euclidean norm** is \( \| x \| = \sqrt{\sum_{j=1}^{n} x_j^2} \).

Maps

In the notes and the textbook, maps (i.e. functions) are written \( f : \mathbb{R}^n \to \mathbb{R}^m \) when the domain of definition of \( f \) contains an open subset of \( \mathbb{R}^n \) and the range of \( f \) is a subset of \( \mathbb{R}^m \). The domain of definition usually is not explicitly denoted.

Derivatives

“Big oh” and “little oh” notation: If \( f : \mathbb{R}^n \to \mathbb{R}^m \) and \( g : \mathbb{R}^n \to \mathbb{R}^m \) are two maps defined (at least) in an open neighbourhood of a point \( x_0 \in \mathbb{R}^n \), and if \( \lim_{x \to x_0} g(x) = 0 \), then we say that

\[
 f(x) = o(g(x)) \text{ as } x \to x_0
\]

(“\( f \) is little oh of \( g \)”’) if \( \lim_{x \to x_0} \| f(x) \|/\| g(x) \| = 0 \) (roughly speaking, if \( f \) goes to zero “faster” than \( g \) does, as \( x \to x_0 \)). For example, for \( n = 1 \) and \( m = 1 \) we have

\[
e^x - 1 - x = o(|x|) \text{ as } x \to 0.
\]

Similarly, we say that

\[
f(x) = O(g(x)) \text{ as } x \to x_0
\]

(“\( f \) is big oh of \( g \)”’) if there is a constant \( C \geq 0 \) such that \( \| f(x) \| \leq C \| g(x) \| \) for all \( x \) sufficiently close to \( x_0 \) (roughly speaking, if \( f \) goes to zero “at least as fast” as \( g \) does, as \( x \to x_0 \)). For example, we have (see Taylor Expansions, below)

\[
e^x - 1 - x = O(|x|^2) \text{ as } x \to 0.
\]

The derivative of a map: If \( f : \mathbb{R}^n \to \mathbb{R}^m \) is a continuous map (i.e. function), then the **derivative** (sometimes called the **Jacobian matrix**) of \( f \) is the \( m \times n \) matrix

\[
f_x(x) = \left( \frac{\partial f_i(x)}{\partial x_j} \right),
\]

where \( i = 1, 2, \ldots, m \) and \( j = 1, 2, \ldots, n \), provided all these partial derivatives exist.

For example, if \( f : \mathbb{R}^3 \to \mathbb{R}^2 \) is given by

\[
f(x) = \begin{pmatrix}
x_1 x_2 x_3 - 1 \\
x_1^2 - x_3
\end{pmatrix},
\]

then its derivative is the \( 2 \times 3 \) matrix of partial derivative expressions

\[
f_x(x) = \begin{pmatrix}
x_2 x_3 & x_1 x_3 & x_1 x_2 \\
x_1^2 & 0 & -1
\end{pmatrix}.
\]
and its derivative evaluated at the point $x_0 = (1,1,1)$ is the constant $2 \times 3$ matrix

$$f_x(x_0) = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & -1 \end{pmatrix}$$

In general, if all first order partial derivatives of all components of $f$ exist at the point $x_0$, then it can be proved that the derivative of $f$ at $x_0$ is the unique matrix $A = f_x(x_0)$ that satisfies

$$f(x_0 + h) = f(x_0) + Ah + R(h)$$

for all $h \in \mathbb{R}^n$ near 0, where $R(h) = o(\|h\|)$ as $h \to 0$. This property is sometimes taken as the definition of $f_x(x_0)$.

If $k \geq 1$ is an integer, we say the map $f$ is $C^k$ if all partial derivatives of all components of $f$, up to and including order $k$, exist and are continuous. We say $f$ is smooth if it is $C^k$ for some integer $k \geq 1$. If $f$ is $C^2$ in an open neighbourhood of $x = x_0$, then it can be proven that

$$f(x_0 + h) = f(x_0) + f_x(x_0)h + R(h)$$

for all $h \in \mathbb{R}^n$ near 0, where $R(h) = O(\|h\|^2)$ as $h \to 0$. Notice that this gives us a little more information than the statement $R(h) = o(\|h\|)$ as $h \to 0$.

Chain Rule

If $g : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^k$ are two maps, their composition is the map $f \circ g : \mathbb{R}^n \to \mathbb{R}^k$ defined by

$$f \circ g(x) = f(g(x)),$$

or

$$f \circ g(x) = f(u), \quad \text{where} \quad u = g(x).$$

By the Chain Rule (Theorem), under suitable conditions we have

$$(f \circ g)_x(x) = f_u(g(x))g_x(x),$$

where the right hand side is the matrix product of the $k \times m$ matrix of partial derivatives of $f$, evaluated at $u = g(x)$, with the $m \times n$ matrix of partial derivatives of $g$, evaluated at $x$.

Inverse Maps

Consider the expression $y = f(x)$, where $f : \mathbb{R}^n \to \mathbb{R}^n$ is a map. We ask whether we can solve uniquely for $x$ in terms of $y$, as $x = g(y)$, in other words, whether $g = f^{-1}$ exists as a well-defined map (i.e. function).

Theorem A.1 (Inverse Function Theorem). Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is a $C^k$ map, $k \geq 1$, in an open neighbourhood of $x_0$, and let $y_0 = f(x_0)$. If the $n \times n$ matrix $f_x(x_0)$ is nonsingular, then there exists a unique, locally defined (i.e. whose domain of definition contains an open neighbourhood of $y_0$ in $\mathbb{R}^n$), $C^k$ map $g : \mathbb{R}^n \to \mathbb{R}^n$, with $x = g(y)$, such that

$$g(f(x)) = x$$

for all $x$ belonging to an open neighbourhood of $x_0$ in $\mathbb{R}^n$.

The unique (locally defined) map $g$ in this theorem is called the inverse map (or inverse function) for $f$, near $(x_0, y_0)$, and is denoted by $g = f^{-1}$. Recall that if $f$ is not 1-to-1, then $g = f^{-1}$ could depend on $x_0$. For example, when $n = 1$ the inverse map for $f(x) = x^2$ that is defined near $(x_0, y_0) = (1,1)$, is different from the inverse map that is defined near $(x_0, y_0) = (-1,1)$.
Implicit Function Theorem (Important!)

For a smooth map $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, we will often want solutions to the equation

$$F(x, y) = 0$$

with $y$ expressed in terms of $x$, as $y = f(x)$, defining a smooth map $f$ near a known specific solution. Suppose that $F(x_0, y_0) = 0$ (i.e. $x = x_0$, $y = y_0$ is a known specific solution), then we consider the $m \times m$ matrix of partial derivatives

$$F_y(x_0, y_0) = \left( \frac{\partial F_i(x_0, y_0)}{\partial y_j} \right).$$

**Theorem A.2** (Implicit Function Theorem). Suppose $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ is a $C^k$ map, $k \geq 1$, in an open neighbourhood of $(x_0, y_0)$ in $\mathbb{R}^n \times \mathbb{R}^m$, and suppose $F(x_0, y_0) = 0$. If the $m \times m$ matrix $F_y(x_0, y_0)$ is nonsingular, then there exists a unique, locally defined $C^k$ map $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $y = f(x)$, such that $f(x_0) = y_0$ and

$$F(x, f(x)) = 0$$

for all $x$ in the domain of definition of $f$ in $\mathbb{R}^n$. Moreover,

$$f_x(x_0) = -[F_y(x_0, y_0)]^{-1} F_x(x_0, y_0).$$

For example, if $F : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ is given by

$$F(x, y_1, y_2) = \begin{pmatrix} xy_1 y_2 - 1 \\ x^2 - y_2 \end{pmatrix},$$

then $F(1, 1, 1) = 0$ in $\mathbb{R}^2$, and its partial derivative with respect to $y = (y_1, y_2) \in \mathbb{R}^2$ is

$$F_y(x, y) = \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} \\ \frac{\partial F_2}{\partial y_1} & \frac{\partial F_2}{\partial y_2} \end{pmatrix} = \begin{pmatrix} xy_2 & xy_1 \\ 0 & -1 \end{pmatrix}$$

which evaluates at the point $(x_0, y_0) = (1, 1, 1)$ to

$$F_y(x_0, y_0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

which is nonsingular, and therefore by the Implicit Function Theorem, $y = (y_1, y_2)$ can be uniquely solved near $x = 1$, $y_1 = 1$, $y_2 = 1$ as some smooth map $f : \mathbb{R} \rightarrow \mathbb{R}^2$, $y = f(x) = (f_1(x), f_2(x))$ with $f_1(1) = 1$ and $f_2(1) = 1$. (In this example, $y_1 = f_1(x)$ and $y_2 = f_2(x)$ can be found explicitly, and the conclusions of the theorem can be checked by explicit calculation.)

Taylor Expansions

A $C^{k+1}$ ($k \geq 1$) map $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ can be approximated, near a point $x_0$ in the interior of its domain, by its Taylor polynomial of degree $k$ at $x_0 = (x_{01}, x_{02}, \ldots, x_{0n})$:

$$f(x) = \sum_{|i| = 0}^{k} \frac{1}{i_1! i_2! \cdots i_n!} \frac{\partial^{[i]} f(x_0)}{\partial x_1^{i_1} \partial x_2^{i_2} \cdots \partial x_n^{i_n}} (x_1 - x_{01})^{i_1} (x_2 - x_{02})^{i_2} \cdots (x_n - x_{0n})^{i_n} + R(x)$$

where $|i| = i_1 + i_2 + \cdots + i_n$ and the remainder $R(x)$ satisfies $R(x) = O(||x - x_0||^{k+1})$. (This is a consequence of the multivariable version of Taylor’s Theorem with Remainder.)

3