MATH 552
A. Review of (Multivariable) Differential Calculus

Euclidean norm

If $x = (\, x_1 \quad x_2 \quad \cdots \quad x_n \,)^{\mathsf{T}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, then its **Euclidean norm** is $\|x\| = \sqrt{\sum_{j=1}^n x_j^2}$.

(If $n = 1$, the Euclidean norm $\|x\|$ is the same as the absolute value $|x|$.)

Maps

In our notes and the textbook, **maps** (i.e. functions) are written

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

where the domain of definition of f *contains an open subset* of $\mathbb{R}^n$ and the range of f is a subset of $\mathbb{R}^m$. Often, the domain of definition is not explicitly specified.

The derivative of a (multivariable) map

"Big oh" and "little oh" notation: Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be two maps defined in an open neighbourhood of a point $x_0 \in \mathbb{R}^n$, and suppose that $\lim_{x \to x_0} g(x) = 0$.

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

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

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, in $\mathbb{R}^1$ we have (see Taylor Expansions, below)

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

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

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

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

For example, if the map $f : \mathbb{R}^3 \to \mathbb{R}^2$ 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 2×3 matrix of real-valued functions

$$f_x(x) = \begin{pmatrix} x_2 x_3 & x_1 x_3 & x_1 x_2 \\ 2x_1^2 & 0 & -1 \end{pmatrix}$$

and its derivative evaluated at the point $x_0 = (1, 1, 1)$ is the 2×3 matrix of constants

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

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 $p \geq 1$ is an integer, we say the map f is C^p if all partial derivatives of all components of f, up to and including order p, exist and are continuous. Often, we say f is **smooth** if it is C^p for some integer $p \geq 1$ and we don't care how large p is. 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 $h \to 0$.

The 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 (you can look them up) 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 or functions

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^p map, $p \geq 1$, 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^p 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 (possibly only 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 $(x_0, y_0) = (-1, 1)$. Notice that the (local) inverse map is at least as smooth as the original map.

The Implicit Function Theorem (Important!)

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

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

with y expressed explicitly 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 $m \times m$ (square) matrix of partial derivatives evaluated at the known solution

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

Theorem A.2 (Implicit Function Theorem)**.** *Suppose $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is a C^p map, $p \geq 1$, 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^p map $f : \mathbb{R}^n \to \mathbb{R}^m$, $y = f(x)$, 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 the map $F : \mathbb{R}^3 \to \mathbb{R}^2$ is given by

$$F(x, y_1, y_2) = \begin{pmatrix} xy_1y_2 - 1 \\ x^2 - y_2 \end{pmatrix},$$

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

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

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 $f : \mathbb{R} \to \mathbb{R}^2$, $y = f(x) = (f_1(x), f_2(x))$ 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 you can – and should – check the conclusions of the theorem by explicit calculation.) Notice that the (local) implicitly defined solution is always at least as smooth as the original map.

Multivariable Taylor polynomial approximations

Any C^p $(p \geq 2)$ map $f : \mathbb{R}^n \to \mathbb{R}^m$ can be approximated, near a point x_0 in the interior of its domain, by its Taylor polynomial of degree $p-1$ at $x_0 = (x_{01}, x_{02}, \ldots x_{0n})$:

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

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