DIFFERENTIAL GEOMETRY
Multivariable Calculus, a refresher

1 Preliminaries

In this section I collect some results from multivariable calculus that play an important role in the course. If you don’t know them yet, you should try to familiarize yourself with them. For details, consult any introductory analysis textbook.

1.1 Differentials

Suppose $U$ is an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^m$. We say that $f$ is differentiable at the point $a \in U$ iff there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to a} \frac{|f(x) - f(a) - A(x-a)|}{|x-a|} = 0.$$  \hspace{1cm} (1)

We will only apply the concept "$f$ is differentiable at $a$" to the case of a function defined in a neighborhood of the point $a$. That is, in this we do not consider differentiability at boundary points of the domain of a function.

Some remarks may be in order, some of them trivial. In the process we introduce, or explain, some notation.

1. If $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we define $|x| = (\sum_{i=1}^{n} x_i^2)^{1/2}$. If $f : U \to \mathbb{R}^m$, $U$ any set, then $f(x) \in \mathbb{R}^m$ for each $x \in U$ (obviously), so $f(x)$ has $m$ components. We then define $f_1(x), \ldots, f_m(x)$ by

$$f(x) = (f_1(x), \ldots, f_m(x)).$$

This also defines $f_i$ as a function from $U$ to $\mathbb{R}$ for each $i = 1, \ldots, m$. We write $f = (f_1, \ldots, f_m)$ if $f(x) = (f_1(x), \ldots, f_m(x))$ for all $x$ in the domain of $f$.

2. Every linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ can be identified with an $m \times n$ matrix $(a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$; its matrix with respect to the canonical bases. We can then write (1) in the much more complicated and messy way:

$$\lim_{x \to a} \frac{\sqrt{\sum_{i=1}^{m} (f_i(x_1, \ldots, x_n) - f_i(a_1, \ldots, a_n) - \sum_{j=1}^{n} a_{ij} x_j)^2}}{\sqrt{\sum_{j=1}^{n} (x_j - a_j)^2}} = 0.$$

3. An equivalent $\epsilon, \delta$ definition of differentiability goes as follows: Suppose $U$ is an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^m$. We say that $f$ is differentiable at the point $a \in U$ iff there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that for every positive real number $\epsilon$ there exists a positive real number $\delta$ with the property that

$$|f(x) - f(a) - A(x-a)| \leq \epsilon|x-a|$$  \hspace{1cm} (2)

whenever $x \in U$ and $|x-a| < \delta$. Notice that in (2) one cannot replace $\leq$ by $<$. This is because we ask the inequality to hold for $|x-a| < \delta$; in particular for $x = a$. For $x = a$, (2) reduces to $0 \leq 0$.

4. If $f : U \to \mathbb{R}^m$ is differentiable at $a \in U$, then the linear map $A$ is uniquely determined. This is easy to prove. For example if $A, B$ are linear maps such that for every $\epsilon > 0$ there exist $\delta, \delta' > 0$ such that $|f(x) - f(a) - A(x-a)| \leq \epsilon|x-a|$ if $x \in U$, $|x-a| < \delta$ and $|f(x) - f(a) - B(x-a)| \leq \epsilon|x-a|$ if $x \in U$ and $|x-a| < \delta'$. Because $U$ is open, there is also $\delta_0 > 0$ such that $|x-a| < \delta_0$ implies $x \in U$. Then letting $\delta'' = \min(\delta, \delta', \delta_0)$ we have that $\delta'' > 0$ and if we define $C = B - A$, then

$$|C(x-a)| = |(f(x) - f(a) - A(x-a)) - (f(x) - f(a) - B(x-a))| \leq |f(x) - f(a) - A(x-a)| + |f(x) - f(a) - B(x-a)| \leq 2\epsilon|x-a|$$

whenever $|x-a| < \delta''$. Now let $u \in \mathbb{R}^n$, $|u| = 1$. Then if we set $x = a + (\delta''/2)u$ we see that $x \in U$ and $|x-a| = \delta''/2 < \delta''$, thus (by linearity of $C$)

$$|C(u)| = \left| C \left( \frac{2}{\delta''}(x-a) \right) \right| = \frac{2}{\delta''} |C(x-a)| \leq \frac{2\epsilon}{\delta''} |x-a| = \epsilon |u| = \epsilon$$
Now $u$ was an arbitrary vector of length 1, $\epsilon > 0$ an arbitrary positive number $\delta''$, which depended on $\epsilon$ disappears in the end; we proved $|Cu| < \epsilon$ for all $u \in \mathbb{R}^n$ with $|u| = 1$. Thus $Cu = 0$ for all $u \in \mathbb{R}^n$ of norm 1, thus (by linearity) for all $u \in \mathbb{R}^n$. Thus $C$ is the 0 map; i.e., $A = B$.

**Definition.** If $U$ is open in $\mathbb{R}^n$, $f : U \to \mathbb{R}^m$, $a \in U$, and $f$ is differentiable at $a$, then the unique linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ is called the **differential of $f$ at $a$** and denoted by $df(a)$.

5. Differentiability is a local property, in the following sense. Let $U, V$ be open in $\mathbb{R}^n$, $a \in U \cap V$, let $f : U \to \mathbb{R}^m$, $g : V \to \mathbb{R}^m$ and assume $f(x) = g(x)$ for all $x \in U \cap V$. Then $f$ is differentiable at $a$ if and only $g$ is differentiable at $a$, and in this case $df(a) = dg(a)$. A particular case of this result is if $V$ is an open ball centered at $a$ contained in $U$. The proof of this is immediate.

6. If $f$ is differentiable at $a$, then $f$ is continuous at $a$. This is essentially because linear maps between finite dimensional vector spaces are continuous. Let $X, Y$ be finite dimensional vector spaces. We will denote by $\mathcal{L}(X, Y)$ the set of all linear maps from $X$ to $Y$. We write $\mathcal{L}(X)$ instead of $\mathcal{L}(X, X)$ if $X = Y$. For the sake of keeping things within bounds, I’ll stay within the usual vector spaces $\mathbb{R}^n$. Let $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Then we define the norm of $A$ by

$$
||A|| = \sup\{|Ax| : x \in \mathbb{R}^n, |x| \leq 1\}.
$$

To see that $||A||$ is finite, let $(a_{ij})$ be the matrix of $A$ with respect to canonical bases. Assume $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then, by Cauchy-Schwarz,

$$
|Ax|^2 = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij} x_j \right)^2 \leq \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij}^2 \right) \left( \sum_{j=1}^{n} x_j^2 \right) = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij}^2 \right) |x|^2
$$

This proves that

$$
|Ax| \leq \left( \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2} < \infty
$$

whenever $|x| \leq 1$. Trying to compute the actual norm of a linear map (except if its matrix is diagonal) can be quite difficult, except if one knows the trick. Luckily, we’ll rarely if ever have to compute the actual norm. However, for the sake of illustration, I’ll do an example below. But first, here are some totally equivalent ways of defining the norm of a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$

(a) $||A|| = \sup\{|Ax| : x \in \mathbb{R}^n, |x| = 1\}$.

(b) $||A|| = \sup\{\frac{|Ax|}{|x|} : x \in \mathbb{R}^n, x \neq 0\}$.

(c) It is obvious that $|Ax| \leq ||A|| |x|$ for all $x \in \mathbb{R}^n$; one can also define

$$
||A|| = \inf\{|M \in \mathbb{R} : |Ax| \leq M|x| \text{ for all } x \in \mathbb{R}^n\}.
$$

An example: Suppose $A : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$
A(x_1, x_2) = (x_1 - 2x_2, x_1 + 4x_2).
$$

Then

$$
||A||^2 = \sup\{(x_1 - 2x_2)^2 + (x_1 + 4x_2)^2 : x_1^2 + x_2^2 = 1\}.
$$

To compute this by brute force we can use $x_1^2 + x_2^2 = 1$ to solve for $x_2$, so $x_2 = \pm \sqrt{1 - x_1^2}$; then the quantity to maximize is

$$
(x_1 - 2x_2)^2 + (x_1 + 4x_2)^2 = 2x_1^2 + 17x_2^2 + 4x_1x_2 = 2x_1^2 + 17(1 - x_1^2) \pm 4x_1\sqrt{1 - x_1^2} = 17 - 15x_1^2 \pm 4x_1\sqrt{1 - x_1^2}.
$$


We can restrict ourselves to the positive square root if we keep in mind that we get the negative square root case by changing the sign of \( x_1 \). So we now try to maximize \( f(x_1) = 17 - 15x_1^2 + 4x_1 \sqrt{1 - x_1^2} \). Notice that because \( |x| = 1 \), we must have \( |x_1| \leq 1 \). Using our calculus 1 skills, we differentiate \( f \) and set the derivative to 0, getting the lovely equation

\[-30x_1 + 4\sqrt{1 - x_1^2} \cdot \frac{4x_1}{\sqrt{1 - x_1^2}} = 0.\]

Working on it, it reduces to the following quadratic equation for \( x_1^2 \):

\[241x_1^4 - 209x_1^2 + 4 = 0.\]

And so it goes. I’m not even sure these computations are right; I didn’t check them. Of course, using brute force for problems such as this one is silly, even for undergraduates. There is a nifty method that’s usually skipped in FAU calculus courses, called Lagrange multipliers that can be used to solve these problems. Using this one gets the following

Let \( A : \mathbb{R}^n \rightarrow \mathbb{R}^m \) be linear, let \( A^T : \mathbb{R}^n \rightarrow \mathbb{R}^m \) be the transpose of \( A \). In terms of matrices, the linear map whose matrix with respect to standard bases is the transpose of the matrix of \( A \). Actually, just assume \( A \) is an \( m \times n \) matrix. Then

\[\|A\| = \max \{ \lambda : \lambda \text{ is an eigenvalue of } A^T A. \}\]

For the example in question, in matrix notation,

\[A = \begin{pmatrix} 1 & -2 \\ 1 & 4 \end{pmatrix}, \quad A^T A = \begin{pmatrix} 5 & -7 \\ -7 & 17 \end{pmatrix}\]

and one sees that the eigenvalues of \( A^T A \) are given by

\[\lambda = 11 \pm \sqrt{85},\]

thus \( \|A\| = 11 + \sqrt{85} \).

But let us return to the beginning of this point, if \( f \) is differentiable at \( a \), then it is continuous at \( a \). Let \( A = df(a) \). Then given \( \epsilon > 0 \), letting \( \delta > 0 \) be related to \( \epsilon \) as in (2), \( |x - a| < \delta \) implies

\[|f(x) - f(a)| \leq |f(x) - f(a) - A(x - a)| + |A(x - a)| \leq (\epsilon + \|A\|)|x - a|\]

and it is clear that \( \lim_{x \to a} f(x) = f(a) \).

7. Two first examples; more later.

(a) Suppose \( A : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is linear, and define \( f(x) = Ax \) (we usually write \( Ax \) rather than \( A(x) \) if \( A \) is a linear map). If \( a \in \mathbb{R}^n \) we obviously have

\[\left|\frac{f(x) - f(a) - A(x - a)}{|x - a|}\right| = \left|\frac{Ax - Aa - A(x - a)}{|x - a|}\right| = 0\]

since \( Ax - a = A(x - a) \). Thus \( f \) is differentiable at each \( a \in \mathbb{R}^n \) and \( df(a) = A \) for all \( a \in \mathbb{R}^n \).

(b) Let \( I \) be an open interval in \( \mathbb{R} \) and let \( f : I \rightarrow \mathbb{R} \). Let \( a \in I \). A linear map \( A : \mathbb{R} \rightarrow \mathbb{R} \) is given by multiplication by a constant; that is \( A : \mathbb{R} \rightarrow \mathbb{R} \) is linear if and only if there exists \( a \in \mathbb{R} \) such that \( Ax = ax \) for all \( x \in \mathbb{R} \). In view of this it is very easy to prove: \( f \) is differentiable at \( a \) in the sense of our definition above if and only if it is differentiable in the calculus 1 or introductory analysis 1 sense and in this case \( df(a) \) is the linear map given by \( df(a) u = f'(a) u \) for all \( u \in \mathbb{R} \).

8. If \( A : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is a linear map we will define the linear maps \( A_i : \mathbb{R}^m \rightarrow \mathbb{R} \) by \( A_i x \) is the \( i \)-th component of \( Ax \). In matrix notation, if the matrix of \( A \) with respect to the standard bases is \((a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \), then

\[A_i(x_1, \ldots, x_n) = \sum_{j=1}^{n} a_{ij} x_j.\]

Suppose \( U \subset \mathbb{R}^n \) is open, \( f = (f_1, \ldots, f_m) : U \rightarrow \mathbb{R}^m \), \( a \in U \). Then \( f \) is differentiable at \( a \) if and only if each \( f : U \rightarrow \mathbb{R} \) is differentiable, in which case \( df_i(a) = (df)_i(a) \). That is, if \( df(a) = A \), then \( df_i(a) = A_i \) for \( i = 1, \ldots, m \).
9. We say \( f \) is differentiable on the open subset \( U \) of \( \mathbb{R}^n \) if \( f \) is differentiable at all points of \( U \). In this case, if \( f : U \to \mathbb{R}^m \), then \( df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \).

10. Let \( U \) be open in \( \mathbb{R}^n \) let \( a \in U \), let \( f, g : U \to \mathbb{R}^m \) be differentiable at \( a \). Then \( \alpha f + \beta g : U \to \mathbb{R}^m \) is differentiable for any choice of \( \alpha, \beta \in \mathbb{R} \); moreover \( d(\alpha f + \beta g)(a) = \alpha df(a) + \beta dg(a) \).

11. (The chain rule) Assume \( f : U \to \mathbb{R}^m \) is differentiable at \( a \in U \), where \( U \) is open in \( \mathbb{R}^n \). Assume \( f(U) \subset V, V \) open in \( \mathbb{R}^m \), let \( g : V \to \mathbb{R}^p \) be differentiable at \( b = f(a) \). Then \( g \circ f : U \to \mathbb{R}^p \) is differentiable at \( a \) and \( d(g \circ f)(a) = dg(b) \circ df(a) \).

In particular, the Jacobian matrix of \( g \) at \( a \) is the product of the Jacobian matrix of \( f \) at \( a \) times the Jacobian matrix of \( g \) at \( b = f(a) \); the matrix of \( f \) on the right, that of \( g \) on the left.

12. We will write \( \{e_1, \ldots, e_n\} \) to denote the canonical base of \( \mathbb{R}^n \). Because it should be clear from the context what the dimension \( n \) of the space is, we won’t overburden the notation to emphasize what \( n \) is. So if \( n = 2 \), then \( e_1 = (1, 0) \); if \( n = 4 \) then \( e_1 = (1, 0, 0, 0) \). In general, \( e_j \) is a vector in \( \mathbb{R}^n \) whose \( j \)-th component is 1, all others are 0. In terms of the so called Kronecker deltas one can define \( \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \)

Let \( f : U \to \mathbb{R}^m \), where \( U \) is open in \( \mathbb{R}^n \). Let \( a \in U \). We say that \( f \) has a partial derivative at \( a \) with respect to the \( j \)-th variable if

\[
\lim_{h \to 0} \frac{1}{h} (f(a + he_j) - f(a))
\]

exists. This limit is, of course a vector in \( \mathbb{R}^m \) and denoted by \( \frac{\partial f}{\partial x_j}(a) \). If \( n = 1 \), it is just the usual calculus 1 derivative \( f'(a) \) of \( f \) at \( a \). It is the derivative of \( f \) with respect to the \( j \)-th variable when all other variables are held fixed. It also obvious that if \( f = (f_1, \ldots, f_m) \), then \( F \) has a partial derivative with respect to the \( j \)-th variable at \( a \) if and only if each \( f_i \) has a partial derivative with respect to the \( j \)-th variable at \( a \), and in this case

\[
\frac{\partial f}{\partial x_j}(a) = \left( \frac{\partial f_1}{\partial x_j}(a), \ldots, \frac{\partial f_m}{\partial x_j}(a) \right).
\]

13. Assume \( f = (f_1, \ldots, f_m) : U \to \mathbb{R}^m \) is differentiable at \( a \in U \), where \( U \) is open in \( \mathbb{R}^n \). Then \( f \) has partial derivatives with respect to all variables at \( a \). Moreover, the matrix of \( df(a) \) with respect to the canonical bases of \( \mathbb{R}^n, \mathbb{R}^m \) is given by

\[
\begin{pmatrix}
\frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\
\frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\
& & \ddots & \\
\frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a)
\end{pmatrix}
\]

This simplifies verifying differentiability. For example, suppose we want to prove that \( f : \mathbb{R}^3 \to \mathbb{R} \) given by \( f(x_1, x_2, x_3) = x_1^2 - 3x_1x_2 + 4x_3^2 \) is differentiable at \( a = (1, 2, 1) \). Thanks to the result of this point, we know what the differential should be; it is the linear map having the \( 1 \times 3 \) matrix

\[
\begin{pmatrix}
\frac{\partial f}{\partial x_1}(1, 2, 1) & \frac{\partial f}{\partial x_2}(1, 2, 1) & \frac{\partial f}{\partial x_3}(1, 2, 1)
\end{pmatrix} = (-4, -3, 12);
\]

that is, \( df(a)u = -4u_1 - 3u_2 + 12u_3 \) if \( u = (u_1, u_2, u_3) \). Either this map works, or nothing does. So we must investigate

\[
f(x) - f(a) - df(a)(x - a) = x_1^2 - 3x_1x_2 + 4x_3^2 - (a) - [-4(x_1 - 1) - 3(x_2 - 2) + 12(x_3 - 1)]
\]

\[
= x_1^2 - 3x_1x_2 + 4x_3^2 - 4x_1 - 3x_2 + 12x_3 + 3.
\]

Writing \( x_1 = (x_1 - 1) + 1, x_2 = (x_2 - 2) + 2, x_3 = (x_3 - 1) + 1, \) expanding and simplifying one gets

\[
f(x) - f(a) - df(a)(x - a) = (x_1 - 1)^2 - 3(x_1 - 1)(x_2 - 2) + 4(x_3 - 1)^3 + 12(x_3 - 1)^2.
\]
Estimating $|x_i - a_i| \leq |x - a|$ one sees that

$$|f(x) - f(a) - df(a)(x - a)| \leq (16|x - a| + 4|x - a|^2)|x - a|$$

so that clearly

$$\lim_{x \to a} \frac{|f(x) - f(a) - df(a)(x - a)|}{|x - a|} = 0.$$ 

14. It would be nice if existence of the partial derivatives at a point implied differentiability; unfortunately that is not so. A typical example is the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

Then

$$\lim_{h \to 0} \frac{f((0, 0) + he_1) - f(0, 0)}{h} = f(h, 0) - f(0, 0) = \frac{h^2}{h} = h$$

thus

$$\lim_{h \to 0} \frac{f((0, 0) + he_1) - f(0, 0)}{h} = 0$$

and $\partial f/\partial x(0, 0)$ exists and is 0. Similarly, $\partial f/\partial y(0, 0) = 0$. If the function is differentiable at $(0, 0)$, its differential at $(0, 0)$ must be the linear map whose matrix is $(0 0)$; i.e., the zero linear map. Thus $f$ is differentiable at $(0, 0)$ if and only if

$$0 = \lim_{(x, y) \to (0, 0)} \frac{|f(x, y) - f(0, 0)|}{\sqrt{x^2 + y^2}} = \lim_{(x, y) \to (0, 0)} \frac{|xy|}{(x^2 + y^2)^{3/2}}.$$ 

But this is false. For example, if $x = y$

$$\frac{|xy|}{(x^2 + y^2)^{3/2}} = \frac{x^2}{2|x|^3} = \frac{1}{|x|},$$

which certainly doesn’t go to 0 as $x \to 0$. It turns out that the function in question is not even continuous at $(0, 0)$.

15. The existence of partials does not imply differentiability. But actually, it almost does. So far everything stated has an immediate proof; in fact, a lot of textbooks might not even state it thinking its just too obvious to mention. The following result is slightly harder to prove.

**Theorem 1** Let $U$ be an open subset of $\mathbb{R}^n$, let $f : U \to \mathbb{R}^m$, let $a \in U$. If $df_x f_j(x)$ exist for $j = 1, \ldots, n$ for all $x$ in a neighborhood of $a$ and are continuous at $a$, then $f$ is differentiable at $a$.

16. Assume again that $f = (f_1, \ldots, f_m) : U \to \mathbb{R}^m$ is differentiable at $a \in U$, where $U$ is open in $\mathbb{R}^n$. The matrix of $df(a)$ with respect to the canonical bases of $\mathbb{R}^n, \mathbb{R}^m$, that is, the matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix},$$

is called the Jacobian matrix of $df$ at $a$. If $n = m$, then the determinant of the Jacobian matrix is called the Jacobian determinant or simply the Jacobian of $f$ at $a$.

17. **Definition.** Let $U$ be an open subset of $\mathbb{R}^n$, let $f : U \to \mathbb{R}^m$. We say $f$ is continuously differentiable on $U$, and write $f \in C^1(U)$, iff $\frac{\partial f}{\partial x_j}(x)$ exist for $j = 1, \ldots, n$ for all $x \in U$ and are continuous on $U$. This means that $f$ is differentiable at all points of $U$. One can provide $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with a metric defining $d(A, B) = \|A - B\|$ for all $A, B \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. One can then see (quite easily) that $f \in C^1(U)$ if and only if $f$ is differentiable at all points of $U$ and the map $df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is continuous. We won’t place too much emphasis on this last way of defining continuous differentiability. However . . .
Identifying each linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ with its matrix $(a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ with respect to canonical bases identifies $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with $\mathbb{R}^{nm}$. If the matrix of $A$ is $(a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$, we can also define the Euclidean norm of $A$ by $|A| = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}$. It is quite easy to prove that then

$$\|A\| \leq |A| \leq \sqrt{n} \|A\|.$$  

In fact, the first inequality is the same as (3). For the second one, notice that $Ae_j = (a_{1j}, \ldots, a_{mj})$, thus

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |Ae_j|^2 \leq \sum_{j=1}^{n} \|A\|^2 |e_j|^2 = \sum_{j=1}^{n} \|A\|^2 = n \|A\|^2.$$  

This shows that notion such as continuity, uniform continuity, are the same for both the first norm $\| \cdot \|$, as well as for the Euclidean norm. This is just a particular case of a theorem that states that in a finite dimensional vector space all norms are equivalent. So, to return to the earlier discussion, if $U$ is open in $\mathbb{R}^n$ and $f : U \to \mathbb{R}^m$ is differentiable, it makes perfect good sense to ask whether $df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is differentiable; if it is we have that

$$d(df) : U \to \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m)) \cong \mathcal{L}(\mathbb{R}^n, \mathbb{R}^{nm}) \cong \mathbb{R}^{nm}.$$  

But we’ll try to avoid this escalation of dimensions. because of the theorem, continuous differentiability can be defined in terms of partial derivatives, and ours is going to be a very continuous, smooth course. So we stick to partials, were possible.

18. As said above, we stick to partial derivatives. And since a function into $\mathbb{R}^m$ is just an $m$-tuple of functions into $\mathbb{R}$, we may assume mostly that $m = 1$ with the proviso that if we defined a concept for real valued functions and then apply the concept without redefining to a vector valued function, it is understood that each component of the function satisfies the definition.

Assume that $f : U \to \mathbb{R}$, $U$ open in $\mathbb{R}^n$, and assume that all partial derivatives of $f$ exist at all points of $U$. These partials are then functions from $U$ to $\mathbb{R}$ and it makes sense to ask whether they have partial derivatives. The order of derivation makes a difference! Thus, in general, $\frac{\partial}{\partial x_j} \left( \frac{\partial f}{\partial x_i} \right) \neq \frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_j} \right)$ if $k \neq j$. Luckily for us, however, they are equal in all cases we have to consider in this course. The result to remember (one can get the same consequence with slightly fewer hypotheses) is if at some point of $U$ all the second partials of $f$ exists and are continuous, then the order in which they were obtained is immaterial.

We thus will say, if $k \in \mathbb{N}$, that $f : U \to \mathbb{R}$ is $k$-times continuously differentiable, and write $f \in C^k(U)$ if and only if all partial derivatives of $f$ up to order of derivation $k$ exist and are continuous on $U$. We also define $C^\infty(U) = \bigcap_{k \in \mathbb{N}} C^k(U)$ and $C^0(U) = C(U)$ equal to the set of all continuous real valued functions on $U$.

In this context a certain notation is quite useful. A multi-index is an $n$-tuple of non-negative integers; what $n$ is, is usually clear from the context. A typical notation for a multi-index is $\alpha = (\alpha_1, \ldots, \alpha_n)$. If $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index, we define the length of $\alpha$ by $|\alpha| = \alpha_1 + \cdots + \alpha_n$. It is the same notation used for the Euclidean norm; once again, it should be clear from the context which is which. Suppose now $f : U \to \mathbb{R}$, where $U$ is open in $\mathbb{R}^n$. Suppose $\alpha_1, \ldots, \alpha_n$ are non-negative integers and we want to refer to the function obtained by differentiating $f$ $\alpha_1$ times with respect to the first variable, $\alpha_2$ times with respect to the second variable, $\ldots$, $\alpha_n$ times with respect to the $n$-th variable. If $f$ is sufficiently many times continuously differentiable (a standing assumption in this course) then, mercifully, the order in which we perform these derivations doesn’t matter, so we can group corresponding derivatives together and denote the end product by

$$\frac{\partial^{\alpha_1 + \cdots + \alpha_n} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$  

Instead, we introduce the multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ and define

$$D^\alpha f = \frac{\partial^{\alpha_1 + \cdots + \alpha_n} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} = \frac{\partial^{\alpha_1} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$  

6
If any of the $\alpha_i$’s equals 0, we interpret $\partial^0/\partial x_i^0$ as doing nothing. Thus, for example, if $n = 3$ and $\alpha = (3,0,1)$, then

$$D^{\alpha}f = \frac{\partial^4 f}{\partial x_1^3 \partial x_3}. $$

In particular, if $\alpha = 0$, the $n$-tuple with all entries equal to 0, we have (or define) $D^0 f = f$.

### 1.2 Two Fundamental Theorems

All the results of the previous subsection were fairly simple. Things get a bit more serious here. I will state (versions of) the inverse function theorem and of the implicit function theorem and try to illustrate a bit their use. For proofs, I refer to any good standard textbook of introductory analysis, for example, Rudin’s *Principles of Mathematical Analysis*.

In some ways it all comes down to solving equations. In linear algebra we learn how to solve a system of equations of the form

\[
\begin{align*}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= y_m
\end{align*}
\]

This system is easier to handle when written out in matrix notation. Let $A = (a_{ij})$ be the $m \times n$ matrix of the system, so we can write the system in the form $Ax = y$, where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$.

By solving it we mean expressing $x_1, \ldots, x_n$ in terms of $y_1, \ldots, y_m$. We learn that very roughly one needs one equation for each variable one tries to solve so that in very, very general terms, the ideal situation is $m = n$. If $m < n$ one will be able, again very roughly, at best to solve for $m$ of the unknowns. If $m > n$ one might have too many equations and there might be a contradiction among the equations. This is, to repeat, very rough. A very precise though at first glance trivial result is: if $m = n$ and the matrix $A$ is invertible, then the system has a unique solution $x = A^{-1} y$. What makes this result less trivial than it seems, is that there are relatively simple ways of determining (pun intended?) that a matrix is invertible; namely by computing its determinant. In fact, we have the following well known linear algebra theorem: *If $m = n$, then the following conditions are equivalent:*

1. $\det A \neq 0$
2. The system $Ax = y$ has a solution for each choice of $y \in \mathbb{R}^n$
3. The only solution of $Ax = 0$ is $x = 0$.
4. Solutions of the system are unique.

The inverse function and the implicit function theorems deal with the more complicated question of solving equations of the form $f(x) = y$ where $f : U \rightarrow \mathbb{R}^m$, $U \subset \mathbb{R}^n$. Since $f = (f_1, \ldots, f_m)$, we have once more $m$ equations and $m$ unknowns. Once again, very roughly, one needs one equation for each unknown we want to solve, so the ideal situation is again $n = m$. If $f$ is differentiable at some point $a$, then the affine map $x \mapsto f(a) + df(a)(x - a)$ is “close” to the map $f$ and it is not unreasonable to assume that the equation $f(x) = y$ and $f(a) + df(a)(x - a) = y$ have similar behaviors. But one has to make clear what is meant by “close.” Anyway, to make a long story less long, here is the inverse function theorem.

**Theorem 2** *(INVFT)* Let $f : U \rightarrow \mathbb{R}^n$ be $k$-times continuously differentiable, where $k \in \mathbb{N} \cup \{\infty\}$ and $U$ is open in $\mathbb{R}^n$. Let $a \in U$ and assume that the differential $df(a)$ is invertible (as a linear map $\mathbb{R}^n \rightarrow \mathbb{R}^n$). There exist then open sets $V \subset U$, $W \subset \mathbb{R}^n$ such that $a \in V$, $f(a) \in W$ and $f|_V$ (the restriction of $f$ to $V$) maps $V$ onto $W$ in a one-to-one fashion. Moreover, $(f|_V)^{-1} : W \rightarrow V$ (the inverse of the restriction of $f$ to $V$) is also $k$-times continuously differentiable.

Because a linear map is invertible if and only if its determinant* is non-zero, the main hypothesis of INVFT is frequently stated as: $J_f(a) \neq 0$, where we denote by $J_f(x)$ the Jacobian determinant of $f$ at $x$; $J_f$ is defined wherever $f$ is differentiable.

*If $X$ is a finite dimensional vector space and $A : X \rightarrow X$ is a linear map, let $\{v_1, \ldots, v_n\}$ be a basis of $X$ and let $M_A$ be the matrix of $A$ with respect to that basis; specifically $M_A = (c_{ij})$ where $Av_j = \sum_{i=1}^n c_{ij}v_i$. Then one defines det $A = \det M_A$. This definition makes sense because all matrix representations of $A$ have the same determinant.*
by
\[ J_f = \det \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}, \]

A definition that makes the statement a bit shorter is the following:

Let \( U, V \) be open in \( \mathbb{R}^n \). A map \( f : U \to V \) is a diffeomorphism if \( f \in C^1(U) \), \( f \) is one-to-one on \( U \) and onto \( V \) (that is \( f(U) = V \)), and the inverse \( f^{-1} : V \to U \) is in \( C^1(V) \). If \( A \subset \mathbb{R}^n \), \( f : A \to \mathbb{R}^n \), \( a \in A \), we say that \( f \) is a local diffeomorphism at \( a \) if there exist open subsets \( U, V \) of \( \mathbb{R}^n \) such that \( a \in U \subset A \) and \( f : U \to V \) is a diffeomorphism. With this we can state INVFT in the following shorter version:

**Theorem 3** Let \( f : U \to \mathbb{R}^n \) be \( k \)-times continuously differentiable, where \( k \in \mathbb{N} \cup \{ \infty \} \) and \( U \) is open in \( \mathbb{R}^n \). Let \( a \in U \) and assume that \( J_f(a) \neq 0 \). Then \( f \) is a local diffeomorphism at \( a \) and the local inverse is also \( k \)-times continuously differentiable.

Here are a few examples and consequences.

1. Let us begin with the simplest situation, \( n = 1 \). If we have a function \( f : I \to \mathbb{R} \), where \( I \) is an open interval in \( \mathbb{R} \) and \( f \) is continuously differentiable, if \( a \in I \) and \( J_f(a) = f'(a) \neq 0 \), then either \( f'(a) > 0 \) or \( f'(a) < 0 \). Say \( f'(a) > 0 \); the case \( f'(a) < 0 \) is quite similar. By continuity of \( f' \), \( f'(x) > 0 \) in some interval \( J = (a - \epsilon, a + \epsilon) \subset [a - \epsilon, a + \epsilon] \subset I \), thus \( f \) is strictly increasing in \( J \) and if we set \( c = f(a - \epsilon), d = f(a + \epsilon) \), then \( f(J) = (c, d) \). Thus \( f \) maps \( J \) onto \( (c, d) \) and it is a standard introductory analysis result that \( f^{-1} : (c, d) \to J \) is differentiable and, in fact,

\[
(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}.
\]

2. Consider the map \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) defined by \( f(x, y) = (e^x \cos y, e^x \sin y) \). This map is clearly in \( C^\infty(\mathbb{R}^2) \). The Jacobian matrix is

\[
\begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix},
\]

the Jacobian determinant is \( J_f(x) = e^{2x} \neq 0 \) for all \((x, y) \in \mathbb{R}^2\). Concentrating at any point of \( \mathbb{R}^2 \), say at \((0, 0)\), we have a local diffeomorphism. There exist open sets \( U, V \) in \( \mathbb{R}^2 \), \((0, 0) \in U\), \((1, 0) = f(0, 0) \in V \) such that \( f(U) = V \), \( f \) is one-to-one on \( V \), and \((f|_V)^{-1} : V \to U \) is \( C^\infty \). The inverse function theorem gives no indication on how large (or small) \( U \) or \( V \) are. We’ll return to this example shortly.

3. Let \( U \) be open in \( \mathbb{R}^n \) and let \( f : U \to \mathbb{R}^n \) be continuously differentiable. Assume \( J_f(x) \neq 0 \) for all \( x \in U \). Then \( f \) is an open map; i.e., if \( V \) is an open subset of \( U \), then \( f(V) \) is an open subset of \( \mathbb{R}^n \). This is, I hope, obvious from INVFT.

4. Let \( U \) be open in \( \mathbb{R}^n \) and let \( f : U \to \mathbb{R}^n \) be continuously differentiable and one-to-one. Assume \( J_f(x) \neq 0 \) for all \( x \in U \). Then \( f \) is a diffeomorphism of \( U \) onto the open set \( f(U) \). The inverse is as many times differentiable as \( f \) is.

5. Returning now to our \( \mathbb{R}^2 \) example, \( f(x, y) = (e^x \cos y, e^x \sin y) \), we can ask ourselves if \( f \) is one-to-one. One sees that because of the periodicity of \( \sin \), \( \cos \), \( f \) is NOT one-to-one, but is one-to-one on horizontal strips of width \( \leq 2\pi \). For example, \( f \) is one-to-one on the open strip \( U = \{(x, y) : |y| < \pi\} \). It is a relatively simple exercise to figure out what \( f(U) \) is and to get an expression for the inverse. But all becomes very simple if one allows complex numbers and writes \( z = x + iy \). Then \( f(z) = e^z, f(U) = C \setminus \{x \in \mathbb{R} : x \leq 0\} \) (as one says in complex analysis, the plane cut along the negative real axis), and the inverse of \( f|_U \) is the principal value logarithm function.

I won’t prove INVFT here. The proof is a bit involved, but not horrifically hard. The main case is the one in which \( k = 1 \); the function \( f \) is only assumed to be continuously differentiable. This is because once it is established that the inverse \( f^{-1} \) is differentiable, its jacobian matrix must be the inverse of the Jacobian matrix of \( f \); using the adjoint form of the inverse of a matrix one can express all entries of the Jacobian matrix of \( f^{-1} \); that is, all partial derivatives of \( f \), as algebraic functions of the partial derivatives of \( f \), hence the partial derivatives of \( f^{-1} \) are (at least) as many times differentiable as those of \( f \).
Let us turn now to the implicit function theorem. This theorem gives conditions under which one can use an equation of the form $f(x) = C$, where $f : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$, $C$ is a constant, and $n > m$, to solve for $m$ of the components of $x$, and to express these as functions of the remaining $n - m$ components. To get into the mood, I want to consider the simplest situation, namely $m = 1, n = 2$. Consider a concrete function, namely $f(x, y) = x^2 + y^2$. We already get some insight into the general situation by seeing what can happen with the equation

$$x^2 + y^2 = C. \tag{5}$$

Let us say we want to use this equation to get $y$ as a function of $x$. An obvious first observation is that if $C < 0$, there is no solution. If $C = 0$ there is a solution if and only if $x = 0$, and then $y = 0$; we can hardly say that we express $y$ in terms of $x$. If $C > 0$, then we can solve as long as $|x| \leq \sqrt{C}$; we get two solutions $y = \pm \sqrt{C - x^2}$, valid for $x \in (-\sqrt{C}, \sqrt{C})$. While this could be called the simplest non-trivial example (where we consider the case in which $f$ is a linear map the trivial case), it already has some of the main features of what happens in general. Let us consider the problem of solving (5) geometrically. The graph of the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2 + y^2$ is a surface in $\mathbb{R}^3$; $G_f = \{(x, y, z) \in \mathbb{R}^3 : z = x^2 + y^2\}$ (called a paraboloid of revolution, but it doesn’t really matter what it is called). The equation (5) represents the points in the intersection of the graph $G_f$ with the plane $z = C$. In general, the graph $G_f$ of a function does not have to have a non-empty intersection with a horizontal plane $z = C$. For our function, this is the case if $C < 0$. If it has an intersection, that intersection could be an isolated point, as happens for our example when $C = 0$. If the intersection is more than isolated points, it still might not be quite possible to express the intersection points in the form $y = \text{function of } x$. Or there may be (as is the case for our example) more than one way of doing this.

We must be more modest. Staying for a while longer in the simplest case, $n = 2, m = 1$, let $f : U \to \mathbb{R}$ be continuously differentiable, $U$ open in $\mathbb{R}^2$, and consider the equation

$$f(x, y) = 0. \tag{6}$$

(For the theory part, one can always assume $C = 0$; if not replace $f$ by $f - C$. That amounts to writing (5) in the form $x^2 + y^2 - C = 0$.)

We are now looking at the intersection of the graph of $f$ with the plane $z = 0$. The first thing we do at this more modest level is to assume there is an intersection, so we only consider the case in which there is $(x_0, y_0)$ such that $f(x_0, y_0) = 0$. And now we modestly inquire: Is there an interval of positive length centered at $x_0$ so that for every $x$ in that interval we can get a unique solution $y$ of the equation (6)? Or, in a more precise way: Does there exist $\epsilon > 0$ and a unique function $\phi : (x_0 - \epsilon, x_0 + \epsilon) \to \mathbb{R}$ such that $\phi(x_0) = y_0$ and $f(x, \phi(x)) = 0$ for all $x \in (x_0 - \epsilon, x_0 + \epsilon)$? Becoming a bit bolder, we can also ask: Will $\phi$ be differentiable? As it turns out, we better add that as a requirement.

Let us see first how this more modest approach works in the case $f(x, y) = x^2 + y^2 - C$, where $C$ is a constant. By considering only cases in which there is a solution of $x^2 + y^2 - C = 0$, we are implicitly (no pun intended) assuming that $C \geq 0$. If $C = 0$, then $(0, 0)$ is a solution. But no $\epsilon > 0$ exists. Assume now $C > 0$, say $C = 1$. Then, for example, $(0, 1)$ solves the equation and we can take $\epsilon = 1$; the function $\phi : (-1, 1) \to \mathbb{R}$, defined by

$$\phi(x) = +\sqrt{1 - x^2},$$

is the one and only unique differentiable function in the whole wide world, defined on $(-1, 1)$, and satisfying both $\phi(0) = 1$ and

$$f(x, \phi(x)) = x^2 + \phi(x)^2 = 1.$$ 

If we drop the differentiability requirement, we can get an infinity of such functions; rather silly functions, but still functions. For example, we could define $\phi(x) = 1$ if $x = 0$, but $\phi(x) = -\sqrt{1 - x^2}$ if $x \neq 0$. Or we could define $\phi(x) = \sqrt{1 - x^2}$ if $x$ is rational, $\phi(x) = -\sqrt{1 - x^2}$ if $x$ is irrational. Or, we could define $\phi(x) = \sqrt{1 - x^2}$ if $|x| < 1/4$, $\phi(x) = -\sqrt{1 - x^2}$ if $1/4 \leq |x| < 1$. Or the possibilities are endless. Notice that none of the alternatives to the unique nice one are even continuous; as it turns out continuity is sufficient for uniqueness.

Still with $f(x, y) = x^2 + y^2 - 1$, consider the solution $(3/5, -4/5)$ of the equation $x^2 + y^2 - 1 = 0$. Here it is easy to see that we can take $\epsilon = 1 - 3/5 = 2/5$, and the function $\phi : (1/5, 1) \to \mathbb{R}$ defined by $\phi(x) = -\sqrt{1 - x^2}$ is the one and only differentiable function such that $\phi(3/5) = -4/5$ and $x^2 + \phi(x)^2 = 1$.

Suppose we now consider the solution $(1, 0)$ of $x^2 + y^2 - 1 = 0$. We see that once more no $\epsilon > 0$ exists. We cannot solve to get $y$ in terms of $x$ when $x > 1$. We know everything about this example because we can solve the equation explicitly. But is
there any way of differentiating between the cases in which we can find \( \epsilon > 0 \) and \( \phi \) without having to solve the equation? Of course there is, otherwise I wouldn’t have asked this question. Let us ask ourselves what can go wrong in a general situation. Suppose again \( f : U \to \mathbb{R} \) is continuously differentiable, \( f(x_0, y_0) = 0 \); what can happen to make it impossible to use the equation \( f(x, y) = 0 \) to get \( y = \phi(x) \) in a neighborhood of \( x_0 \). Case 1. \((x_0, y_0)\) is an isolated point of the intersection of the graph of \( f \) with the \((x, y)\)-plane. But that means that the value 0 is either a local maximum or a local minimum of \( f \); in either case \( \frac{\partial f}{\partial y}(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0) = 0 \). Case 2. The intersection is a large set, but it is still not possible to describe it in the form \((x, \phi(x))\). One problem could be that near \( x_0 \), every vertical line intersects the set of points \(\{(x, y) : f(x, y) = 0\}\) either not at all, or at least twice. In these cases one can argue that one will have \( \frac{\partial f}{\partial y}(x_0, y_0) \neq 0 \). What one sees is that in every case in which this \( \epsilon > 0 \) does not exist, \( \frac{\partial f}{\partial y}(x_0, y_0) = 0 \).

But let us be positive; suppose \( ddx f y(x_0, y_0) \neq 0 \), say \( \frac{\partial f}{\partial y}(x_0, y_0) > 0 \). By continuity, \( \frac{\partial f}{\partial y}(x, y) > 0 \) for \((x, y)\) near \((x_0, y_0)\). That means that if \( x \) is near \( x_0 \), then the map \( y \mapsto f(x, y) \) is strictly increasing near \( y_0 \). For the case \( x = x_0 \), because \( f(x_0, y_0) = 0 \), \( y = y_0 \) will be the only \( y \) solving \( f(x_0, y) = 0 \) and \( f(x_0, y) < 0 \) for \( y < y_0 \). \( f(x_0, y) > 0 \) for \( y > y_0 \). If \( x \) is very close to \( x_0 \), the map \( y \mapsto f(x, y) \) will also go from positive to negative and, being strictly increasing there will be a single solution \( y = \phi(x) \) of the equation \( f(x, y) = 0 \). Geometrically, what happens is that the surface of equation \( z = f(x, y) \) intersects the \((x, y)\)-plane in a nice curve describable in the form \( y = \phi(x) \).

What happens in more general situations, when the dimension is higher? The answer is that again the function tends to behave near a point like its differential. Looking at the Jacobian matrix, assuming \( m \leq n \), its rank can at most be \( m \). The implicit function theorem assumes the rank is \( m \). In this case there is an \( m \times m \) minor that is different from zero. At least one, there could be several. The rows of the Jacobian matrix correspond to the components of the function \( f \), the columns to the variables. Say that \( m \times m \) minor corresponding to the variables \( x_{j_1}, \ldots, x_{j_m} \) is not zero. Then these variables can be expressed locally and smoothly in terms of the remaining ones. That is the implicit function theorem. To be able to write out the precise version it may be convenient to introduce a bit more notation.

Assume \( f_1, f_2, \ldots, f_m : U \to \mathbb{R} \) are functions, \( U \subset \mathbb{R}^n \). Assume \( m \leq n \) and let \( 1 \leq j_1 < j_2 < \cdots < j_m \leq n \). We define the function \( \frac{\partial(f_1, \ldots, f_m)}{\partial(x_{j_1}, \ldots, x_{j_m})} \) on the set of points \( x \in U \) where all the partial derivatives in question are defined by

\[
\frac{\partial(f_1, \ldots, f_m)}{\partial(x_{j_1}, \ldots, x_{j_m})}(x) = \det \begin{pmatrix}
\frac{\partial f_1}{\partial x_{j_1}}(x) & \frac{\partial f_1}{\partial x_{j_2}}(x) & \cdots & \frac{\partial f_1}{\partial x_{j_m}}(x) \\
\frac{\partial f_2}{\partial x_{j_1}}(x) & \frac{\partial f_2}{\partial x_{j_2}}(x) & \cdots & \frac{\partial f_2}{\partial x_{j_m}}(x) \\
\cdots & \cdots & \cdots & \cdots \\
\frac{\partial f_m}{\partial x_{j_1}}(x) & \frac{\partial f_m}{\partial x_{j_2}}(x) & \cdots & \frac{\partial f_m}{\partial x_{j_m}}(x)
\end{pmatrix}
\]

As \((j_1, \ldots, j_m)\) varies over all possible \( \binom{n}{m} \) choices, we get all the minors of the Jacobian matrix of \( f = (f_1, \ldots, f_m) \). In particular,

\[
J_f = \frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_n)}
\]

if \( m = n \). The notation will still be messy, if one tries to be general. I will now state two versions of the implicit function theorem. A messy one, and then the standard one in which an assumption is made to avoid a lot of the mess. Here is one extra notational definition: If \( m < n \), if \( j_1, \ldots, j_m \) is a set of integers such that \( 1 \leq j_1 < j_2 < \cdots < j_m \leq n \), we will call \((k_1, \ldots, k_{n-m})\) the complement of \((j_1, \ldots, j_m)\) if \( 1 \leq k_1 < \cdots < k_{n-m} \leq n \) and \( \{j_1, \ldots, j_m\} \cup \{k_1, \ldots, k_{n-m}\} = \{1, \ldots, n\} \). Suppose now we have an \( m \)-tuple \((y_{j_1}, \ldots, y_{j_m})\) and an \((n-m)\)-tuple \(x_{k_1}, \ldots, x_{k_{n-m}}\).

For example, if \( n = 4 \), \( m = 1 \) and \( j_1 = 3 \), then the complement is \( k_1 = 1, k_2 = 2, k_3 = 4 \). If \( n = 5, m = 3 \) and \((j_1, j_2, j_3) = (2, 3, 5)\), then the complement is \( k_1 = 1, k_2 = 4 \). If

**Theorem 4 (IMPFT1)** Let \( U \) be open in \( \mathbb{R}^n \), assume \( f : U \to \mathbb{R}^m \) is continuously differentiable, \( m < n \). Let \( a \in U \), let \( j_1, \ldots, j_m \) be such that \( 1 \leq j_1 < \cdots < j_m \leq n \) and

\[
\frac{\partial(f_1, \ldots, f_m)}{\partial(x_{j_1}, \ldots, x_{j_m})}(a) \neq 0.
\]
Let \((k_1, \ldots, k_{n-m})\) be the complement of \((j_1, \ldots, j_m)\).

There exist open sets \(V, W\) in \(\mathbb{R}^n, \mathbb{R}^{n-m}\), respectively such that

\[
(a_{k_1}, \ldots, a_{k_{n-m}}) \in V, \quad (a_{j_1}, \ldots, a_{j_m}) \in W, \quad V \times W \subset U,
\]

and a unique differentiable \(\phi = (\phi_1, \ldots, \phi_m) : V \to W\) such that if \(x\) is such that \((x_{k_1}, \ldots, x_{k_{n-m}}) \in V, (x_{j_1}, \ldots, x_{j_m}) \in W\), then \(f(x) = 0\) if and only if

\[
x_{\ell} = \phi_\ell(x_{k_1}, \ldots, x_{k_{n-m}}), \quad \ell = 1, \ldots, m.
\]

Moreover, if in addition \(f \in C^k(U)\) for some \(k \geq 2\), then \(\phi \in C^k(V)\).

To avoid all these many subindices, most textbooks will assume that the non-zero minor is indexed by the last \(m\) components of the variable; that is, is \(\frac{\partial(f_1, \ldots, f_m)}{\partial(x_{n-m+1}, \ldots, x_n)}\), and say that one can always reduce the problem to that case by relabeling the variables. In addition they will distinguish the first \(n-m\) variables from the last \(m\), so that instead of having \(x = (x_1, \ldots, x_n)\) one works with \((x, y)\) where \(x = (x_1, \ldots, x_{n-m})\) and \(y = (y_1, \ldots, y_m)\). But we may as well relabel the dimensions, use \(n\) instead of \(n-m\). So now the theorem looks as follows.

**Theorem 5 (IMPFT2)** Let \(U\) be open in \(\mathbb{R}^n\), \(V\) be open in \(\mathbb{R}^m\), let \(f : U \times V \to \mathbb{R}^m\) (notice: same \(m\) as the space containing the open set \(V\)), assume \(f \in C^k(U \times V)\) for some \(k \geq 1\). Points of \(U\) will be denoted generically by \(x = (x_1, \ldots, x_n)\), of \(V\) by \(y = (y_1, \ldots, y_m)\), so we write \(f(x, y)\) for the value of \(f\) at a general point of \(U \times V\). Let \(a = (a_1, \ldots, a_n) \in U\), \(b = (b_1, \ldots, b_m) \in V\) and assume

\[
\frac{\partial(f_1, \ldots, f_m)}{\partial(y_1, \ldots, y_m)}(a, b) \neq 0. \quad \text{(End of the hypotheses)}
\]

There exist open sets \(U_0 \subset U, V_0 \subset V\) such that \(a \in U_0, b \in V_0\), and a unique differentiable \(\phi : U_0 \to V_0\) such that if \((x, y) \in U_0 \times V_0\), then \(f(x, y) = 0\), if and only if \(y = \phi(x)\). Moreover, \(\phi\) is at least as many times continuously differentiable as \(f\).

Notice that for every \(x \in U_0\), \((x, \phi(x)) \in U_0 \times V_0\), thus \(f(x, \phi(x)) = 0\) for all \(x \in U_0\). Since \((a, b) \in U_0 \times V_0\) and \(f(a, b) = 0\), we have \(\phi(a) = b\).

Both theorems, INVFT and IMPFT are equivalent. One can reduce the proof of either of them to the other. Some textbooks derive implicit function as a consequence of inverse function, others do the opposite.