A Review of Differential Calculus

S. Kumaresan
School of Math. and Stat.
University of Hyderabad
Hyderabad 500046
kumaresa@gmail.com

The aim of the article is to briefly recall the definitions and the important theorems the reader must have learnt in a course on calculus of several variables. Here our concern will be to bring out the geometric ideas behind the concepts and the theorems. Hence there will be a lot of examples but no proofs. The reader is referred to my book "A Cours ein Differential geometrty and Lie groups" for details.

Let us begin with the one variable set-up. Let $f\colon (a,b) \to \mathbb{R}$ be a function. Let $a < x < b$. We say that f is differentiable at x if the limit $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ exists. If the limit exists it is usually denoted by $f'(x)$ and called the *derivative* of f at x. Another way of saying this is as follows: There exists a unique real number $f'(x)$ such that

$$f(x+h) - f(x) = f'(x)h + o(|h|) \tag{1}$$

where $|h|$ is a read as *little oh of mod* h and is a quantity which goes to 0 faster than $|h|$ as $h \to 0$, i.e., $\lim_{h\to 0}(o(|h|)/|h|) = 0$.

What is so important about this formulation? It tells us that the derivative of f at x can be thought of as the linear map from $\mathbb{R}$ to $\mathbb{R}$ given by $h \mapsto f'(x)h$ and that this linear map is a good approximate for f at x. Thus a complicated function f can be described (at a point) by a linear function (such as $h \to f'(x)h$), a very nice function next only to constants!

The basic idea of differential calculus is to study the local behaviour of a function at a point by its first order (linear) approximation at the same point. In case the reader finds this vague we suggest that he bears with us for some more time after which he will understand the meaning of this sentence. such words! Geometrically if we think of f as its graph $\{(x, f(x)) : x \in \text{ domain of } f\} \subset \mathbb{R}^2$, then the derivative of f at x gives the slope of the tangent line at $(x, f(x))$ to the curve, i.e., "the line corresponding to the linear map above", which is the best approximate at that point. Now this formulation (1) of derivative of a function f is easily adapted to maps $f\colon U \subset \mathbf{E} \to \mathbf{F}$ where $\mathbf{E}$ and $\mathbf{F}$ are real Banach spaces and U an open subset of $\mathbf{E}$. For the definition of Banach spaces the reader may refer to the appendix to this chapter. To understand most of what follows it is not necessary that they are infinite dimensional. You may as well assume that $\mathbf{E} = \mathbb{R}^m$ and $\mathbf{F} = \mathbb{R}^n$.

We want to imitate (1) in formulating the concept of derivative of f: "$f(x+h) - f(x) = A(h) + o(h)$", for $x \in U$, h in a sufficiently small neighbourhood W of x so that $x + W := \{y \in \mathbf{E} : \exists w \in W \text{ such that } y = x + w\}$. Now, clearly, we wish A to be a linear map approximating

f at x. Since $f(x+h) - f(x) \in \mathbf{F}$, and $h \in \mathbf{E}$, we see that A must be a map from $\mathbf{E}$ to $\mathbf{F}$. If $\mathbf{E}$ and $\mathbf{F}$ are finite dimensional then no further condition on A need be imposed. However if one of them is infinite dimensional then we require A to be continuous (and linear).

With the above notation, we say that f is *(Fréchet) differentiable* at x if there exists a continuous linear map $A\colon \mathbf{E} \to \mathbf{F}$ such that

$$f(x+h) - f(x) = A(h) + o(\|h\|) \tag{2}$$

holds for all $h \in \mathbf{E}$ with sufficiently small norm. An equivalent way is to say that

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

If f is differentiable and if B satisfies (2) with B replacing A then $A = B$, so that such an A is unique. (Exercise: Prove this.) We call A the *Fréchet derivative* of f at x, and denote it either by $f'(x)$ or by $Df(x)$.

Ex. 1. If f is differentiable at x then it is continuous at x. (In some text-books this is given as the motivation for the definition above thereby distorting the geometric meaning behind it!)

We shall look at some examples.

Example 2. Let $f := A\colon \mathbf{E} \to \mathbf{F}$ be a continuous and linear map. Since it is already linear it is its own best linear approximation at any point of $\mathbf{E}$ and hence we should expect $f'(x) = A$, for all $x \in \mathbf{E}$. We shall convince ourselves of this: $f(x+h) - f(x) = f(x) + f(h) - f(x) = f(h)$, since f is linear. Hence we can take $f'(x)(h) = f(h)$ so that $o(\|h\|) = 0$ in this situation.

Question: If f is a constant map what is $f'(x)$?

Question: Can you define the second derivative $f''(x)$ of a function? What is $f''(x)$ for f as in Example 1?

Example 3. Let $\mathbf{E} := M(n, \mathbb{R}) := \{X = (x_{ij}) : x_{ij} \in \mathbb{R}, 1 \leq i, j \leq n\}$, the set of all real $n \times n$ matrices. Then $\mathbf{E}$ is a finite dimensional vector space over $\mathbb{R}$. In fact, the map $X := (x_{ij}) \mapsto (x_{11}, x_{12}, \ldots, x_{nn-1}, x_{nn}) \in \mathbb{R}^{n^2}$ establishes a vector space isomorphism. We make $\mathbf{E}$ a Banach space taking the norm of X to be the operator norm of X i.e., by setting

$$\|X\| := \sup_{\{v \in \mathbf{E} : \|v\| = 1\}} \|Xv\|.$$

We now consider the map $f\colon \mathbf{E} \to \mathbf{E}$ given by $f(X) := X^2$. Here $X^2 = X \cdot X$, the matrix multiplication. We wish to show that this map is differentiable on all of $\mathbf{E}$ and compute its derivative. Let $H \in \mathbf{E}$. We have

$$\begin{aligned} f(X+H) - f(X) &= (X+H)^2 - X^2 \\ &= X^2 + XH + HX + H^2 - X^2 \\ &= (XH + HX) + o(\|H\|). \end{aligned}$$

So if we define $f'(X)(H) = XH + HX$, then $H \mapsto XH + HX$ is continuous and linear, and $f'(x)$ satisfies (2).

Example 4. Suppose $\mathbf{E}$, $\mathbf{F}$ and $\mathbf{G}$ are real Banach spaces and $f\colon \mathbf{E}\times\mathbf{F}\to\mathbf{G}$ is a continuous bilinear map. Let $x,\ h\in\mathbf{E}$ and $y,\ k\in\mathbf{F}$. Then we have

$$\begin{aligned} f(x+h,y+k)-f(x,y) &= f(x,y)+f(x,k)+f(h,y)+f(h,k)\\ &\quad -f(x,y)\\ &= f(x,k)+f(h,y)+f(h,k)\\ &= f(x,k)+f(h,y)+o(\|(h,k)\|). \end{aligned}$$

Thus we find that $f'(x,y)(h,k)=f(x,k)+f(h,y)$. Notice that the right side is continuous and linear on the space $\mathbf{E}\times\mathbf{F}$.

The above easily generalises to continuous multilinear maps on $\mathbf{E}_1\times\cdots\times\mathbf{E}_n$ to another Banach space $\mathbf{F}$. If f is such a map then one has:

$$f'(x_1,\ldots,x_n)(h_1,\ldots,h_n)=?$$

Work this out immediately as this is needed below soon.

We specialise this to a situation which may be familiar to the reader.

Let $\mathbf{E}_i=\mathbb{R}^n$, for $1\le i\le n$ and $x_i\in\mathbf{E}_i$. Then the determinant function

$$f\colon=\det\colon \mathbf{E}_1\times\cdots\times\mathbf{E}_n\to\mathbb{R}$$

is defined by $\det(x_1,\ldots,x_n):=\det(x_{ij})$ where $x_i=\sum_j x_{ij}e_j$. Here e_j is the usual basis vector $(0,\ldots,1,\ldots,0)$. Recall that det is an alternating multilinear function of its variables:

$$\det(x_1,\ldots,x_n)=sign(\sigma)\det(x_{\sigma(1)},\ldots,x_{\sigma(n)})$$

for σ a permutation of $\{1,\ldots,n\}$. Using the above exercise we find the familiar rule for the differentiation of the determinants:

$$f'(x_1,\ldots,x_n)(h_1,\ldots,h_n)=\sum_i\det(x_1,\ldots,h_i,\ldots,x_n). \tag{3}$$

The reader is strongly urged to check this. The reader might have seen this when the entries are differentiable functions in the following form: if $F:=\det(f_{ij}):=\det(F_1,\ldots,F_n)$ then

$$F'=\sum_i\det(F_1,\ldots,F_i',\ldots,F_n),$$

in an obvious notation.

We can reformulate the above in a different set-up. As in Example 3 we can identify $\mathbb{R}^n\times\cdots\times\mathbb{R}^n$ (n-times) with $M(n,\mathbb{R})$ so that the above map is $f\colon M(n,\mathbb{R})\to\mathbb{R}$ given by $f(X)=\det(X)$. The tuple $(e_1,\ldots,e_n)$ of the canonical basis vectors goes under this identification to I, the identity matrix. Now what is $f'(I)(H)$? Let $H:=(h_{ij})\in M(n,\mathbb{R})$. We use (3) where $h_i=(h_{i1},\ldots,h_{in})\in\mathbf{E}=\mathbb{R}^n$ and $x_i=e_i$. We then get

$$\det(x_1,\ldots,h_i,\ldots,x_n)=\det\begin{pmatrix}1&0&0&\ldots&0&0\\0&1&0&\ldots&0&0\\\vdots&\vdots&\ddots&\ldots&0&0\\h_{i1}&h_{i2}&\ldots&h_{ii}&\ldots&h_{in}\\\vdots&\vdots&\vdots&\vdots&\vdots&\vdots\\0&0&0&0&0&1\end{pmatrix}=h_{ii}.$$

Thus we have

$$f'(I)(H) = \sum_i h_{ii} = \operatorname{tr}(H). \tag{4}$$

We shall return to this example later.

Example 5. We shall now consider U, the set of all invertible continuous linear operators on a Banach space $\mathbf{E}$. We claim that U is an open subset of $BL(\mathbf{E})$, the Banach space of all bounded (i.e., continuous) linear operators on $\mathbf{E}$. (The norm on $BL(\mathbf{E})$ is the operator norm; see Example 2.) We first observe that if $\mathbf{E}$ is finite dimensional, say $\mathbf{E} = \mathbb{R}^n$ then we have

$$\begin{aligned} U &= GL(n, \mathbb{R}) \ = \{A \in M(n, \mathbb{R}) : A^{-1} \text{ exists i.e., } \det(A) \neq 0\} \\ &= \det{}^{-1}(\mathbb{R} \setminus \{0\}). \end{aligned}$$

Since det is a polynomial function in the entries x_{ij}, it is continuous. Hence U is open in $M(n, \mathbb{R}) \cong \mathbb{R}^{n^2}$. We shall give a different proof of this fact which works in all cases.

Let $A \in U$. To show that U is open we must exhibit an open set W of 0 in $BL(\mathbf{E})$ such that $A + W \subset U$. Let $H \in \mathbf{E}$. We shall try to find $(A + H)^{-1}$ formally:

$$(A + H)^{-1} = (A(I + A^{-1}H))^{-1} = (I + A^{-1}H)^{-1}A^{-1}.$$

Now $(I + A^{-1}H)^{-1}$ looks like $(1 + x)^{-1}$, which has the binomial expansion $\sum_k (-1)^k x^k$, provided that $|x| < 1$. This suggests that we consider the series

$$\sum_k (-1)^k (A^{-1}H)^k = I - (A^{-1}H) + (A^{-1}H)^2 - \cdots \tag{5}$$

The series of operator norms $\|I\| - \|A^{-1}H\| + \left\|(A^{-1}H)^2\right\| - \cdots$ is absolutely convergent *if* $\|A^{-1}H\| < 1$. (since $\|AB\| \leq \|A\| \|B\|$ for all A, $B \in BL(\mathbf{E})$. - Exercise) The above condition is achieved if $\|H\| < \|A^{-1}\|^{-1}$. Hence the series (5) is norm convergent in $BL(\mathbf{E})$ to an element B such that $(A + H)B = I = B(A + H)$. Thus if we take $W = \{H \in BL(\mathbf{E}) : \|H\| < \|A^{-1}\|^{-1}\}$, then $A + W \subset U$, i.e., U is open.

We now consider the map $f \colon U \to U$ given by $A \mapsto A^{-1}$. We want to show that f is differentiable and compute its derivative. Let $A \in U$ and $H \in BL(\mathbf{E})$. We have

$$\begin{aligned} f(A + H) - f(A) &= (A + H)^{-1} - A^{-1} \\ &= (I - (A^{-1}H) + o(\|H\|))\, A^{-1} - A^{-1} \\ &= -A^{-1}HA^{-1} + o(\|H\|). \end{aligned}$$

Exercise: Justify the above set of equations. Hence we find that

$$f'(A)(H) = -A^{-1}HA^{-1} = -L_{A^{-1}} \circ R_{A^{-1}}(H)$$

where $L_B(X) = BX$, the left multiplication by B on $BL(\mathbf{E})$, etc.

Example 6. In all the previous examples and definitions we worked with real Banach spaces and real linear maps. We wish to consider $f \colon U \subset \mathbb{C} \to \mathbb{C}$, a complex analytic (i.e., holomorphic) function. We say that f as above is holomorphic on U if for all $z \in U$ there is $f'(z) \in \mathbb{C}$

such that $\lim_{h\to 0}\frac{f(z+h)-f(z)}{h} = f'(z)$, and $z \to f'(z)$ is continuous. In other words, for all $z \in U$ there exists $f'(z) \in \mathbb{C}$ such that

$$f(z+h) - f(z) = f'(z)h + o(|h|) \tag{6}$$

and $z \mapsto f'(z)$ is continuous. We consider f as a map from $U \subset \mathbb{R}^2$ to $\mathbb{R}^2$. Let us denote this map as F so as to avoid confusion. Now we ask: Does $F'(x,y)$ exist and what is it for $(x,y) \mapsto x+iy := z \in \mathbb{C}$? Notice that (6) implies that the $\mathbb{C}$-linear map $h \mapsto f'(z)h$ is a $\mathbb{C}$-linear approximation to f at z. Hence we must expect that the map $h \mapsto f'(z)h$ which is apriori $\mathbb{R}$-linear "is $F'(x,y)$." We shall explain this in detail.

Let $f'(z) = a+ib$, and $h = h_1 + ih_2$. Then we have

$$f'(z)h = (a+ib)(h_1+ih_2) = (ah_1 - bh_2) + i(bh_1 + ah_2).$$

Thus the $\mathbb{C}$-linear map from $\mathbb{C}$ to $\mathbb{C}$ given by $h \mapsto f'(z)h$ has as its underlying $\mathbb{R}$-linear map, the map

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \mapsto \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = {}^t(ah_1 - bh_2, bh_1 + ah_2).$$

Thus the underlying real linear map is the map

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \mapsto \begin{pmatrix} ah_1 - bh_2 \\ bh_1 + ah_2 \end{pmatrix}$$

Thus we would like to claim that $F'(x,y) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. If we write $f = u+iv$ in the standard notation, then

$$F'(x,y) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \text{ at } (x,y).$$

Since f is holomorphic we have the Cauchy-Riemann equations

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \text{ and } \frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}.$$

Hence we have

$$F'(x,y) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ -\frac{\partial u}{\partial y} & \frac{\partial u}{\partial x} \end{pmatrix}.$$

But then $\frac{\partial u}{\partial x} = a$ and $\frac{\partial u}{\partial y} = -b$. This implies that $f'(z) = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}$ at z as claimed. Thus we have shown that $F'(x,y)$ is the $\mathbb{R}$-linear map underlying the $\mathbb{C}$-linear map $\zeta \mapsto f'(z)\zeta$.

On first reading the reader may skip the following and go to the next example.

Conversely, let $F\colon U \subset \mathbb{R}^2 \to \mathbb{R}^2$ be C^1 i.e., $F'(x,y)$ exists and is continuous as a map from U to $M(2,\mathbb{R})$. This is same as saying that the partial derivatives of $F = (u,v)$ exist and are continuous. Then the corresponding map $f\colon U \subset \mathbb{C} \to \mathbb{C}$ given by $f(z) = u(x,y) + iv(x,y)$ for $z = x+iy$ is complex analytic if the $\mathbb{R}$-linear maps $F'(x,y)$ are "$\mathbb{C}$-linear", for all $(x,y) \in U$.

The last statement requires some explanation. Given $\mathbb{R}^2$, how do we "recognize" it as $\mathbb{C}$? $\mathbb{R}^2$ can be recognized as a one dimensional vector space over $\mathbb{C}$ if we endow $\mathbb{R}^2$ with a

$\mathbb{C}$-multiplication by $i := \sqrt{-1}$, a fixed square root of -1. i.e., $j\colon x+iy \mapsto i(x+iy) = -y+ix$ is the scalar multiplication by i on $\mathbb{C}$. This is a $\mathbb{C}$-linear map and corresponds to the $\mathbb{R}$-linear map $J\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$J\colon \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}.$$

Now viewing $\mathbb{R}^2$ as a $\mathbb{C}$-vector space is determined as soon as we know $(\mathbb{R}^2, J)$ for the following reason: Define

$$\begin{aligned} (a+ib)(x+iy) &= \text{“}a(x,y) + J(b(x,y))\text{”} \\ &= \text{“}(ax, ay) + (-by, bx)\text{”} \\ &= \text{“}(ax-by, ay+bx)\text{”} \\ &= (ax-by) + i(ay+bx). \end{aligned}$$

What happens is this. Secretly, we think of (x,y) as $x+iy$ so that $(a+ib)(x+iy) = (ax-by)+i(ay+bx)$, which corresponds to $(ax-by, ay+bx)$. Thus the multiplication by i corresponds to the $\mathbb{R}$-linear map J. Now given an $\mathbb{R}$-linear map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ it corresponds to a $\mathbb{C}$-linear map iff “$A(iz) = iA(z)$”, i.e., iff $A\circ J(x,y) = J\circ A(x,y)$ for all $(x,y) \in \mathbb{R}^2$. i.e., iff $A\circ J = J\circ A$. Let us record our finding as

Lemma 7. *An $\mathbb{R}$-linear map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ is $\mathbb{C}$-linear iff $A\circ J = J\circ A$.* □

If you have really followed us up to this point the reward is near. We apply the above lemma to $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ to conclude:

$$F'(x,y) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \text{ is } \mathbb{C}\text{−linear iff } F'(x,y)\circ J = J\circ F'(x,y).$$

This means that

$$\begin{pmatrix} \frac{\partial u}{\partial y} & +\frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial y} & -\frac{\partial v}{\partial x} \end{pmatrix} = \begin{pmatrix} -\frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \\ +\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{pmatrix} \quad \text{i.e., iff } \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \text{ and } \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

That is, $F = (u,v)$ is holomorphic iff “F satisfies Cauchy-Riemann equations.” We summarize all these in the following

Theorem 8. *Let $F\colon U \subset \mathbb{R}^2 \to \mathbb{R}^2$ be C^1. Let us write the function F as $F(x,y) = (u(x,y), v(x,y))$ and set $f(z) = u(x,y) + iv(x,y)$ for $z := x+iy$. Then f is holomorphic on U iff the Fréchet derivative $F'(x,y)$ is $\mathbb{C}$-linear for all $(x,y) \in U$.* □

Example 9. After all this heady stuff, let us return to the classical setting: $F\colon U \subset \mathbb{R}^m \to \mathbb{R}^n$. If we write $F := (f_1, \ldots, f_n)$, then $F'(x)$ as a linear operator from $\mathbb{R}^m$ to $\mathbb{R}^n$ has the matrix representation $\left(\frac{\partial f_i}{\partial x_j}\right)_{\substack{1\le i\le n \\ 1\le j\le m}}$ with respect to the usual basis of $\mathbb{R}^m$ etc. Since this is the approach which is very well explained in the text books we shall not go any further into this, other than explicating some hidden geometric consequences below. (However see Examples 7 and 8.)

Enough of examples. Let us look into the geometric aspects of differential calculus. We start with the following question: Given f, U, $\mathbf{E}$, and $\mathbf{F}$ etc., $f'(x)$ acts on $v \in \mathbf{E}$. What is the geometric meaning of v or $\mathbf{E}$?

It is instructive to look at the case when $\mathbf{E} = \mathbb{R}^n$ and $\mathbf{F} = \mathbb{R}$. Given $x \in \mathbf{E}$, $f'(x)$ is a linear functional on $\mathbb{R}^n$. i.e., $f'(x) \in (\mathbb{R}^n)^*$, the dual of $\mathbb{R}^n$. Since $\mathbb{R}^n$ is an inner product space, this linear functional is given by inner product with a vector $v \in \mathbb{R}^n$. In calculus course you must have learnt that this vector is grad $f\colon = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n})$. Thus one has

$$f'(x)(v) = \left\langle \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}, v \right\rangle = \sum_i v_i \frac{\partial f}{\partial x_i}.$$

Here $\langle , \rangle$ is the Euclidean inner product and $v = \sum_i v_i e_i$. Now the linear functional $f'(x)$ is known as soon as we know $f'(x)(e_i)$ for $1 \leq i \leq n$. But $f'(x)(e_i) = \frac{\partial f}{\partial x_i}(x)$. The geometric meaning of $\frac{\partial f}{\partial x_i}(x)$ is that it is the derivative of f in the direction of e_i at the point x.

More generally for $v \in \mathbb{R}^n$ if we set

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

then the limit exists since $f'(x)$ exists. This $D_v f(x)$ can be thought of as the derivative of f in the direction (and magnitude) of v or the directional derivative of f in the direction of v. We then have $D_v f(x) = f'(x)(v)$.

Now if we look at the definition of $D_v f(x)$, the quantity $D_v f(x)$ is nothing but the derivative of the function of one variable $t \mapsto f(x+tv)$. Hidden here is the crucial observation that we are restricting the function to the curve γ given by $t \mapsto \gamma(t)$ for t sufficiently small and the fact that this curve γ passes through the point $x = \gamma(0)$ and that the tangent vector to the curve γ at x is $\gamma'(0) = v$. Recall that the tangent vector to a curve $\sigma\colon (-\varepsilon, \varepsilon) \to \mathbf{E}$ at 0 or at $\sigma(0)$ is $\frac{d}{dt}(\sigma)\,|_{t=0}$. This is usually denoted by $\sigma'(0)$. If we write $\sigma = (\sigma_1, \ldots, \sigma_n)$, then $\sigma'(0) = (\sigma_1'(0), \ldots, \sigma_n'(0))$.

The chain rule tells us that we can use any curve σ such that $\sigma(0) = x$ and $\sigma'(x) = v$ to calculate $D_v f(x)$:

$$D_v f(x) = \frac{d}{dt}(f \circ \sigma(t))\,|_{t=0}\ . \tag{7}$$

Exercise: Check this.

Thus the vectors v on which the derivative $f'(x)$ acts can be thought of as tangent vectors to curves passing through x. (††)

The significance of the statement (7) is overlooked in (almost) all text books on differential calculus. To impress it on the minds of the reader we have devised the following example:

Let us consider $f\colon GL(n, \mathbb{R}) = \{A \in M(n, \mathbb{R}) : \det(A) \neq 0\} \to \mathbb{R}$ given by $f(A) = \det(A)$. We have already computed the derivative of f at I: $f'(I)(X) = \operatorname{Tr} X$, the linear functional 'tr' on $M(n, \mathbb{R})$. Now we consider the map $\exp\colon M(n, \mathbb{R}) \to GL(n, \mathbb{R})$ given by

$$e^X := \exp(X) := \sum_{k=0}^{\infty} \frac{X^k}{k!}.$$

For more on this map see the exercise set at the end of this chapter. Then the curve $\gamma(t) := \exp(tX) = e^{tX}$ satisfies $\gamma(0) = I$ and $\gamma'(t) = X$. Hence it follows from Example 3) and (7) that

$$\mathrm{Tr}(X) = f'(I)(X) = \frac{d}{dt} f \circ \gamma(t)\,|_{t=0} = \frac{d}{dt}\det(e^{tX})\,|_{t=0}\,. \tag{8}$$

We now use this to prove

$$\det(e^X) = e^{\mathrm{Tr}(X)} \text{ i.e., } \det(\exp(X)) = \exp(\mathrm{Tr}(X)). \tag{9}$$

Before going into a proof let us ask ourselves whether it sounds plausible. If X is a diagonal matrix, say $X = (x_1, \ldots, x_n)$, i.e., $x_{ij} = \delta_{ij} x_i$. (δ_{ij} is the Kronecker delta which is 1 if $i = j$ and 0 otherwise.) Then $\exp(X) = (e^{x_1}, \ldots, e^{x_n})$ so that (9) is true for all diagonal matrices. Hence it is true for all *diagonalisable* matrices. (Check this.) (A matrix X is diagonalisable if there exist invertible matrices A such that AXA^{-1} is diagonal.) But unfortunately not all matrices are diagonalisable. If you remember your linear algebra well, the second best thing to test (9) is on the matrices which are in the Jordan Canonical form. It is easily verified that (9) remains valid in this case too and hence it is true for all matrices. We shall not go into the details of a proof along these lines. Instead, we shall give a Calculus proof based on " $\det'(I) = \mathrm{Tr}$."

Let us consider the function $g(t) := \det(e^{tX})$ for a mixed $X \in M(n, \mathbb{R})$. We then have

$$\begin{aligned} g'(s) &= \frac{d}{dt} g(s+t) = \frac{d}{dt}\det(e^{(s+t)X})\,|_{t=0} \\ &= \frac{d}{dt}(\det(e^{sX})\ \det(e^{tX}))\,|_{t=0} \\ &= \det(e^{sX})\,\frac{d}{dt}\left(\det(e^{tX})\right)|_{t=0} \\ &= g(s)\,\mathrm{Tr}(X) \end{aligned}$$

by (8). Thus we have $g'(t) = g(t)\,\mathrm{Tr}(X)$. It follows therefore that $g(u) = g(0)\,e^{u\,\mathrm{Tr}(X)}$ for all $u \in \mathbb{R}$. In particular by taking $u = 1$ we get (9).

We do not stop here! We shall give another beautiful application of the *geometric principle* above. We use (9) to find $\det'(A)(X)$, for $A \in GL(n, \mathbb{R})$ and $X \in M(n, \mathbb{R})$. One could proceed as in Example 3 to find a formula for $f'(A)(X)$ in terms of coordinates of $A = (a_{ij})$ and $X = (x_{ij})$. But it is unlikely that you will get the nice formula for $f'(A)(X)$ that we are going to derive below!

After this sales talk let us get on with our business. So, given $X \in M(n, \mathbb{R})$, we want to find a curve σ such that $\sigma(0) = A$ and $\sigma'(0) = X$. On first impulse one may try $\sigma(t) = A\,e^{tX}$ so that $\sigma(0) = A$. But however notice that $\sigma'(0) = AX$ *not* X. But this situation is easily remedied. How about setting $\gamma(t) = A\,e^{tA^{-1}X}$? Verify it works! Thus we have

$$\begin{aligned} f'(A)(X) &= \frac{d}{dt} f \circ \gamma(t)\,|_{t=0} \\ &= \frac{d}{dt}\det(A\,e^{tA^{-1}X})\,|_{t=0} \\ &= \det(A)\,\frac{d}{dt}|_{t=0}\ e^{t\ \mathrm{Tr}(A^{-1}X)} \\ &= \det(A)\ \mathrm{Tr}(A^{-1}X). \end{aligned}$$

You should work this out using the approach in Example 3 and verify that the expression obtained that way is the same as the one on the right side of the last equation. This will convince you of the merit of the geometric principle enunciated above.

There is one more thing that we would like to point out: viz., the light that the geometric principle sheds on the Taylor's formula for an $\mathbb{R}$-valued function of several variables. Let U be a convex or star-shaped open set in $\mathbb{R}^n$. Let $x \in U$ and $y \in \mathbb{R}^n$. Assume that $x + y \in U$ and $f : U \to \mathbb{R}$ is $(k+1)$-times continuously differentiable function. Then the Taylor's formula gives us an expression for $f(x+y)$ in terms of f and its derivatives *in the direction of* y. The *increment* in the variable x is y and we wish to approximate (express) the *increment* in the value i.e., $f(x+y) - f(x)$ by means of the change of f at x in the direction of y. This intuitive idea suggests us that we consider the function $g(t) := f(x+ty) = f \circ \gamma(t)$, say, in an obvious notation. Then clearly g is $(k+1)$-times continuously differentiable function of t. We can therefore apply the Taylor's formula for functions of one variable to g to get:

$$g(t) = g(0) + \sum_{0 \leq j \leq k} \frac{t^j}{j!} g^{(k)}(0) + O(t^{k+1})$$

and hence we get

$$f(x+ty) = f(x) + \sum_{0 \leq j \leq k} \frac{t^j}{j!} D_y^j g(x) + O(t^{k+1}).$$

This is the Taylor's formula for real valued functions of a vector variable. This insight brings out a couple of points: 1) Considering the one variable function g is natural, not " a hat trick" and 2) the Taylor coefficients of the function f are the directional derivatives of f at x along y. Of course in applications of analysis one needs various forms of the remainder term in this formula. Our only point here is to exhibit the underlying geometric content of this result.

One may wonder why we started with a not necessarily finite dimensional real Banach space. The next couple of examples deals with the infinite dimensional set-up.

Example 10. Let $\mathbf{E}$ be a real Hilbert space. Consider $f(x) := \langle x, x \rangle$ for $x \in \mathbf{E}$. What is $f'(x)(h)$ for $h \in \mathbf{E}$? Question: What can you say if we take $g(x) := \langle x, x \rangle^{1/2}$? Investigate what happens when f is as above but $\mathbf{E}$ is a Hilbert space over $\mathbb{C}$.

Example 11. Let $\mathbf{E} := C_0^1[0,1]$ be the space of all real valued C^1-functions on [0,1] such that $f(0) = 0$. Let the norm be given by:

$$\| f \| := \sup_{0 \leq t \leq 1} |f(t)| + \sup_{0 \leq t \leq 1} |f'(t)|.$$

Then it is easy to see that $\mathbf{E}$ is a real Banach space. Let $\mathbf{F} := C[0,1]$ be the Banach space of continuous functions g on [0,1] with the supremum norm $\| g \| := \sup_{0 \leq t \leq 1} |g(t)|$. Consider the map $T : \mathbf{E} \to \mathbf{F}$, given by $Tf = f' + f^k$, for $k \in \mathbb{N}$. Here $f' := \frac{d}{dt} f$. Then $T'(f)(g) = g' + k\, f^{(k-1)}\, g$. (Exercise.)

Most often differential calculus is used to solve non-linear problems by linearising them i.e., by "taking derivatives " and applying either the implicit function theorem or the inverse function theorem. Before giving a simple minded example of such an application, we take this opportune moment to review the inverse mapping theorem (I.M.T., for short).

Theorem 12 (Inverse Mapping Theorem)**.** *Let* $\mathbf{E}$ *and* $\mathbf{F}$ *be Banach spaces,* $U \subset \mathbf{E}$ *be an open set. Let* $f\colon U \to \mathbf{F}$ *be* C^k *with* $1 \le k \le \infty$*. Assume that* $p \in U$ *is such that* $f'(p)$ *is bijective (so that* $f'(p)^{-1}$ *exists and is continuous). Then there is an open subset* V *of* U *containing the point* p *with the following properties: i) on* V *the map is one-one, ii) the image* $f(V)$ *is an open neighborhood of* $f(p)$ *and iii)* f^{-1} *is* C^k *on* $f(V)$ *with* $Df^{-1}(y) = Df(f^{-1}(y))^{-1}$ *for all* $y \in V$*.* □

Even though we have not defined the second, third derivatives etc., the reader should have no difficulty in formulating these notions (or consult the appendix). If you have seen a proof in the finite dimensional case which used the contraction mapping theorem then it is trivial to see that the same proof works in the case of any Banach space too. In any case we shall give a proof of a slightly more general version of the Implicit Function Theorem in the Banach space set-up from which it is easy to obtain the I.M.T.

First of all we shall try to explain the geometric meaning of I.M.T. in the finite dimensional case. Next we shall indicate by means of a simple example how I.M.T. helps in nonlinear problems.

Let $f\colon \mathbb{R}^{n+1} \to \mathbb{R}$ be C^1. Assume that 0 lies in the image of f, without loss of generality. In general, the set $S := f^{-1}(0)$ does not have any nice geometric property. However, if we assume that

$$\text{for all } p \in S, \quad f'(p) := \operatorname{grad} f(p) \neq 0 \tag{10}$$

then S "looks locally like a hyperplane." Of course this needs explanation! Given $p \in S$, since $f'(p) \neq 0$, we assume without loss of generality that $\frac{\partial f}{\partial x_{n+1}} \neq 0$. Then consider the map $\Phi\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ given by $(x_1, \dots, x_n, x_{n+1}) \mapsto (x_1, \dots, x_n, f(x))$. Then $\Phi'(p)$ has the Jacobian

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \dots & \frac{\partial f}{\partial x_{n+1}} \end{pmatrix}.$$

The determinant of this matrix is $\pm\frac{\partial f}{\partial x_{n+1}} \neq 0$ and hence $\Phi'(p)$ is invertible. Thus Φ is a "C^1-diffeomorphism" of an open neighborhood of V of p in $\mathbb{R}^{n+1}$ onto an open set $\Phi(V) \subset \mathbb{R}^{n+1}$, by I.M.T. (By a C^1-diffeomorphism we mean a map F which is C^1, one-one on its domain U and $F(U)$ is open and F^{-1} is also C^1 on $F(U)$.) Now we introduce a new set of coordinates on V by setting $y_i(r) := u_i \circ \Phi(r)$, where u_i are the 'usual' coordinates on $\mathbb{R}^{n+1}$: $u_i(x) := x_i$. In plain language this is:

$$y_i(r) = \begin{cases} r_i & \text{if } 1 \le i \le n \\ f(r) & \text{if } i = n+1. \end{cases}$$

With respect to this new set of coordinates y_i, S has a local description around p on $V \cap S$: it is the "hyperplane" $\{y_{n+1} = 0\}$. Thus by taking a suitable system of coordinates we "straighten" the hypersurface to a hyperplane locally.

To see how this change of coordinates can help us we pose the following question: Suppose the hypersurface S is also described around p as $g^{-1}(0)$. That is, there exists an open set $U \ni p$ such that $S \cap U = g^{-1}(0)$, with $g\colon U \to \mathbb{R}$ being a C^1-function. Is g divisible by f at least locally around p? That is, does there exist another function h defined in an open set containing p on which we can write $g = fh$?

Let $F := f \circ \Phi^{-1}$ and $G := g \circ \Phi^{-1}$. Then it follows that $\Phi(V \cap S) = \{y_{n+1} = 0\} \cap \Phi(V)$. The above question then reduces to an equivalent one: Is G divisible by $F = y_{n+1}$ locally around 0? This is certainly easy to answer (in the affirmative, by Taylor expansion).

The moral therefore is that the I.M.T. allows us to use a coordinate system that is most convenient or that simplifies the geometric problem on hand.

We now return to the infinite dimensional Example 8. We have $DT(0)(g) = g'$. That is, $DT(0) = \frac{d}{dt}\colon C_0^1[0,1] \to C[0,1]$. This is a continuous one-one onto linear map, the inverse being given by the indefinite integral I:

$$I(h) := \text{the function } t \mapsto \int_0^t h(s)\, ds.$$

(See the fundamental theorem of calculus in the appendix.) Hence T maps an open ball centered at 0 in the domain space onto a neighborhood of 0 in the range space bijectively. In particular, there exists a $\delta > 0$ such that if $g \in C[0,1]$ and $\|g\| < \delta$, then there is an $f \in C_0^1[0,1]$ such that the equation $Tf = g$ holds. Thus we have solved the nonlinear problem $Tf = g$ for g with $\|g\| < \delta$.

Example *cum* **exercise:** Let $k \in C(I \times I)$ where $I := [0,1]$. Then the operator K defined by

$$Kf(x) := \int_0^1 k(x,y)\, f(y)\, dy$$

is a continuous linear map on the Banach space $C(I)$. Assume that $\lambda \notin \operatorname{spec}(K) := \{\lambda \in \mathbb{C} : (K - \lambda I)^{-1} \notin BL(C(I))\}$. Then there exists an $\varepsilon > 0$ such that the nonlinear equation

$$\lambda f(x) = \int_0^1 k(x,y)\left(f(y) + [f(y)]^2\right) dy + g(x) \tag{11}$$

has a solution in $C(I)$, for all $g \in C(I)$ with $\|g\| < \varepsilon$. This is easy. Consider the operator A given by $Af :=$ the left side $-$ the right side of the equation (11). Then $A'(0) = \lambda I - K$, which is invertible. Complete the details.

Example *cum* **exercise:** Let $Af(x) := \int_0^1 k(x,y,f(y))\, dy$ for $f \in C(I)$. Assume that k, $\frac{dk}{du}$, the partial derivative with respect to the third variable are continuous as functions from $I \times I \times \mathbb{C} \to \mathbb{C}$. Then show that $A'(f)(h) = \int_0^1 \frac{dk}{du}(x,y,f(x))\, h(y)\, dy$.