# LECTURE 8

**Date of Lecture:** February 10, 2020

As always, $\mathbf{K} \in \{\mathbf{R}, \mathbf{C}\}$.

The symbol is for flagging a cautionary comment or a tricky argument. It occurs in the margins and is Knuth's version of Bourbaki's "dangerous bend symbol".

An $n$-tuple $(x_1, \dots, x_n)$ of symbols ($x_i$ not necessarily real or complex numbers) will also be written as a column vector when convenient. Thus

$$(x_1, \dots, x_n) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

A map $\boldsymbol{f}$ from a set $S$ to a product set $T_1 \times \cdots \times T_n$ will often be written as an $n$-tuple $\boldsymbol{f} = (f_1, \dots, f_n)$, with $f_i$ a map from $S$ to $T_i$, and hence, by the above convention, as a column vector

$$\boldsymbol{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}.$$

(See Remark 2.2.2 of Lecture 5.)

The default norm on Euclidean spaces of the form $\mathbf{R}^n$ is the Euclidean norm $\| \,\|_2$ and we will simply denote it as $\| \,\|$.

Note that $(x_1, \dots, x_n) \neq [x_1 \dots x_n]$. Each side is the transpose of the other.

## 1. **Notations and basic results**

### 1.1. **Line segments in $\mathbf{R}^n$ and the Mean Value Theorem.**

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be distinct points in $\mathbf{R}^n$. We write $[\boldsymbol{u}, \boldsymbol{v}]$ for the line segment joining $\boldsymbol{u}$ and $\boldsymbol{v}$. Thus

$$[\boldsymbol{u}, \boldsymbol{v}] = [\boldsymbol{v}, \boldsymbol{u}] = \{\boldsymbol{u} + t(\boldsymbol{v} - \boldsymbol{u}) \mid t \in [0, 1]\}. \tag{1.1.1}$$

An *interior point* of $[\boldsymbol{u}, \boldsymbol{v}]$ is point in $[\boldsymbol{u}, \boldsymbol{v}] \smallsetminus \{\boldsymbol{u}, \boldsymbol{v}\}$. Thus $\boldsymbol{\xi}$ is an interior point of $[\boldsymbol{u}, \boldsymbol{v}]$ if $\boldsymbol{\xi} = \boldsymbol{u} + t(\boldsymbol{v} - \boldsymbol{u})$ for some $t \in (0, 1)$.

Suppose $U$ is open in $\mathbf{R}^n$ and $f\colon U \to \mathbf{R}$ is a map such that for some $j \in \{1, \dots, n\}$, $f$ is continuous on a segment $[\boldsymbol{a}, \boldsymbol{a} + s\boldsymbol{e}_j] \subset U$ where $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ is the standard basis of $\mathbf{R}^n$ and that $D_j\boldsymbol{f}$ exists for every interior point of $[\boldsymbol{a}, \boldsymbol{a} + s\boldsymbol{e}_j]$. Then it is clear from the standard mean value theorem that there exists in interior point $\boldsymbol{\xi}$ of $[\boldsymbol{a}, \boldsymbol{a} + s\boldsymbol{e}_j]$ such that

$$f(\boldsymbol{a} + s\boldsymbol{e}_j) - f(\boldsymbol{a}) = sD_jf(\boldsymbol{\xi}). \tag{1.1.2}$$

## 2. Characterising $\mathscr{C}^1$ maps

2.1. **Higher derivatives again.** Suppose $\boldsymbol{f}\colon U \to \mathbf{R}^m$ is a map with $U$ an open subset of $\mathbf{R}^n$ and $k$ a positive integer. Recall from §§3.1 of Lecture 6 that we have the notion of $\boldsymbol{f}$ being $k$-times differentiable on $U$, with $k^{\text{th}}$ order derivative $D^k\boldsymbol{f}$. Note that the image space of $D^k$ has, in a certain sense, a complicated description. For example, we know that if $D\boldsymbol{f}$ exists on $U$ then $D\boldsymbol{f}$ takes values in $L(\mathbf{R}^n, R^m)$, or, equivalently, in $M_{mn}$, the space of $M \times n$ matrices (identifying $L(\mathbf{R}^n, \mathbf{R}^m)$ with $M_{mn}$). Then $D^2\boldsymbol{f}$, if it exists, takes values in $L(\mathbf{R}^n, L(\mathbf{R}^n, \mathbf{R}^m))$. In general, if $W_0 = \mathbf{R}^m$, and we define $W_k$ recursively, for $k > 1$, by the formula $W_k = L(\mathbf{R}^n, W_{k-1})$, then for a $k$-times differentiable function $\boldsymbol{f}$ , the map $D^k\boldsymbol{f}$ takes values in $W_k$, so that $D^k\boldsymbol{f}\colon U \to W_k$. One can impose any norm on $W_k$ since they are equivalent, $W_k$ being a finite dimensional $\mathbf{R}$-vector space. Usually, on $W_1 = L(\mathbf{R}^n\, \mathbf{R}^m)$, it is convenient to impose the operator norm, with $\mathbf{R}^n$ and $\mathbf{R}^m$ being given the Euclidean norm. However, sometimes it may be more convenient to impose other norms. As always, we identify $L(\mathbf{R}^n, \mathbf{R}^m)$ with $M_{nm}$.

Recall, again from §§3.1 of Lecture 6, that $\boldsymbol{f}$ as above is said to be in $\mathscr{C}^k$, or $\mathscr{C}^k(U)$ if we wish to be precise, if $\boldsymbol{f}$ is $k$-times differentiable and $D^k\boldsymbol{f}$ is continuous.

2.2. **Continuous partial derivatives.** Suppose $U$ is an open subset of $\mathbf{R}^n$ and $\boldsymbol{f}\colon U \to \mathbf{R}^m$ is $\mathscr{C}^1$. As usual, we write $f_i$ for the $i^{\text{th}}$ component of $\boldsymbol{f}$, so that $\boldsymbol{f} = (f_1, \dots, f_m)$. By Theorem 2.1.3 of Lecture 6 we see that this implies that the partial derivatives $D_j f_i$ exist and are *continuous* on $U$ for $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$. It turns out that these conditions characterise $\mathscr{C}^1$ maps. Here is the precise statement.

**Theorem 2.2.1.** *Let $\boldsymbol{f}\colon U \to \mathbf{R}^m$ be a function on an open subset $U$ of $\mathbf{R}^n$. Then $\boldsymbol{f}$ is in $\mathscr{C}^1(U)$ if and only if all partial derivatives $D_j\boldsymbol{f}$, $1 \le j \le n$, exist and are continuous on $U$*

*Proof.* We have already seen, in the discussion above the statement of the theorem, that if $\boldsymbol{f}$ is in $\mathscr{C}^1$ then the partial derivatives $D_j\boldsymbol{f}$ exist and are continuous.

Conversely suppose all the $D_j\boldsymbol{f}$ exist on $U$ and are continuous. Let us first assume $m = 1$, so that $\boldsymbol{f} = f_1 = f$ (say). Let $A\colon U \to M_{1n}$ be given by the formula

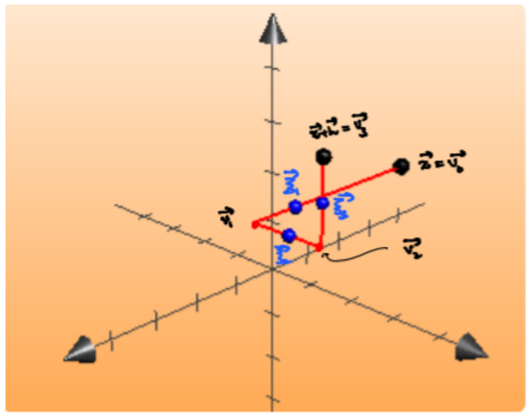
$$(\lozenge) \qquad A(\boldsymbol{x}) = \begin{bmatrix} D_1 f(\boldsymbol{x}) & D_2 f(\boldsymbol{x}) & \dots & D_n f(\boldsymbol{x}) \end{bmatrix} \qquad (\boldsymbol{x} \in U).$$

Fix $\boldsymbol{x} = (x_1, \dots, x_n)$ in $U$ and let $\boldsymbol{h} = (h_1, \dots, h_n) \in \mathbf{R}^n$ be of small enough magnitude that $\boldsymbol{x} + \boldsymbol{h} \in U$. Set

$$\boldsymbol{v}_0 = \boldsymbol{x}, \quad \text{and} \quad \boldsymbol{v}_k = (h_1, \dots, h_k, 0, \dots, 0) + (x_1, \dots, x_n), \qquad (1 \le k \le n).$$

If $\|\boldsymbol{h}\|$ is small enough, $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in U$. By (1.1.2) we can find an interior point $\boldsymbol{\xi}_k$ of $[\boldsymbol{v}_{k-1}, \boldsymbol{v}_k]$, $k = 1, \dots, n$, such that

$$(\dagger) \qquad f(\boldsymbol{v}_k) - f(\boldsymbol{v}_{k-1}) = h_k \cdot D_k f(\boldsymbol{\xi}_k) \qquad (k = 1, \dots, n).$$



In the picture above $n = 3$, $m = 1$, the two black dots are $\boldsymbol{x}$ and $\boldsymbol{x} + \boldsymbol{h}$, the blue dots are $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, and $\boldsymbol{\xi}_3$. For definiteness, take the black dot on the top to be $\boldsymbol{x} + \boldsymbol{h}$.

Let $\boldsymbol{w} = (D_1 f(\boldsymbol{\xi}_1), \dots, D_n f(\boldsymbol{\xi}_n))$ and $\nabla f(\boldsymbol{x}) = (D_1 f(\boldsymbol{x}), \dots, D_n f(\boldsymbol{x}_n))$. Note that $\boldsymbol{\xi}_k \to \boldsymbol{x}$ as $\boldsymbol{h} \to \boldsymbol{0}$ for $k = 1, \dots, n$. Since $D_1 f, \dots, D_n f$ are all continuous by our hypothesis, it follows that

$$(\spadesuit) \qquad \lim_{\boldsymbol{h} \to \boldsymbol{0}} \boldsymbol{w} = \nabla f(\boldsymbol{x}).$$

Summing (†) from $k = 1$ to $k = n$ we get

$$(\ddagger) \qquad f(\boldsymbol{x} + \boldsymbol{h}) - f(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{h} \rangle,$$

where, as before, $\langle \; , \; \rangle$ is the standard inner product on $\mathbf{R}^n$.

Let $A(\boldsymbol{x})$ be as in ($\Diamond$). We have

$$(\Diamond\Diamond) \qquad A(\boldsymbol{x})(\boldsymbol{h}) = \sum_{k=1}^{n} D_k f(\boldsymbol{x}) h_k = \langle \nabla f(\boldsymbol{x}), \boldsymbol{h} \rangle.$$

From (‡), ($\Diamond\Diamond$), and Cauchy-Schwarz, we get

$$\begin{aligned} \frac{1}{\|\boldsymbol{h}\|} \| f(\boldsymbol{x} + \boldsymbol{h}) - f(\boldsymbol{x}) - A(\boldsymbol{x})\boldsymbol{h} \| &= \frac{1}{\|\boldsymbol{h}\|} \| \langle \boldsymbol{w} - \nabla f(\boldsymbol{x}), \boldsymbol{h} \rangle \| \\ &\le \frac{1}{\|\boldsymbol{h}\|} \| \boldsymbol{w} - \nabla f(\boldsymbol{x}) \| \| \boldsymbol{h} \| \\ &= \| \boldsymbol{w} - \nabla f(\boldsymbol{x}) \|. \end{aligned}$$

By ($\spadesuit$) we get $\lim_{\boldsymbol{h} \to \boldsymbol{0}} \| \boldsymbol{w} - \nabla f(\boldsymbol{x}) \| = 0$. It follows that

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{1}{\|\boldsymbol{h}\|} \| f(\boldsymbol{x} + \boldsymbol{h}) - f(\boldsymbol{x}) - A(\boldsymbol{x})\boldsymbol{h} \| = 0,$$

whence $f$ is differentiable at $\boldsymbol{x}$ and the matrix of $f'(\boldsymbol{x})$ is $A(\boldsymbol{x})$.

Now suppose $m \ge 1$ and $\boldsymbol{f} = (f_1, \dots, f_m)$. We have just proved that each $f_j$ is differentiable. For each $\boldsymbol{x} \in U$, define

$$T(\boldsymbol{x}) \colon \mathbf{R}^n \longrightarrow \mathbf{R}^m$$

by the formula

$$T(\boldsymbol{x})\boldsymbol{h} = \begin{bmatrix} f_1'(\boldsymbol{x})\boldsymbol{h} \\ f_2'(\boldsymbol{x})\boldsymbol{h} \\ \vdots \\ f_m'(\boldsymbol{x})\boldsymbol{h} \end{bmatrix} \qquad (\boldsymbol{h} \in \mathbf{R}^n).$$

It is clear that $T(\boldsymbol{x})$ is linear for each $\boldsymbol{x} \in U$ since each $f_j'(\boldsymbol{x})\colon \mathbf{R}^n \to \mathbf{R}$ is. As $\boldsymbol{x}$ varies over $U$, $T(\boldsymbol{x})$ varies over $L(\mathbf{R}^n, \mathbf{R}^m)$ and this obviously defines a map

$$T\colon U \longrightarrow L(\mathbf{R}^n, \mathbf{R}^m).$$

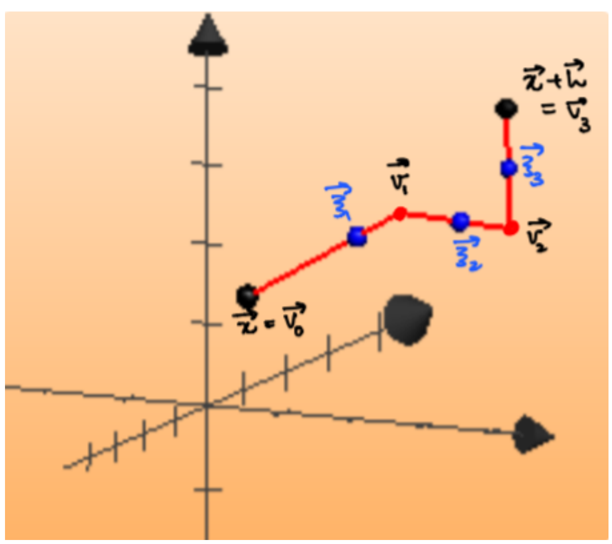
As we will see, $T = \boldsymbol{f}'$.

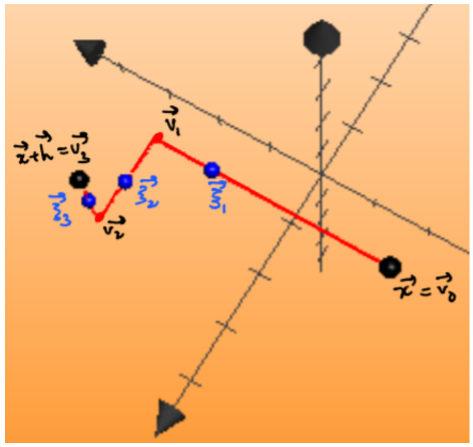
For $\boldsymbol{h} \neq \mathbf{0}$ such that $\boldsymbol{x} + \boldsymbol{h} \in U$, we have

$$\frac{1}{\|\boldsymbol{h}\|}\Big(\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{x}) - T(\boldsymbol{x})\boldsymbol{h}\Big) = \frac{1}{\|\boldsymbol{h}\|}\begin{bmatrix} f_1(\boldsymbol{x}+\boldsymbol{h}) - f_1(\boldsymbol{x}) - f_1'(\boldsymbol{x})\boldsymbol{h} \\ f_2(\boldsymbol{x}+\boldsymbol{h}) - f_2(\boldsymbol{x}) - f_2'(\boldsymbol{x})\boldsymbol{h} \\ \vdots \\ f_m(\boldsymbol{x}+\boldsymbol{h}) - f_m(\boldsymbol{x}) - f_m'(\boldsymbol{x})\boldsymbol{h} \end{bmatrix}.$$

The limit of the right side as $\boldsymbol{h} \to \mathbf{0}$ is $\mathbf{0}$ since each component tends to 0. This means the left side $\to \mathbf{0}$ as $\boldsymbol{h} \to \mathbf{0}$. This shows that $\boldsymbol{f}$ is differentiable on $U$. In fact we have shown that $\boldsymbol{f}' = T$. $\square$

Here are some more pictures illustrating a key idea in the proof






### 3. Towards the inverse function theorem

Throughout we assume $U$ is a convex open set in $\mathbf{R}^n$, $\mathbf{0} \in U$ and $\boldsymbol{f}\colon U \to \mathbf{R}^n$ is a $\mathscr{C}^1$ map with $\boldsymbol{f}(\mathbf{0}) = \mathbf{0}$ and $\boldsymbol{f}'(\mathbf{0}) = I$, the identity linear transformation on $\mathbf{R}^n$.

3.1. **$\boldsymbol{f}$ is one-to-one.** Since $\boldsymbol{f}'$ is continuous on $U$ and the operator norm $\| \ \|$ is continuous on $L(\mathbf{R}^n, \mathbf{R}^m)$, therefore we have an open neighbourhood of $\mathbf{0}$, which we can take to be convex, on which $\|I - \boldsymbol{f}'(\boldsymbol{x})\| \leq \frac{1}{2}$. By replacing $U$ by this neighbourhood, we assume that $\|I - \boldsymbol{f}'(\boldsymbol{x})\| \leq \frac{1}{2}$ for all $\boldsymbol{x} \in U$. Then

1. $\boldsymbol{f}'(\boldsymbol{x}) = I - (I - \boldsymbol{f}'(\boldsymbol{x}))$ is invertible, for $\boldsymbol{x} \in U$, by part (b) of Theorem 2.1.1 of Lecture 7. Indeed $\|I - \boldsymbol{f}'(\boldsymbol{x})\| < 1$ for $\boldsymbol{x} \in U$.
2. Let $\boldsymbol{g}\colon U \to \mathbf{R}^n$ be the map given by $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}$. Then $\boldsymbol{g}$ is $\mathscr{C}^1$ and $\|\boldsymbol{g}'(\boldsymbol{x})\| = \|\boldsymbol{f}'(\boldsymbol{x}) - I\| \leq \frac{1}{2}$ for all $\boldsymbol{x} \in U$. In particular, by the version of the mean value theorem in Theorem 3.1.2 of Lecture 7 we get
$$\|\boldsymbol{g}(\boldsymbol{x}_1) - \boldsymbol{g}(\boldsymbol{x}_2)\| \leq \tfrac{1}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \qquad (\boldsymbol{x}_1, \boldsymbol{x}_2 \in U).$$

For $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $U$, from 2. above we get
$$\begin{aligned} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| - \|\boldsymbol{f}(\boldsymbol{x}_1) - \boldsymbol{f}(\boldsymbol{x}_2)\| &\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2 - (\boldsymbol{f}(\boldsymbol{x}_1) - \boldsymbol{f}(\boldsymbol{x}_2))\| \\ &= \|\boldsymbol{g}(\boldsymbol{x}_1) - \boldsymbol{g}(\boldsymbol{x}_2)\| \\ &\leq \tfrac{1}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \end{aligned}$$
whence
$$\tfrac{1}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \leq \|\boldsymbol{f}(\boldsymbol{x}_1) - \boldsymbol{f}(\boldsymbol{x}_2)\|. \tag{3.1.1}$$
From (3.1.1) it is clear that $\boldsymbol{f}(\boldsymbol{x}_1) = \boldsymbol{f}(\boldsymbol{x}_2)$ if and only if $\boldsymbol{x}_1 = \boldsymbol{x}_2$. Thus $\boldsymbol{f}$ is one-to-one.

**About these notes.** These course notes are a reasonably faithful record of the lectures given at the Chennai Mathematical Institute (CMI) in the January-April semester of 2019-20. The course is Analysis II, a core course for first year undergraduates at CMI. For more material related to this course, visit `https://www.cmi.ac.in/~pramath/teaching.html#ANA2`. If you have comments on these notes, or on related course material, please send an email to pramath@cmi.ac.in.

## References


[R] W. Rudin, *Principles of Mathematical Analysis*, (Third Edition), McGraw-Hill, New Delhi, 1976.

[S] M. Spivak, *Calculus on Manifolds (A modern approach to classical theorems of Advanced Calculus)*, Addison-Wesley, Reading, Massachusetts, 1965.