v: 2013-09-13

REAL ANALYSIS LECTURE NOTES

RASUL SHAFIKOV

1. Functions of several variables: differentiation

1.1. **Vector Space $\mathbb{R}^n$.** We view $\mathbb{R}^n$ as a n-dimensional vector space over the field of real numbers with the usual addition of vectors and multiplication of scalars. The scalar or dot product of two vectors $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ is defined as

$$x \cdot y = \sum_{i=1}^{n} x_i y_i. \tag{1}$$

Together with this dot product $\mathbb{R}^n$ forms and n-dimensional *Euclidean space*. The norm of a vector is then defined as

$$|x| = \sqrt{x \cdot x}.$$

This norm satisfies the following three properties:

(i) $|x| \geq 0$, $|x| = 0$ iff $x = 0$;
(ii) $|cx| = |c||x|$, for all $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$;
(iii) $|x + y| \leq |x| + |y|$, for all $x, y \in \mathbb{R}^n$.

A vector space with a norm satisfying the above three properties is called a *normed space*. The normed also induces a metric on $\mathbb{R}^n$ given by

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2},$$

which is, of course, the standard Euclidean distance in $\mathbb{R}^n$. One can verify that this metric satisfies all three required properties: (i) $d(x, y) \geq 0$, $d(x, y) = 0$ iff $x = y$; (ii) $d(x, y) = d(y, x)$; (iii) $d(x, y) \leq d(x, z) + d(z, y)$ for any $x, y, z \in \mathbb{R}^n$. To prove properties (iii) of the norm and of the metric in $\mathbb{R}^n$ one can use so-called *Minkowski's inequality*:

$$\left(\sum_{i=1}^{n} (a_i + b_i)^k \right)^{1/k} \leq \left(\sum_{i=1}^{n} a_i^k \right)^{1/k} + \left(\sum_{i=1}^{n} b_i^k \right)^{1/k},$$

where $a_i, b_i \geq 0$, and $k > 1$. In fact, Minkowski's inequality is a special case of the *Hölder inequality*:

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1} a_i^p \right)^{1/p} \cdot \left(\sum_{i=1} b_i^q \right)^{1/q},$$

where $a_i, b_i \geq 0$, $p, q > 1$, $1/p + 1/q = 1$. Note that when $p = q = 2$ the Hölder inequality can be written in the form $a \cdot b \leq |a| \cdot |b|$.

The topology on $\mathbb{R}^n$ is induced by the metric: an open set contains a point x together with a small ball

$$\mathbb{B}(x, \varepsilon) = \{ y \in \mathbb{R}^n : |x - y| < \varepsilon \}, \quad \varepsilon > 0.$$

This topology defines $\mathbb{R}^n$ as a *complete* metric space, i.e, every Cauchy sequence with respect to the metric converges to an element of the space. Further, $(\mathbb{R}^n, |\cdot|)$ is a *Banach* space, i.e., a complete normed space. A Banach space is called a *Hilbert* space if its norm comes from a scalar product. Thus, $\mathbb{R}^n$ is a Hilbert space with the scalar product defined by (1).

A map $A : \mathbb{R}^n \to \mathbb{R}^m$ is called *linear* if $A(ax+by) = aA(x)+bA(y)$ for all $x, y \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$. A linear map can be identified with a $n \times m$ matrix with real coefficients. We define the norm of a linear map as

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

It follows immediately from the definition that $|Ah| \leq ||A|| \cdot |h|$ for all $h \in \mathbb{R}^n$.

A map $f : \mathbb{R}^n \to \mathbb{R}^m$ is called *affine* if $f(x) = Ax + B$, where A is a linear map, and B is a constant vector.

1.2. **Continuity.** A domain $\Omega \subset \mathbb{R}^n$ is a connected open set. Given a function $f : \Omega \to \mathbb{R}$ and point $x_0 \in \Omega$, we say that f is continuous at x_0 if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(x_0)| < \varepsilon$ whenever $|x - x_0| < \delta$.

Theorem 1.1. *A function f is continuous at x_0 if and only if $\lim_{j\to\infty} f(x^j) = f(x_0)$ for any sequence of point $(x^j) \to x_0$.*

The proof of $\Rightarrow$ follows from the definition of continuity. To prove the converse formulate the negation of continuity of a function and get a contradiction with the assumption.

Example 1.1. The function $f(x, y) = \frac{xy}{x^2+y^2}$ does not have a limit as $x, y \to 0$, and thus does not admit continuous extension to the origin. On the other hand, the function $g(x, y) = \frac{x^2y}{x^2+y^2}$ has limit equal to 0 as $x, y \to 0$, which follows from the estimate

$$\left|\frac{x^2y}{x^2+y^2}\right| = \left|\frac{xy}{x^2+y^2}\right| |x| \leq \frac{1}{2}|x|.$$

Hence, g become continuous at the origin after setting $g(0) = 0$. $\diamond$

Continuity of maps $f : \mathbb{R}^n \to \mathbb{R}^m$ is defined similarly.

1.3. **Differtiability.** Recall that for $n = 1$, a function $f : \mathbb{R} \to \mathbb{R}$ is called *differentiable* at a point x if the limit

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

exists. This implies that

$$f(x+h) - f(x) = f'(x) \cdot h + r(h),$$

where $r(h) = o(h)$, i.e., $r(h)/h \to 0$ as $h \to 0$. The definition of differentiability in higher dimensions is defined similarly.

Definition 1.2. *Let $\Omega \subset \mathbb{R}^n$ be a domain, $f : \Omega \to \mathbb{R}^m$ be a map, $x \in \Omega$. If there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{h\to 0} \frac{|f(x+h) - f(x) - Ah|}{|h|} = 0, \tag{2}$$

then we say that f is differentiable at x and we write $f'(x) = A$. If f is differentiable at every point of Ω, then we say that f is differentiable in Ω. The map A is called the differential of f at x, and the corresponding matrix is called the Jacobian matrix of f.

Theorem 1.3. *If the above definition holds for $A = A_1$ and $A = A_2$ then $A_1 = A_2$.*

Proof. Let $B = A_1 - A_2$. Then

$$|Bh| \leq |f(x+h) - f(x) - A_1 h| + |f(x+h) - f(x) - A_2 h|.$$

Hence, by differentiability of f, we have $\frac{|Bh|}{|h|} \to 0$ as $h \to 0$. It is a straightforward exercise to verify that for a linear map B this implies that $B \equiv 0$. $\square$

Example 1.2. The Jacobian matrix of a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ coincides with the matrix that represents the map A, i.e, $A'(x) = A$ for any $x \in \mathbb{R}^n$. $\diamond$

Theorem 1.4 (The Chain Rule)**.** *Let $\Omega \subset \mathbb{R}^n$ be a domain, and $f : \Omega \to \mathbb{R}^m$ be a differentiable map at $a \in \Omega$. Suppose that $g : f(\Omega) \to \mathbb{R}^l$ be a map differentiable at $f(a)$. Then the map $F = g \circ f = g(f)$ is differentiable at a and*

$$F'(a) = g'(f(a)) \cdot f'(a).$$

Note that the product in the above formula is just the matrix multiplication of the Jacobian matrices g' and f'.

Proof. Let $b = f(a)$. We set $A = f'(a)$, $B = g'(b)$, $U(h) = f(a+h) - f(a) - Ah$, and $V(k) = g(b+k) - g(b) - Bk$, where $h \in \mathbb{R}^n$ and $k \in \mathbb{R}^m$. Then

$$\nu(h) = \frac{|U(h)|}{|h|} \to 0 \quad \text{as } h \to 0, \quad \mu(k) = \frac{|V(k)|}{|k|} \to 0, \quad \text{as } k \to 0. \tag{3}$$

Given a vector h we set $k = f(a+h) - f(a)$. Then

$$|k| = |Ah + U(h)| \leq (||A|| + \nu(h))\,|h|, \tag{4}$$

and

$$F(a+h) - F(a) - BAh = g(b+k) - g(b) - BAh = B(k - Ah) + V(k) = BU(h) + V(k).$$

Hence, (3) and (4) imply that for $h \neq 0$,

$$\frac{|F(a+h) - F(a) - BAh|}{|h|} \leq (||B||\nu(h) + (||A|| + \nu(h))\,\mu(k).$$

Letting $h \to 0$ we have $\nu(h) \to 0$, and $k \to 0$ by (4), so $\mu(k) \to 0$. From this it follows that $F'(a) = BA$ as required. $\square$

Example 1.3. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is a differentiable map at $a \in \mathbb{R}^n$ such that in a neighbourhood of $f(a)$ the map f^{-1} is defined and differentiable. Then the composition of $f^{-1} \circ f$ is a differentiable map whose differential at a by the Chain Rule equals

$$(f^{-1} \circ f)'(a) = (f^{-1})'(f(a)) \cdot f'(a).$$

On the other hand, the differential of the identity map is the identity, and we conclude that the matrix corresponding to $(f^{-1})'(f(a))$ is the inverse matrix to that of $f'(a)$. $\diamond$

Let $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, ... $e_n = (0, \ldots, 0, 1)$ be the standard basis in $\mathbb{R}^n$, and denote by $\{u_1, \ldots, u_m\}$ a similar basis in $\mathbb{R}^m$. For a domain $\Omega \subset \mathbb{R}^n$ the map $f : \Omega \to \mathbb{R}^m$ can be written in the form

$$f(x) = \sum_{i=1}^{n} f_i(x) u_i = (f_1(x), \ldots, f_m(x)), \tag{5}$$

where each $f_i : \Omega \to \mathbb{R}$ is a function. For a function $f : \Omega \to \mathbb{R}$ the limit

$$(D_j f)(x) = D_{x_j} f(x) = f_{x_j}(x) = \frac{\partial f}{\partial x_j} = \lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t},$$

if exists, is called a *partial* derivative with respect to variable x_j. Unlike functions of one variable, existence of partial derivatives does not imply in general that a function is differentiable, for example the function $f(x, y)$ in Example 1.1 has partial derivatives everywhere with respect to variables x and y, but is not even continuous at the origin. Examples of continuous functions that have partial derivatives but are not differentiable also exist.

Applying partial derivatives with respect to variable x_j to components f_i of a map $f : \Omega \to \mathbb{R}^m$ we obtain a matrix $(\frac{\partial f_i}{\partial x_j})$. As it turns out, if f is differentiable at a point $x \in \Omega$, then all partial derivatives exist.

Theorem 1.5. *Suppose $f : \Omega \to \mathbb{R}^m$ is differentiable at $x \in \Omega$. Then $(D_j f_i)(x)$ exist for all i, j and*

$$f'(x)e_j = \sum_{i=1}^{m} (D_j f_i)(x) u_i = \left(\frac{\partial f_1}{\partial x_j}, \frac{\partial f_2}{\partial x_j}, \dots, \frac{\partial f_m}{\partial x_j} \right).$$

Proof. Fix j. Since f is differentiable at x, $f(x+te_j) - f(x) = f'(x)(te_j) + r(te_j)$, where $r(te_j) = o(t)$. Then by the linearity of $f'(x)$,

$$\lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t} = f'(x)e_j.$$

If now f is represented in terms of components as in (5), we have

$$\lim_{t \to 0} \sum_{i=1}^{m} \frac{f_i(x + te_j) - f_i(x)}{t} u_i = f'(x)e_j.$$

Thus, each coefficient in front of u_i has a limit, which shows existence of the partial derivatives of f. □

It follows from the above theorem that the Jacobian matrix $f'(x)$ is given by

$$f'(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \cdots & & \cdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Example 1.4. Let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be a differentiable map, $\gamma = (\gamma_1, \dots, \gamma_n)$. Its image in $\mathbb{R}^n$ is a called a (parametrized) smooth curve. Its differential is a column or an $n \times 1$ matrix of the form

$$D\gamma = \left(\frac{d\gamma_1}{dt}, \dots, \frac{d\gamma_n}{dt} \right)^T.$$

Note that we used the usual sign of derivative because each component of γ is a function of one variable t. (T indicates transposition of a matrix.) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Then its differential is the $1 \times n$ matrix

$$Df = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right).$$

The composition function $g = f \circ \gamma$ is a usual function of one variable. By the Chain Rule its derivative can be computed as

$$\frac{dg}{dt}(t) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right) \cdot \left(\frac{d\gamma_1}{dt}, \dots, \frac{d\gamma_n}{dt}\right)^T = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\gamma(t)) \frac{d\gamma_i}{dt}(t). \tag{6}$$

$\diamond$

The above example has an important generalization. For a differentiable function f, define ∇f, the gradient of f, to be the vector given by

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right) = \sum_{i=1}^{n} (D_i f)(x)\, e_i.$$

Then (6) can be written in the form $g'(t) = \nabla f(\gamma(t)) \cdot \gamma'(t)$, where the dot indicates the dot product in $\mathbb{R}^n$. Let u now be a unit vector, and let $\gamma(t) = x + tu$ be the line in the direction of u. Then $\gamma'(t) = u$ for all t, and so $g'(0) = (\nabla f(x))u$. On the other hand, $g(t) - g(0) = f(x + tu) - f(x)$, hence,

$$\lim_{t\to 0} \frac{f(x+tu) - f(x)}{t} = \nabla f(x) \cdot u.$$

This is called the directional derivative of f at x in the direction of vector u, denoted sometimes by $D_u f(x)$ or $\frac{\partial f}{\partial u}$. For a fixed f and x it is clear that the directional derivative attains its maximum if u is a positive multiple of ∇f. So ∇f gives the direction of the fastest growth of the function f.

Theorem 1.6. *If a function $f : \Omega \to \mathbb{R}$ has continuous partial derivatives $\frac{\partial f}{\partial x_j}$ at a point a for $j = 1, \dots, n$, then f is differentiable at a.*

Proof. For simplicity of notation we assume that $\Omega \subset \mathbb{R}^2$, the proof in the general case is the same. We need to show that there exists a linear map $A : \mathbb{R}^2 \to \mathbb{R}$ such that (2) holds. The clear choice for A is the $(D_1 f, D_2 f)$. Let $a = (a_1, a_2)$. For a fixed $h = (h_1, h_2)$ we have

$$\Delta f = f(a+h) - f(a) = [f(a+h) - f(a_1, a_2 + h_2)] + [f(a_1, a_2 + h_2) - f(a)].$$

We apply the Mean Value theorem to two expressions on the right to obtain

$$\Delta f = h_1 \frac{\partial f}{\partial x_1}(a_1 + \theta_1 h_1, a_2 + h_2) + h_2 \frac{\partial f}{\partial x_2}(a_1, a_2 + \theta_2 h_2)$$

for some numbers $\theta_i \in (0, 1)$. Hence,

$$\Delta f = h_1 \frac{\partial f}{\partial x_1}(a) + h_2 \frac{\partial f}{\partial x_2}(a) + \varepsilon(h),$$

where

$$\varepsilon(h) = h_1 \left(\frac{\partial f}{\partial x_1}(a_1 + \theta_1 h_1, a_2 + h_2) - \frac{\partial f}{\partial x_1}(a)\right) + h_2 \left(\frac{\partial f}{\partial x_2}(a_1, a_2 + \theta_2 h_2) - \frac{\partial f}{\partial x_2}(a)\right).$$

By continuity of partial derivatives we obtain

$$\frac{|\Delta f - D_1 f(a) h_1 - D_2 f(a) h_2|}{|h|} = \frac{|\varepsilon(h)|}{|h|} \to 0, \quad \text{as } h \to 0,$$

which is the required statement. $\square$

Definition 1.7. *A map $f : \Omega \to \mathbb{R}^m$ is called continuously differentiable, or of class $C^1(\Omega)$, if $f'(x)$ is a continuous function on Ω, i.e., for every $\varepsilon > 0$ there exists $\delta > 0$ such that $||f'(y) - f'(x)|| < \varepsilon$ whenever $|x - y| < \delta$.*

Theorem 1.8. *$f \in C^1(\Omega)$ if and only if all partial derivatives exist and continuous on Ω.*

Proof. Suppose that $f \in C^1(\Omega)$. Then

$$(D_j f_i)(x) = [f'(x)e_j]u_i = \left[\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \dots & & \dots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} e_j\right] \cdot u_j,$$

for all i, j and $x \in \Omega$ (we continue to use the notation $\{u_j\}$ for the standard basis in the target domain). Then

$$(D_j f_i)(y) - (D_j f_i)(x) = \left[(f'(y) - f'(x))e_j\right] u_i$$

Since $|u_i| = |e_j| = 1$, we have

$$|(D_j f_i(y) - (D_j f_i)(x)| \leq ||f'(y) - f'(x)||,$$

which shows continuity of $D_j f_i$.

For the proof of the theorem in the other direction it suffices to consider the case $m = 1$, i.e., when f is a function. But this is exactly the content of Theorem 1.6. $\square$

Suppose that $f : \Omega \to \mathbb{R}$ with partial derivatives $D_1 f, \dots, D_n f$. If the functions $D_j f$ are themselves differentiable, then the second-order partial derivatives of f are defined by

$$D_{ij} f = D_i D_j f = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (i, j = 1, \dots, n).$$

If all functions $D_{ij} f$ are continuous on Ω, we say that f is of class $C^2(\Omega)$ and write $f \in C^2(\Omega)$. In the same way we define derivatives of any order and functions of class $C^k(\Omega)$, $k = 1, 2, \dots, \infty$. A map $f \in C^k(\Omega)$ if every component of f is of class $C^k(\Omega)$.

It may happen that $D_{ij} f \neq D_{ji} f$ at some point where both derivatives exist. But if both derivatives are continuous, the following holds.

Theorem 1.9. *Suppose f is defined in $\Omega \subset \mathbb{R}^n$, and $D_1 f$, $D_{21} f$ and $D_2 f$ exist at every point of Ω, and $D_{21} f, D_{12} f$ are continuous at some point $a \in E$. Then*

$$(D_{12} f)(a) = (D_{21} f)(a).$$

In particular $D_{12} f = D_{21} f$ for $f \in C^2(\Omega)$.

Proof. For simplicity assume that $n = 2$, as the proof for a general n is the same. Let $a = (a_1, a_2)$. Consider the expression

$$\Delta = \frac{f(a_1 + h, a_2 + k) - f(a_1 + h, a_2) - f(a_1, a_2 + k) + f(a)}{hk},$$

where h, k are nonzero, say positive, and sufficiently small. The auxiliary function

$$\phi(x_1) = \frac{f(x_1, a_2 + k) - f(x_1, a_2)}{k}$$

is, by the assumptions of the theorem, differentiable on the interval $[a_1, a_1 + h]$ with

$$\phi'(x_1) = \frac{D_1 f(x_1, a_2 + k) - D_1 f(x_1, a_2)}{k},$$

in particular, it is continuous. Then Δ can be written in the form

$$\Delta = \frac{\phi(a_1 + h) - \phi(a_1)}{h}.$$

By the Mean Value theorem applied to $\phi(x_1)$ on $[a_1, a_1+h]$ we get for $0<\theta<1$,

$$\Delta = \phi'(a_1+\theta h) = \frac{D_1 f(a_1+\theta h, a_2+k) - D_1 f(a_1+\theta h, a_2)}{k}.$$

Since $D_{12}f$ exists, we apply the Mean Value theorem to $D_1 f(a_1+\theta h, x_2)$ on the interval $[a_2, a_2+k]$ to get for $0<\theta_1<1$,

$$\Delta = D_{21} f(a_1+\theta h, a_2+\theta_1 k). \tag{7}$$

By symmetry in Δ we can interchange the variables in the auxiliary function by considering

$$\psi(x_2) = \frac{f(a_1+h, x_2) - f(a_1, x_2)}{h},$$

and obtain analogously that for $0<\theta_2, \theta_3<1$

$$\Delta = D_{12} f(a_1+\theta_2 h, a_2+\theta_3 k). \tag{8}$$

Comparing (7) and (8) we conclude that

$$D_{21} f(a_1+\theta h, a_2+\theta_1 k) = D_{12} f(a_1+\theta_2 h, a_2+\theta_3 k).$$

By letting $h, k \to 0$, and from continuity of the second order derivatives we conclude the result. $\square$

Definition similar to (1.2) also works for general Banach spaces of arbitrary dimension. We say that a map $f : V \to W$ between two Banach spaces is called differentiable at a point $a \in V$, if there exists a continuous linear map (operator) $A := Df(a) : V \to W$ such that

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

Note that the requirement is that the map A is linear and *continuous* which is essential for infinite dimensional space. Sometimes Df is called the Fréchet derivative of the map f. Proofs similar to those of this section show that Theorem 1.3 and the Chain Rule also hold in this general setting. If the derivative of f exists at every point of V, then Df becomes the map

$$Df : V \to B(V, W); \quad x \to Df(x).$$

Here $B(V, W)$ denotes the space of continuous linear operators from V to W, a Banach space itself. The map f is called continuously differentiable if Df is continuous. Note that this is not the same as the map $Df(x)$ being continuous for every x, which is part of the definition of differentiability of f. From here higher order derivatives are defined inductively.