Summary of Real Analysis 2 – Semester 2 (2009-10)

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

An Important Note: I do not have a hand-written manuscript when I type these notes. Also, I rarely proof-read. Please bring typos/mistakes to my notice.

Reference Books:

1. M. Spivak, *Calculus on Manifolds*
2. J. Munkres, *Analysis on Manifolds*,
3. Tom Apostol, *Mathematical Analysis*, Narosa Publishing House.
4. W. Rudin, *Principles of Mathematical Analysis*, Wiley International.

Preliminaries

1. Linear Algebra is the foundation of Differential Calculus. We very briefly review some basic concepts.

2. Let V and W be (finite dimensional) real vector spaces. Let $\{v_i : 1 \leq i \leq m\}$ (resp. $\{w_j : 1 \leq j \leq n\}$) be a basis of V (resp. of W). Let $T\colon V \to W$ be a linear map. Then the matrix A of T w.r.t. to these bases is an $n \times m$-matrix (a_{ij}) where the i-th column is the coefficients (a_{ij}) where $Tv_i = a_{1i}w_1 + a_{2v_i}w_2 + \cdots + a_{ni}w_n = \sum_{j=1}^n a_{ji}w_j$. Note the way the coefficients are indexed.

3. If $A = (a_{ij})$ is a real $n \times m$ matrix, then we have an associated linear map $T\colon \mathbb{R}^m \to \mathbb{R}^n$ given by $T\colon x \mapsto Ax$ where $x = (x_1, \dots, x_m)^t \in \mathbb{R}^m$ and Ax is the product of an $n \times m$ matrix A with the $m \times 1$ matrix x. The matrix of T w.r.t. the standard bases of $\mathbb{R}^m$ and $\mathbb{R}^n$ is A. Note that Ae_i is the i-th column of A.

   As a general rule, we write vectors in $\mathbb{R}^n$ as a column vector (i.e., a matrix of size $n \times 1$). A 1×1 real matrix is identified with real number which is its unique entry.

4. As examples, we wrote down the matrices of the linear maps:

   (a) $T\colon \mathbb{R}^2 \to \mathbb{R}^4$ given by $T(x, y) = (2x + 3y, x + y, x - y, y)$.
   (b) $T\colon \mathbb{R}^n \to \mathbb{R}$ given by $T(x_1, \dots, x_n) := \sum_i a_i x_i$.

5. Any linear map $T\colon \mathbb{R} \to \mathbb{R}$ is of the form $Tx = cx$ where $c = T(1)$.

6. Any linear map $T\colon \mathbb{R}^n \to \mathbb{R}$ is of the form $Tx = \sum_{i=1}^n c_i x_i$ where $c_i = T(e_i)$. *Hint:* Write $x = \sum_i x_i e_i$ and apply T to both sides.

7. We recalled the definition of an inner product on a **real** vector space. As an example, we looked at $\mathbb{R}^n$ with the standard inner product $(x, y) \mapsto \sum_{i=1}^n x_i y_i$. It is also known as the Euclidean inner product or the standard dot product. Note that using our convention in Item 3, the dot product can be written as $x \cdot y = y^t x$, the product of matrices of type $1 \times n$ and $n \times 1$.

8. The result in Item 6 can be reformulated as follows: Any linear map $T\colon \mathbb{R}^n \to \mathbb{R}$ is of the form $T(x) = x \cdot c$ where the vector $c = (T(e_1), \ldots, T(e_n))$.

9. If $(V, \langle\ ,\ \rangle)$ is an inner product space, then $\|x\| := \langle x, x\rangle^{1/2}$.

10. The most important inequality is the Cauchy-Schwarz inequality:

$$|\langle v, w\rangle| \leq \|v\|\,\|w\| \text{ for all } v, w \in V. \tag{1}$$

The equality holds iff one of the vectors is a scalar multiple of the other.

11. The norm $\|\ \| : V := \mathbb{R}^n \to \mathbb{R}$ has the following properties:

   (a) $\|x\| \geq 0$ and $\|x\| = 0$ iff $x = 0$.
   (b) $\|tx\| = |t|\,\|x\|$ for all $t \in \mathbb{R}$ and $x \in V$.
   (c) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.

12. Any function $\|\ \| : V \to \mathbb{R}$ from any real vector space to $\mathbb{R}$ satisfying the properties listed in Item 11 is called a norm on V.

13. We showed the following are norms on $\mathbb{R}^n$:

   (a) $x = (x_1, \ldots, x_n) \mapsto \sqrt{(x_1^2 + \cdots + x_n^2)}$. This is called the standard or Euclidean norm.
   (b) $x \mapsto \sum_{i=1}^n |x_i|$. This is called the L^1-norm and is denoted by $\|x\|_1$.
   (c) $x \mapsto \max\{|x_i| : 1 \leq i \leq n\}$. This is called the max norm or L^∞-norm. It is denoted by $\|x\|_\infty$.

14. A norm on a vector space V gives rise to a metric on V as follows: $d(x, y) := \|x - y\|$. We checked that d is a metric on X.

   The metrics on $\mathbb{R}^n$ induced by the norms the standard $\|\ \|$, $\|\ \|_1$ and $\|\ \|_\infty$ will be denoted by d, d_1 and d_∞.

   Whenever we talk of distances in $\mathbb{R}^n$, it will be with reference to the standard/Euclidean metric d.

Items 1–14 were done on 22 December 2009 (14:25 — 16:00).

15. Let V be a finite dimensional real vector space. Let $\{v_i : 1 \leq i \leq m\}$ be an orthonormal (O.N.) basis of V. Then any $v = \sum_{i=1}^m a_i v_i$, where $a_k = \langle v, v_k\rangle$. *Hint:* Take inner product of both sides with v_k.

16. Keep the notation of the last item. Let W be another inner product space with $\{w_j : 1 \leq n\}$ as an O.N. basis.Let $T\colon V \to W$ be a linear map. Then the matrix $A = (a_{ij}$ of T w.r.t. to these O.N. bases can be explicitly written. We have $a_{rs} = \langle Tv_s, w_r\rangle$. *Hint:* Let $Tv_i = \sum_{j=1}^n a_{ji}w_j$. Take inner product of both sides with the vector w_r.

Items 15–16 were done on 18 December 2009 (14:00 — 15:00).

17. We defined open balls in a metric space (X, d). The open ball with centre $a \in X$ and radius $r > 0$ is defined as $B(a, r) := \{x \in X : d(x, a) < r\}$. We looked at $B((a, b), r)$ in $\mathbb{R}^2$.

18. We defined a sequence in a metric space (X, d). A sequence (x_n) in (X, d) converges to $x \in X$ if for each $\varepsilon > 0$, we can find $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have $d(x_{,n}, x) < \varepsilon$, that is, $x_n \in B(x, \varepsilon)$.

19. We proved the uniqueness of the limit of convergent sequences in a metric space.

20. We looked at some examples:

    (a) In any metric space, a constant sequence is always convergent.

    (b) In $\mathbb{R}^n$, we saw that a sequence $\vec{x}_k := (x_{k1}, \dots, x_{kn})$ is convergent to a vector $x = (x_1, \dots, x_n)$ in any of the metrics d, d_1 or d_∞ iff for each $1 \leq i \leq n$, the sequence (x_{ki}) of real numbers converge to x_i. Thus, $\vec{x_k} \to \vec{x}$ iff the sequence "converges coordinate-wise".

    The reason for this fact about convergence is the 'equivalence of norms':

$$\frac{1}{n}\|x\|_1 \leq \frac{1}{\sqrt{n}}\|x\|_2 \leq \|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 .$$

    (c) We looked at some examples of convergent sequences and also some examples of non-convergent sequences in $\mathbb{R}^2$.

    (d) If we fix a vector $a = (a_1, \dots, a_n) \in \mathbb{R}^n$. Then the sequence $x_n := (1/n)a$ converges to the zero vector.

21. Let (x_k), (y_k) be sequences in $\mathbb{R}^n$ converging to x and y respectively. Let $c \in \mathbb{R}$. We proved the following:

    (a) The sequence (z_k) where $z_k = x_k + y_k$ of vectors in $\mathbb{R}^n$ converges to $x + y$ (w.r.t. any one of the metrics d, d_1 and d_∞).

    (b) The sequence (cx_k) converges to cx.

    (c) The sequence $(x_k \cdot y_k)$ of real numbers converges to $x \cdot y$.

22. We say that that a function $f\colon (X, d) \to (Y, d)$ is continuous at a point $a \in X$ if for *every* sequence (x_n) in X converging to a, the sequence $(f(x_n)$ converges to $f(a)$ in Y. The function f is said to be continuous on X if f is continuous at each point of X.

23. We looked at some examples of continuous functions.

    (a) Fix $y_0 \in Y$. The the constant function $f(x) = y_0$ for $x \in X$ is continuous.

    (b) The identity function $f(x) = x$ is a continuous function from (X, d) to itself.

(c) Given a metric space (X,d) (having at least two points), there exist non-constant real valued continuous functions. For example, fix $a \in X$ and look at $f(x) := d(x,a)$.

(d) Given a linear map $T\colon \mathbb{R}^m \to \mathbb{R}^n$ there exist a constant $C > 0$ such that $\|Tx\| \leq C\|x\|$ for $x \in \mathbb{R}^m$. From this it followed that T is *uniformly continuous.*

In fact, the proof showed the following: any linear map from a finite dimensional eulidean space $\mathbb{R}^m$ to any vector space equipped with a norm is uniformly continuous.

Items 17–23 were done on 23 December 2009 (14:30 — 16:00).

24. We recalled the definition of uniform continuity of a function $f\colon (X,d) \to (Y,d)$. Given a continuous function $f\colon [a,b] \to \mathbb{R}$, while proving the Riemann integrability of f on $[a,b]$, one needs the fact such an f is uniformly continuous on $[a,b]$.

25. We proved that any linear map $T\colon \mathbb{R}^n \to V$ to any normed linear space (NLS, that is, a vector space equipped with a norm) is uniformly continuous.

26. The vector addition $(x,y) \mapsto x+y$ from $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is continuous.

27. The dot product $(x,y) \mapsto x \cdot y$ from $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is continuous.

28. We showed that if $f\colon (X,d) \to (Y,d)$ is continuous at $a \in X$ and if $g\colon (Y,d) \to (Z,d)$ is continuous at $b := f(a)$, then the composite function $g \circ f$ is continuous at a.

29. Applications of the last item.

    (a) The functions $x = (x_1, \dots, x_n) \mapsto x_i$ from $\mathbb{R}^n$ to $\mathbb{R}$ is continuous. Hence any polynomial functions in the variables $x_1, \dots, x_n$ is continuous on $\mathbb{R}^n$.

    (b) Let (X,d) be a metric space. Let $f,g\colon X \to \mathbb{R}^n$ be continuous. Then the map $f+g\colon X \to \mathbb{R}^n$ defined by $(f+g)(x) := f(x)+g(x)$ is continuos. It is the composite of the maps $x \mapsto (f(x),g(x))$ and $(u,v) \mapsto u+v$.

    (c) With the notaion of the last item, the map $h(x) := f(x) \cdot g(x)$ from X to $\mathbb{R}$ is continuous. It is the composite of the maps $x \mapsto (f(x),g(x))$ and $(u,v) \mapsto u \cdot v$.

    (d) Two special cases of the last item: if f,g are real valued continuous functions on a metric space (X,d), then their sum $f+g$ and the product fg are continuous.

    Hence the set $C(X,\mathbb{R})$ of all real valued continuous functions on a metric space X is a real vector space and is a ring. (It is an algebra, if you know what this means.)

    (e) Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be linear. The map $(x,y) \mapsto Tx \cdot y$ from $\mathbb{R}^m \times \mathbb{R}^n$ to $\mathbb{R}$ is continuous. It is the composite of the maps $(x,y) \mapsto (Tx,y)$ and $(u,v) \mapsto u \cdot v$.

    (f) The map $x \mapsto \|x\|$ from an NLS to $\mathbb{R}$ is continuous. (We used the triangle inequality: $|\,\|x\| - \|y\|\,| \leq \|x-y\|$.) Also the map $x \mapsto 1/\|x\|$ is continuous from the set of nonzero vectors to $\mathbb{R}$. It is the composite of the map $x \mapsto \|x\|$ and $t \mapsto 1/t$.

    (g) If $f\colon (X,d) \mapsto \mathbb{R}^*$ is a continuous function, then the function $g\colon X \to \mathbb{R}^*$ defined by $g(x) := 1/f(x)$ is continuous.

30. Let $A \subset (X, d)$. A point $p \in X$ is said to be a cluster point of A if, for any $r > 0$, the intersection $B(p, r) \cap A$ contains a point other than p. The subset $\mathbb{Z} \subset \mathbb{R}$ has no cluster point. Any real number is a cluster point of the set $\mathbb{Q} \subset \mathbb{R}$.

31. Let (X, d) be a metric space. Let $a \in X$ be a cluster point of X. Let $f : X \setminus \{a\} \to (Y, d)$ be a function. (Note that a need not be in the domain of f, that is, $f(a)$ may not be defined.) Then we say that $b \in Y$ is the limit of f as $x \to a$ in X if for any given $\varepsilon > 0$, there exists $\delta > 0$ such that for any $x \in B(a, \delta) \cap E$ with $x \neq a$, we have $f(x) \in B(b, \varepsilon)$, that is, $d(f(x), b) < \varepsilon$. Such a b, if exists, is unique (here we needed the fact that a is a cluster point of X) and is denoted by $\lim_{x \to a} f(x) = b$.

32. Let $f : (X, d) \to (Y, d)$ be a function. Assume that $a \in X$ be a cluster point of X. Then f is continuous at a iff (i) $\lim_{x \to a} f(x)$ exists, say, $b \in Y$ and (ii) $b = f(a)$.

33. Let d be the absolute value metric on $\mathbb{Z} \subset \mathbb{R}$. Any convergent sequence in $(\mathbb{Z}, d)$ is eventually constant. As a consequence, any function $f : (\mathbb{Z}, d) \to (Y, d)$ is continuous.

34. Let $M_{n \times m}(\mathbb{R}) \cong M_{n \times m}$ denote the vector space of all real matrices of size $n \times m$. We let $M(n, \mathbb{R}) := M_{n \times n}(\mathbb{R})$. The map

$$A = (a_{ij}) \mapsto (a_{11}, \ldots, a_{1m}, a_{21}, \ldots, a_{2m}, \ldots, a_{n1}, \ldots, a_{nm})$$

is a linear isomorphism of $M_{n \times m}$ with $\mathbb{R}^{nm}$. This isomorphism induces at least three natural norms on $M_{n \times m}$:

$$\|A\| := \left(\sum_{ij} |a_{ij}|^2 \right)^{1/2}, \|A\| := \max_{i,j} \{|a_{ij}|\}, \|A\| := \sum_{i,j} |a_{ij}| .$$

35. The matrix product $(A, B) \mapsto AB$ from $M_{n \times k} \times M_{k \times m} \to M_{n \times m}$ is continuous. In particular, the map $A \mapsto A^2$ from $M(n, \mathbb{R})$ to itself is continuous.

36. The map $A \mapsto \det(A)$ from $M(2, \mathbb{R})$ to $\mathbb{R}$ is continuous. In fact, the map $A \mapsto \det A$ is continuous from $M(n, \mathbb{R})$ to $\mathbb{R}$. We used the Laplace expansion:

$$\det A = \sum_{\sigma \in S_n} \text{sign}\,(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

Observe that $\det : M(n, \mathbb{R}) \to \mathbb{R}$ is a ploynomial function in the variables a_{ij}.

37. A function $f : (X, d) \to (Y, d)$ is continuous at $a \in X$ iff for any $r > 0$, the inverse image $f^{-1}(B(f(a), r))$ is open in X. It is continuous on X iff for every open set $V \subset Y$, the inverse image $f^{-1}(V)$ is open in X. (The proof was learnt in Real Analysis 1.)

38. The set $GL(n, \mathbb{R})$ of all invertible matrices is an open subset of $M(n, \mathbb{R})$. Hence the set of singular matrices of size $n \times n$ is closed in $M(n, \mathbb{R})$.

Items 24–38 were done on 29 December 2009 (14:30 — 16:00).

39. Given $f : \mathbb{R}^m \to \mathbb{R}^n$, we can write it as $f(x) = (f_1(x), \ldots, f_n(x))$ where f_i is the composite $\pi_i \circ f$. We proved that f is continuous iff each f_i is continuous. (You were asked to write the proof after I explained the proof and we 'jointly corrected' your writing!)

40. Given a linear map $A\colon \mathbb{R}^m \to \mathbb{R}^n$ or an $n \times m$ matrix A we define the operator norm $\|A\|$ as follows:

$$\|A\| := \inf\{C : C \geq 0 \text{ and } \|Ax\| \leq C\,\|x\| \text{ for all } x \in \mathbb{R}^m\}.$$

Note that it makes sense, since the subset of real numbers on the right side is non-empty and bounded below by 0. Before we prove that this is a norm on the vector space $L(\mathbb{R}^m, \mathbb{R}^n)$ or on $M_{n\times m}$, we looked at some examples.

41. Operator norms of some linear maps or matrices:

 (a) $\|0\| = 0$.
 (b) $\|I_{n\times n}\| = 1$.
 (c) $\|cI_{n\times n}\| = |c|$.
 (d) If $A = \operatorname{diag}(c_1, \ldots, c_n)$ is the diagonal matrix, then $\|A\| = \max\{|c_i| : 1 \leq i \leq n\}$.
 (e) We shall prove later that if A is a real symmetric matrix, then $\|A\|$ is the maximum of the absolute values of the eigenvalues of A. Note that the last item is a special case of this result.

42. We proved that the operator norm is indeed a norm on $L(\mathbb{R}^m, \mathbb{R}^n)$ and on $M_{n\times m}$. Unless otherwise specified, whenever we talk of norm of a linear map or of a matrix, it will refer to the operator norm.

43. The operator norm has an important property: $\|B \circ A\| \leq \|B\|\,\|A\|$, where $A\colon \mathbb{R}^m \to \mathbb{R}^n$ and $B\colon \mathbb{R}^n \to \mathbb{R}^k$ are linear or B is of type $k \times n$ and A is of type $n \times m$.

 Strict inequality may occur: Consider a nonzero linear map A such that $A^2 = 0$.

Items 39–43 were done on 30 December 2009 (14:30 — 15:50).

44. We claim that a sequence (A_k) in $M(n, \mathbb{R})$ converges to A in the operator norm iff it converges 'entry-wise' or coordinate-wise. To see this, we show that $\|A\| \leq \sqrt{n}\,\|A\|_{\max}$:

$$\|Ax\|^2 = \sum_i \left(\sum_j a_{ij}x_j\right)^2 \leq \sum_i \left(\sum_j |a_{ij}||x_j|\right)^2 \leq \sum_i \|A\|_{\max}^2 \|x\|^2 = n\,\|A\|_{\max}^2 \|x\|^2.$$

 This inequality shows that if $A_k \to A$ coordinate-wise, then it converges in max norm and hence in operator norm also.

 To see the other way, note that $\|A_k e_i - Ae_i\| \leq \|A_k - A\|$ so that $A_k e_i \to Ae_i$. In particular, their entries also converge.

45. Consider the map $A \mapsto A^{-1}$ from $GL(2, \mathbb{R})$ to itself. Since

$$A^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

 it is clear that the inversion map is continuous. The same result is true for $GL(n, \mathbb{R})$. One uses the formula for A^{-1} in terms of co-factors of A.

Differentiation

46. Given the sequences $(1/n)$, $(1/n^2)$ and $(1/2^n)$, you guessed which goes (converges) to zero fastest. This notion was formulated precisely as follows: given two sequences (x_n) and (y_n) both converging to 0, we say (x_n) goes to zero much faster than (y_n) if $x_n/y_n \to 0$.

47. This led us to the following: if $f, g \colon (X, d) \to \mathbb{R}$ are such that $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, then we say that f goes to zero much faster than g as $x \to a$ if $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$.

48. In several variable calculus, when we talk of differentiability, the domain of the function is always assumed to be an open set in some $\mathbb{R}^n$. We shall see later the reason for this. (Compare this with the uniqueness part in Item 31.

49. Let $U \subset \mathbb{R}^m$ be open, $a \in U$. Let $f \colon U \to \mathbb{R}$ be given. Given $x \in U$, we may write $x = a + h$. Then h is called the increment in the independent variable and $f(x) - f(a) \equiv f(a+h) - f(a)$ is called the increment in the dependent variable. We say that f is differentiable at a if we can control the increment in the dependent variable by means of a linear map $A \colon \mathbb{R}^m \to \mathbb{R}$: $f(a+h) - f(a) \approx Ah$, read as $f(a+h) - f(a)$ is approximately equal to Ah. Note that this is same as saying that for x near a, the value $f(x)$ is *approximately equal to* $f(a) + A(x - a)$. Whenever we approximate like this, we need to have a control on the error we are making. The error is $E(h) := f(a+h) - f(a) - Ah$. An obvious first requirement is $E(h) \to 0$ as $h \to 0$.

    Let us look at an example. Consider $f \colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. Let $a \in \mathbb{R}$. Then we observe

    $$f(a+h) - f(a) = 2ah + h^2.$$

    Since we know all the linear maps from $\mathbb{R}$ to itself, we defined $Ah = 2ah$ so that $E(h) = h^2$ which goes to zero faster than h going to 0.

    Going back to the general case, we require that $E(h)$ goes to 0 much faster than h, that is, $\lim_{h\to 0} E(h)/\|h\| = 0$.

50. The discussion in the last item led us to the following definition. Let $f \colon U \subset \mathbb{R}^m \to \mathbb{R}^n$. Let $a \in U$. We say that f is differentiable at a if there exists a linear map $A \colon \mathbb{R}^m \to \mathbb{R}^n$ such that for $x \in U$, if we write $f(x) = f(a) + A(x-a) + E(x-a)$ then $\lim_{x\to a} \frac{\|E(x-a)\|}{\|x-a\|} = 0$.

    Such a linear map, if it exists, is unique (to be proved later). It is denoted by $Df(a)$ and called the (total or Frechet) derivative of f at a.

51. Examples:

    (a) Consider $f \colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^n$ where $n \in \mathbb{N}$. Key observation:

    $$\begin{aligned} f(a+h) - f(a) = (a+h)^n - a^n &= \binom{n}{1} a^{n-1}h + \binom{n}{2} a^{n-2}h^2 + \cdots + \binom{n}{n} a^0 h^n \\ &\leq na^{n-1}h + h^2(\text{Constant}), \end{aligned}$$

    where we assumed that $|h| \leq 1$. Hence $Df(a)h = na^{n-1}h$.

(b) $f(x) = e^x$ for $x \in \mathbb{R}$. Key observation:

$$\begin{aligned} e^{a+h} - e^a &= e^a(e^h - 1) \\ &= e^a([1 + \frac{h}{1!} + \frac{h^2}{2!} + \cdots] - 1) \\ &\leq e^a h + e^a h^2(\text{Constant}). \end{aligned}$$

Hence $Df(a)(h) = e^a h$.

(c) Consider $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = xy$. Key observation:

$$f(a+h, b+k) - f(a,b) = ak + bh + hk.$$

Note that $\frac{|hk|}{\|(h,k)\|} \leq \frac{\|(h,k)\|\|(h,k)\|}{\|(h,k)\|}$ and we have $Df(a,b)(h,k) := ak + bh$.

(d) Consider $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = x + y$. You found that $Df(a,b)(h,k) = h + k$.

(e) Let $f\colon \mathbb{R}^m \to \mathbb{R}^n$ be linear. Then $Df(a)(h) = f(h)$, that is, $Df(a) = f$. Note that the last example is a special case of this result.

(f) Consider $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by $f(x,y)x \cdot y$. Key Observation:

$$f(a+h, b+k) - f(a,b) = a \cdot k + h \cdot b + h \cdot k.$$

Hence $DF(a,b)(h,k) = a \cdot k + b \cdot h$ as $\frac{|h \cdot k|}{\|(h,k)\|} \leq \frac{\|h\|\|k\|}{\|(h,k)\|} \leq \frac{\|(h,k)\|^2}{\|(h,k)\|}$.

(g) Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be linear. Consider $f\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ defined by $f(x,y) = Tx \cdot y$. Then we have $Df(a,b) = Ta \cdot k + Th \cdot b$.

52. We now formualte the definition of differentiability at a point in terms of ε-δ. If we let $\varphi(h) := \frac{\|E(h)\|}{\|h\|}$, then $\lim_{h \to 0} \frac{\|E(h)\|}{\|h\|} = 0$ is same as saying that $\lim_{h \to 0} \varphi(h) = 0$. That is, given $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $0 \neq h \in B(0, \delta)$, we must have $\varphi(h) \in (-\varepsilon, \varepsilon)$. Since $\varphi(h) \geq 0$, this is same as requiring that $\varphi(h) < \varepsilon$. This leads us to the following equivalent defintion:

f is differentiable at a iff there exists a linear map $A\colon \mathbb{R}^m \to \mathbb{R}^n$ such that for each ε there exists $\delta > 0$ such that for $0 < \|h\| < \delta$, we must have $\|f(a+h) - f(a) - Ah\| < \varepsilon \|h\|$.

Items 45–52 were done on 31 December 2009 (14:30 — 16:05). Happy New Year!

53. Exercise:

(a) Consider $f(x,y) := x^2 y$ for $(x,y) \in \mathbb{R}^2$. Then f is differentiable at (a,b) with $Df(a,b)(h,k) = 2abh + a^2 k$.

(b) Consider $f(x,y) := (x, y, xy)$ from $\mathbb{R}^2$ to $\mathbb{R}^3$. Then $Df(a,b)(h,k) = (h, k, ak + bh)$.

(c) Consider $f\colon \mathbb{R}^m \to \mathbb{R}$ defined by $f(x) = x \cdot x$. Then f is differentiable at a and $Df(a)(h) = 2a \cdot h$.

(d) Let A be an $n \times n$ matrix. Let $f(x) := Ax \cdot x$ for (column) vectors $x \in \mathbb{R}^n$. Then $Df(a)(h) = Aa \cdot h + Ah \cdot a$. In particular, if A is symmetric, then $Df(a)(h) = 2Aa \cdot h$. The last item is a special case of this result.

(e) Let $f\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^k$ be a bilinear map. Then one finds $Df(a,b)(h,k) = f(a,k) + f(h,b)$. Note that Items 51c, 51f and 51g are special cases of this result.

(f) Consider $f(X) := X^2$ for $X \in M(n, \mathbb{R})$. Then f is differentiable at $A \in M(n, \mathbb{R})$ with $Df(A)(H) = AH + HA$.

You solved (a), (b) and (e) in the class.

54. We explained why we insisted on the domain being open.

55. We proved the uniqueness of the linear map in the definition of of differentiability. We saw how the set U being open was needed in the proof.

56. Let $J \subset \mathbb{R}$ be an open interval and $a \in J$. Then $f\colon J \to \mathbb{R}$ is differentiable at $a \in J$ iff the limit $\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$ exists. If f is differentiable, then we have $Df(a)(1) = f'(a)$.

    (For a complete proof of the last two items, refer to my article "A conceptual Introduction to Multi-variable Calculus".)

57. Keep the notation of the last item. Then f is differentiable at a iff there exists $f_1\colon J \to \mathbb{R}$ such that (i) f_1 is continuous at a and (ii) $f(x) = f(a) + f_1(x)(x-a)$ for all $x \in J$.

    We saw a couple of uses of this result. We arrived at the analogue of this result for the case of $f\colon U \subset \mathbb{R}^m \to \mathbb{R}^n$: f is differentiable at $a \in U$ iff there exists $f_1\colon U \to L(\mathbb{R}^m, \mathbb{R}^n) \cong M_{n\times m}$ such that (i) f_1 is continuous at a and (ii) $f(x) = f(a) + f_1(x)(x-a)$.

    Hint: Consider
$$f_1(x) := \begin{cases} A + \frac{1}{\|x-a\|^2} E(x)(x-a)^t, & x \neq a \\ A, & x = a \end{cases}.$$
    where $E(x)(x-a)^t$ is the matrix product of the $n \times 1$ matrix $E(x)$ with the $1 \times m$ matrix $(x-a)$.

    We shall prove it in the next class.

Items 53–57 were done on 1 January 2010 (14:30 — 16:00).

58. We proved the result of the last item. The next few items are some of the typical applications.

59. If f is differentiable at a, then f is continuous at a.

60. If $f, g\colon U \subset \mathbb{R}^m \to \mathbb{R}^n$ is differentiable at a, then $f+g$ is differentiable at a with $D(f+g)(a) = Df(a) + Dg(a)$.

61. If $f, g\colon U \subset \mathbb{R}^m \to \mathbb{R}$ is differentiable at a, then $h = fg$ is differentiable at a with $Dh(a) = f(a)Dg(a) + g(a)Df(a)$.

62. **Chain Rule:** Let $f\colon U \subset \mathbb{R}^m \to \mathbb{R}^n$ be differentiable at $a \in U$. Let $g\colon V \subset \mathbb{R}^n \to \mathbb{R}^k$ be differentiable at $b = f(a) \in V$. Then $h := g \circ f$ is differentiable at a with $Dh(a) = Dg(b) \circ Df(a)$.

63. If $f\colon J \subset \mathbb{R} \to \mathbb{R}^n$ is differentiable, then we think of it as a (parametrized) curve in $\mathbb{R}^n$. In such a case, we use the notation γ or c in place of f. The point $\gamma(t)$ is a vector, usually denoted by $(x_1(t), \ldots, x_n(t))$ and is called the position vector of the point $\gamma(t)$. We may think of γ as the trajectory of a particle as it moves along the time interval J.

64. $f\colon U \subset \mathbb{R}^m \to \mathbb{R}^n$ is differentiable at $a \in U$ iff each $f_i = \pi \circ f$ is differentiable at a. (Here π_i is the projection of $\mathbb{R}^n$ on its i-th factor.) We have $Df_i(a) = \pi_i \circ Df(a)$.

65. Keep the notation of the last item. The matrix of the linear map $\pi_i\colon \mathbb{R}^n \to \mathbb{R}$ is the $1 \times n$ matrix $E_i := (0, \ldots, 0, 1, 0, \ldots, 0)$ where 1 is at the i-th place. Hence from the chain rule, if A_i is the $1 \times m$ matrix representing $Df_i(a)$ (of course w.r.t. the standard bases), then $A_i = E_i A$, that is, A_i is the i-th row of A.

66. The most important trick in several variable calculus is to reduce the problem to one variable calculus. The key observation is that if $a \in U$, then for any $v \in \mathbb{R}^M$, there exists $\eta > 0$ such that for $t \in (-\eta, \eta)$, the vector $a + tv \in U$. That is, the line segment $\{a + tv : |t| < \eta\} \subset U$. By restricting f to this open line segment gives rise to a function on $(-\eta, \eta)$ as follows:
$$g_v(t) = f(a + tv), \quad t \in (-\eta, \eta).$$

67. If f is differentiable at $a \in U$, then $g_v'(0)$ exists and we have $g_v'(0) = Df(a)(v)$. More explicitly, we have
$$\lim_{t \to 0} \frac{f(a + tv) - f(a)}{t} = Df(a)(v), \quad \text{for all } v \in \mathbb{R}^m.$$
The limit on the right side is called the directional derivative of f at a in the direction of v and is denoted by $D_v f(a)$. Hence, the displayed formula says $D_v f(a) = Df(a)(v)$.

    Loosely speaking, if f is differentiable at some point, then all its directional derivatives exist at that point.

68. A very important special case of the notion of directional derivatives is when we take $v = e_i$, i-th vector in the standard basis. In this case, the standard notation is $D_{e_i} f(a) = \frac{\partial f}{\partial x_i}(a)$, the i-th partial derivative.

69. Let $f\colon U \subset \mathbb{R}^m \to \mathbb{R}$ be differentiable at $a \in U$. We know that $Df(a)(h) = \sum_{i=1}^m c_i h_i$ where $c_i = Df(a)(e_i)$. In view of the last item, it follows that $c_i := \frac{\partial f}{\partial x_i}(a)$. The vector $\operatorname{grad} f(a) := \left(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_m}(a)\right)$ is known as the gradient of f at a. Note that we have

$$Df(a)(h) = \operatorname{grad} f(a) \cdot h.$$

70. Let $f\colon U \subset \mathbb{R}^m \to \mathbb{R}^n$ be differentiable at a. Putting the observations made in Items 64, 65, 69 we see that the i-th row of $A = Df(a)$ is

$$A_i = Df_i(a) = \left(\frac{\partial f_i}{\partial x_1}(a), \ldots, \frac{\partial f_i}{\partial x_m}(a)\right).$$

Hence, the matrix A is given by

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \ldots & \frac{\partial f_1}{\partial x_m}(a) \\ \vdots & & \vdots \\ \frac{\partial f_i}{\partial x_1}(a) & \ldots & \frac{\partial f_i}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(a) & \ldots & \frac{\partial f_n}{\partial x_n}(a) \end{pmatrix},$$

and is known as the Jacobian matrix of f at a.

Items 58–70 were done on 5 January 2010 (14:30 — 16:00).

71. Exercise:

(a) The converse of Item 67 is not true in general. Consider $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & (x, y) \neq (0, 0) \\ 0 & \text{otherwise.} \end{cases}$$

Then all its directional derivatives at $(0, 0)$ exist. However, f is not even continuous at $(0, 0)$ (and hence is certainly not differentiable at $(0, 0)$).

(b) Consider $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$f(\begin{pmatrix} u \\ v \end{pmatrix}) = \begin{pmatrix} u + v \\ u - v \\ u^2 - v^2 \end{pmatrix}$$

and $g\colon \mathbb{R}^3 \to \mathbb{R}$ given by $g(x, y, z) = x^2 + y^2 + z^2$. Find the Jacobian matrix of $D(g \circ f)$ at $\begin{pmatrix} a \\ b \end{pmatrix}$.

(c) Let $f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \\ r \end{pmatrix}$ and $w = g(x, y, z) = \sqrt{x^2 + y^2 + z^2}$. Find $\frac{\partial w}{\partial r}$ and $\frac{\partial w}{\partial \theta}$ using the chain rule. Check the result by direct substitution.

(d) Let $f, g : (a,b) \to \mathbb{R}^n$ be differentiable. Let $\phi(t) := \langle f(t), g(t)\rangle$. Compute $\phi'(t)$.

(e) Let $f\colon \mathbb{R}^m \to \mathbb{R}^k$ and $g : \mathbb{R}^n \to \mathbb{R}^k$ be differentiable. Let $\phi(x,y) := \langle f(x), g(y)\rangle$. Show that ϕ is differentiable on $\mathbb{R}^m \times \mathbb{R}^n$.

(f) Let $c\colon (a,b) \to \mathbb{R}^n$ be differentiable. We think of c as a curve in $\mathbb{R}^n$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be differentiable. Prove that $g(t) := f \circ c(t)$ is differentiable and $g'(t) = \langle \text{grad } f(c(t)), c'(t)\rangle$. Here $c'(t) = \begin{pmatrix} c_1'(t) \\ \vdots \\ c_n'(t)\end{pmatrix} = Dc(t)(1)$ is the tangent vector to c at t. Note that $g'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(c(t)) \cdot c_i'(t)$.

(g) A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is said to be *homogeneous of degree* k if $f(tx) = t^k x$ for all $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Let f be homogeneous of degree k and differentiable on $\mathbb{R}^n$. Show that

$$D_x f(x) = \langle x, \text{grad } f(x)\rangle = \sum x_i \frac{\partial f}{\partial x_i}(x) = k f(x).$$

This is known as Euler's theorem. Prove also the converse. *Hint for both:* Consider $g(t) = f(tx)$ for the first part and $t^{-k} g(t)$ for the converse.

(h) Find the derivatives of the following functions:
  (1) $f(x,y) = x^y$.
  (2) $f(x,y) = \sin(xy)$.
  (3) $f(x,y) = \int_a^{x+y} g$.
  (4) $f(x,y) = \int_a^{xy} g$.
  (5) $f(x,y) = \int_x^{y} g$.

  In (3) to (5), assume that $g\colon \mathbb{R} \to \mathbb{R}$ is continuous.

(i) Compute the Jacobian matrix of the following functions:
  (1) $(x,y) \mapsto (e^x \cos y, e^x \sin y)$.
  (2) $(x,y) \mapsto (x+y, xy, x-y)$.
  (3) $x \in \mathbb{R}^n \mapsto \langle Ax, x\rangle$ where $A\colon \mathbb{R}^n \to \mathbb{R}^n$ is linear.

(j) Let $c\colon (a,b) \to \mathbb{R}^n$ be differentiable such that $\|c(t)\| = 1$ for $t \in (a,b)$. Prove that $c'(t)$ is perpendicular to $c(t)$ for $t \in (a,b)$. Interpret this result geometrically in terms of spheres and tangent planes.

(k) Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be differentiable. Let 0 be a value of f so that $f^{-1}(0)$ is non-empty. Let $c\colon (a,b) \to \mathbb{R}^n$ be a differentiable curve such that $c(t) \in f^{-1}(0)$ for all $t \in (a,b)$. Show that $\langle c'(t), \text{grad}\, f(c(t))\rangle = 0$. Specialize to $f\colon \mathbb{R}^3 \to \mathbb{R}$ and understand the geometry behind this exercise.

72. We brought Item 71a to your attention.

73. We gave two more applications of the principle stated in Item 66.

74. Given $f\colon U \subset \mathbb{R}^m \to \mathbb{R}$, a point $a \in U$ is said to be a local maximum if there exists $r > 0$ such that (i) $B(a,r) \subset U$ and (ii) for all $x \in B(a,r)$, we have $f(x) \le f(a)$.

    A local minimum is defined similarly.

75. Let $f\colon U \subset \mathbb{R}^m \to \mathbb{R}$, $a \in U$. Assume that all the directional derivatives of f at a exist. Assume further that f has a local maximum/minimum at a. Then $D_v f(a) = 0$ for $v \in \mathbb{R}^m$. In particular, if f is differentiable at a, then $Df(a) = 0$. *Hint:* Fix v. Consider $g(t) := f(a + tv)$.

76. Given a real vector space V, and points $x, y \in V$, we defined the line segment

$$[x, y] := \{x + t(y - x) : 0 \leq t \leq 1\} \equiv \{(1 - t)x + ty : 0 \leq t \leq 1\}.$$

77. We defined a convex subset in a normed linear space. Any open (or closed) ball in such a space is convex.

78. The obvious formulation for the mean value theorem would be: Let $U \subset \mathbb{R}^m$ be convex. Let $f\colon U \to \mathbb{R}^n$ be differentiable. For any $x, y \in U$, there exists z in the line segment joining x and y such that $f(y) - f(x) = DF(z)(y - x)$ is false.

    Example: Consider $f\colon \mathbb{R} \to \mathbb{R}^2$ given by $f(t) = (\cos t, \sin t)$. Look at $f(2\pi) - f(0)$.

79. The third application of the principle of Item 68 is the following form of the **mean value theorem:**

    Let $U \subset \mathbb{R}^m$ be convex. Let $f\colon U \to \mathbb{R}^n$ be differentiable. Let $v \in \mathbb{R}^n$. Then there exists $z \in [x, y]$ such that

$$\langle f(y) - f(x), v\rangle = \langle Df(x)(y - x), v\rangle\,. \tag{2}$$

    Hint: Consider $g(t) := \langle f(x + t(y - x), v\rangle$. It is the composite of $t \mapsto x + t(y - x)$, f and $y' \mapsto \langle y', v\rangle$. Mean value theorem of one variable calculus can be applied to g to get $g(1) - g(0) = g'(t_0)(1 - 0)$. Use chain rule to find $g'(t)$.

80. In several variable calculus, more useful than the mean value theorem is the following mean value inequality:

    Keep the notation of the last item. Then

$$\| f(y) - f(x) \| \leq \sup_{0 \leq t \leq 1} \| Df(x + t(y - x)) \| \, \| y - x \|\,, \tag{3}$$

    assuming the supremum exists. To arrive at it, observe that in an inner product space $\| x \| = \sup\{\langle x, u\rangle : \| u \| = 1\}$.

81. As an application of the mean value inequality, we proved that if $f\colon U \subset \mathbb{R}^n$ has zero derivative on U, then f is locally constant on U, that is, for each $x \in U$, there exists $r_x > 0$ such $B(x, r_x) \subset U$ and f is a constant on $B(x, r_x)$.

82. The derivative of a function being zero does **not** imply the function is a constant. Let $U := (-\infty, -1) \cup (1, \infty)$. Then U is an open set in $\mathbb{R}$. The function $f\colon U \to \mathbb{R}$ defined as $f(x) = -1$ if $x < -1$ and $f(x) = 1$ if $x > 1$ is differentiable with zero derivative.

    A subset of a (metric) space is said to be connected if any locally constant function is a constant. (This is equivalent to the standard definition you will learn in your topology course.)

    Hence we conclude if the domain of f is *connected*, then f is a constant.

83. Exercise: Let $f\colon U \subset \mathbb{R}^m \to \mathbb{R}^n$ be differentiable. Assume that there exists $M > 0$ such that $\| Df(x) \| \leq M$ for $x \in U$. Then f is uniformly continuous on U. *Hint:* Use the mean value inequality.

Items 72–83 were done on 6 January 2010 (14:00 – 15:00).

84. We went through some of the items of the last two classes again.

    (a) Mean Value Theorem; especially mean value inequality as many did not understand the right side of (3).

    (b) Many wanted to understand Item 82. We considered $U := (-\infty, -1) \cup (1, \infty)$ and $f(x) = -1$ if $x \in (-\infty - 1)$ and $f(x) = 1$ if $x \in (1, \infty)$. I explained why U is open, the continuity of f, the differentiability of f and that the derivative is zero.

Item 84 was done on 7 January 2010 (14:30 – 16:00).

85. A powerful tool in calculus is the chain rule. We looked at some typical applications.

86. Let $f, g \colon \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then $h := fg$ is the composite of $x \mapsto (f(x), g(x)) \mapsto f(x)g(x)$. We obtain $\operatorname{grad} fg(a) = f(a) \operatorname{grad} g(a) + g(a) \operatorname{grad} f(a)$.

87. More generally, if $f, g \colon \mathbb{R}^m \to \mathbb{R}^n$ are differentiable, then $h(x) := \langle f(x), g(x) \rangle$ is the composite of $x \mapsto (f(x), g(x)) \mapsto \langle f(x), g(x) \rangle$. As we have already computed the derivatives of these functions, chain rule can be applied to obtain $Dh \colon \mathbb{R}^m \to \mathbb{R}$.

88. Let $\gamma \colon J \subset \mathbb{R} \to \mathbb{R}^n$ and $f \colon \mathbb{R}^n \to \mathbb{R}$ be differentiable. If $g(t) := f \circ \gamma(t)$, then $g'(t) = \operatorname{grad} f(\gamma(t)) \cdot \gamma'(t)$.

89. Consider $f \colon GL(n, \mathbb{R}) \to GL(n, \mathbb{R})$ given by $f(X) = X^{-1}$. Because of the formula for A^{-1} in terms of the cofcators of A, we know f is differentiable. To find its derivative, we applied chain rule to the map $x \mapsto XX^{-1} = I$. This is composite of the maps $X \mapsto (X, X^{-1})$ and $(X, Y) \mapsto XY$.

90. Let A be an $n \times n$ symmetric matrix. Consider $f \colon \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ given by $f(x) = \frac{Ax \cdot x}{x \cdot x}$.

91. Let $f \colon \mathbb{R}^m \to \mathbb{R}^n$ be differentiable. Then to find $Df(a)$, it suffices to find $Df(a)(v)$ for any $v \in \mathbb{R}^m$. But the later is $D_v f(a)$. Let now $\gamma \colon (-\varepsilon, \varepsilon) \to \mathbb{R}^m$ be differentiable with two *initial* conditions: (i) $\gamma(0) = a$ and $\gamma'(0) = v$. (Recall that if $\gamma(t) = (x_1(t), \ldots, x_m(t)$, then $\gamma'(t) = D\gamma(t)(1) = (Dx_1(t)(1), \ldots, Dx_m(t)(1)) = (x_1'(t), \ldots, x_m'(t))$.) Then $D_v f(a) = f \circ \gamma'(0)$. Thus to compute the directional derivative $D_v f(a)$, we can use any curve γ which satisfies the initial conditions! $\gamma'(t)$ is the tangent (or velocity) vector at t to the curve γ. Note that $c := f \circ \gamma$ is a differentiable curve in $\mathbb{R}^n$ such that $c(0) = f(a)$. Hence the directional derivative $D_v f(a)$ is the tangent vector to the curve c at 0.

    This yields the following geometric interpretation of the derivative map. $Df(a)$ maps to the tangent vectors at a to the tangent vectors at $f(a)$.

Items 85–91 were done on 8 January 2010 (14:30 – 16:00).

92. Let $f \colon \mathbb{R}^\ell \times \mathbb{R}^m \to \mathbb{R}^n$ be a bilinear map. Let $(a, b) \in \mathbb{R}^\ell \times \mathbb{R}^m$. Then $Df(a, b)(h, k) = f(a, k) + f(h, b)$. The error term $E(h, k) = f(h, k) = \sum_{i,j} h_i k_j f(u_i, v_j)$ where $\{u_i : 1 \leq i \leq \ell\}$ and $\{v_j : 1 \leq j \leq m\}$ are standard bases of $\mathbb{R}^\ell$ and $\mathbb{R}^m$ resp. Hence

$$\| f(h, k) \| \leq \sum_{i,j} \| h \| \, \| k \| \, M, \text{ where } M = \max\{\| f(u_i, v_j) \|\}.$$

93. More generally, if $f\colon \mathbb{R}^{n_1} \times \cdots \mathbb{R}^{n_k} \to \mathbb{R}^n$ is a k-linear map, then

$$Df(a_1, \ldots, a_k)(h_1, \ldots, h_k) = \sum_{i=1}^{k} f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_k).$$

94. We used the last item to compute the derivative of the determinant function $f\colon M(n, \mathbb{R}) \to \mathbb{R}$ given by $f(X) = \det X$. We found that $Df(I)(H) = \operatorname{Tr}(H)$.

95. Let $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ have continuous partial derivatives on U. Then f is differentiable on U.

96. We say that $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable or C^1 on U if $x \mapsto Df(x)$ as a map from $U \to L(\mathbb{R}^n, \mathbb{R})$ is continuous. Since the matrix of $Df(x)$ is $\left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$, this is equivalent to saying that the partial derivatives of f exist and are continuous on U. (We used Items 69 and Item ??

97. More generally, we say f is C^k if all its partial derivatives of order less than or equal to k exist and are continuous, that is, $\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}$ exist for $i_1, \ldots, i_k \in \{1, \ldots, n\}$. Just to make sure you get it right, you were asked to write all partial derivatives of f where $n = 2$ and $k = 3$.

98. We proved the following result:
Let $f\colon U \subseteq \mathbb{R}^2 \to \mathbb{R}$. Assume that $D_1 f$, $D_2 f$, $D_1 D_2 f$ and $D_2 D_1 f$ exist and are continuous. Then $D_1 D_2 f = D_2 D_1 f$. Here $D_i f = D_{e_i} f$ are the partial derivatives. *Hint:* Consider $g_1(x) = f(x, y+k) - f(x, y)$ and $g_2(y) = f(x+h, y) - f(x, y)$. Apply mean value theorem to $g_1(x+h) - g_1(x)$ and once again to the result which is a function of y. Carry out a similar approach to $g_2(y+k) - g_2(y)$ and use continuity of $D_1 D_2 f$ and $D_2 D_1 f$.

99. More generally, if $f\colon \mathbb{R}^n \to \mathbb{R}$ is C^2, then $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ for any i, j.

100. Still more generally, if $f\colon \mathbb{R}^n \to \mathbb{R}$ is C^k, then any partial derivative of order, say, $r \le k$ can be written of the form $\frac{\partial^r f}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}}$ where $i_1 + \cdots + i_n = r$.

Items 92–100 were done on 11 January 2010 (14:30 – 16:00).

101. We recalled that the function $f(t) = \begin{cases} 0 & \text{for } t \le 0 \\ \exp(-1/t) & \text{for } t > 0, \end{cases}$ is C^∞. Since $f^{(n)}(0) = 0$ for all $n \ge 0$, its Taylor expansion at $t = 0$ is 0. Thus this is a smooth (C^∞) function whose Taylor series converges but NOT to the function.

We also drew the graph of f.

102. Let f be as in the last item. Let $\varepsilon > 0$ be given. Define $g_\varepsilon(t) := f(t)/(f(t) + f(\varepsilon - t))$ for $t \in \mathbb{R}$. Then g_ε is C^∞, $0 \le g_\varepsilon \le 1$, $g_\varepsilon(t) = 0$ iff $t \le 0$ and $g_\varepsilon(t) = 1$ iff $t \ge \varepsilon$.

What is the graph of g_ε?

103. Let f, g be as in the last two items. For $r > 0$ and $x \in \mathbb{R}^n$, define $\varphi(x) := 1 - g_\varepsilon(\|x\| - r)$. Then φ is smooth (Why is φ smooth at 0?) and has the following properties: (i) $0 \leq \varphi \leq 1$, (ii) $\varphi(x) = 1$ iff $\|x\| \leq r$ and $\varphi(x) = 0$ iff $\|x\| \geq r + \varepsilon$.

In particular, if $0 < r < R$, there exist smooth functions f such that $f = 1$ on $B(a, r)$ and 0 outside $B(a, R)$.

Can you visualize the graph of such functions?

104. We reviewed the Taylor expansion for real valued functions of a real variable. If $f \colon J \subset \mathbb{R} \to \mathbb{R}$ is C^k, then

$$f(x) = f(a) + \sum_{j=1}^{k-1} \frac{f^{(j)}}{j!}(a)(x-a)^j + \frac{f^{(k)}}{k!}(y). \tag{4}$$

The most important fact we should know about the remainder term $R_k(x)$ is that $\lim_{x \to a} \frac{R_k(x)}{(x-a)^{k-1}} = 0$.

105. Let $f \colon U \subset \mathbb{R}^n \to \mathbb{R}$ be C^k. We get a Taylor expansion for f by considering $g(t) = f(a + th)$. (This is again an instance of our principle of reduction to one-dimensional case.) To keep the notation simple, we shall assume that $a = 0 \in U$. We use x in place of h.

Then g is the composite of $t \mapsto tx \mapsto f(tx)$. Hence

$$\begin{aligned} g'(t) &= \operatorname{grad} f(tx) \cdot x = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(tx) x_i \\ g''(t) &= \sum_{i=1}^n \operatorname{grad} \frac{\partial f}{\partial x_i}(tx_j) x_j x_i \\ &= \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(tx) x_i x_j, \end{aligned}$$

and so on.

106. In particular, if f is C^2, then

$$f(x) = f(0) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(tx) x_i + \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(t_0 x) x_i x_j, \tag{5}$$

for some $0 \leq t_0 \leq 1$.

107. We reviewed the proof of sufficient conditions for the local extrema of functions $f \colon \mathbb{R} \to \mathbb{R}$. The proof suggested the following item.

108. If 0 is a local minimum of f, then for all x in a neighbourhood of 0, we must have $\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(t_0 x) x_i x_j > 0$. Analogous condition for a local maximum.

109. Since we want the condition on the derivatives of f at 0, we arrive at the following sufficient conditions: Let $f \colon U \to \mathbb{R}$ be C^2. Assume that (i) $Df(a) = 0$ and (ii)

$\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(a) x_i x_j > 0$ for all x with $\|x\|$ sufficiently small. Then a is a local minimum of f.

Similar result for a local maximum.

110. We need to modify the proof of the sufficient condition in the one-dimensional case, as the t_0 of Item 106 will depend on the x and hence we may not be able to find a $\delta > 0$ which will say that for all x with $\|x\| < \delta$, we have $f(a+x) \geq f(a)$. This is explained in the next class. See Item 115.

111. Let $A = (a_{ij})$ be an $n \times n$ symmetric matrix. We say that A is positive definite if $Ax \cdot x > 0$ for $x \in \mathbb{R}^n$ with $x \neq 0$.

 There are two well-known criteria for the positive definiteness of a symmetric matrix $A = (a_{ij})$. (i) All the eigenvalue values of A are positive. (ii) All the matrices $(a_{ij})_{1 \leq i,j \leq k}$ for $1 \leq k \leq n$ have positive determinants.

 The proofs of the second criterion for $n = 2$ and the first criterion for all n are easy.

112. One defines similarly negative definiteness of symmetric matrices. The second condition for negative definiteness reads as follows: the determinants of the principal minors alternate in sign beginning with negative.

Items 101–112 were done on 12 January 2010 (14:30 – 16:00).

113. Let $A = (a_{ij})$ be an $n \times n$ symmetric matrix. The function $f \colon x \mapsto Ax \cdot x$ is continuous (we proved this in two ways!). The unit sphere $S := \{x \in \mathbb{R}^n : \|x\| = 1\}$ is a closed and bounded subset of $\mathbb{R}^n$ and hence by Heine-Borel theorem, S is compact. Hence, the function f attains its maxima and minima on S.

114. Assume that A (of the last item) is positive definite. Then $m := \min_{x \in S} f(x) > 0$. We use this information to solve the problem raised in Item 110.

115. Keep the notation of Items 105-106. Let $A := (\frac{\partial^2 f}{\partial x_i x_j}(a))$ be the Hessian matrix. Let m be as in the last item.

$$f(a+h) = f(a) + 0 + \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(a) h_i h_j + \left(\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(a + t_0 h) - \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i} \right) h_i h_j. \tag{6}$$

 Given $\varepsilon := m/2$, by the continuity of the second order partial derivatives, there exists $\delta > 0$ such that the operator norm $\left\| (\frac{\partial^2 f}{\partial x_i x_j}(a + t_0 h)) - A \right\| < \varepsilon$.

 Using this information in (6), we see that the sign of

$$\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(a) h_i h_j + \left(\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(a + t_0 h) - \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i} \right) h_i h_j$$

 is the same as that of $\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(a) h_i h_j$ for $\|h\| < \delta$.

 This proves the sufficiency of the conditions for the local minimum at a. Similarly, one proves the sufficiency of the conditions for the local maximum at a.

116. Note that the above condtions are sufficient condtions and not necessary condtions. For example, look at $f(x) = x^4$ for $x \in \mathbb{R}$. Clearly $x = 0$ is a point of locally (and also global) minimum for f, but $f'(0) = 0 = f''(0)$.

117. We then made a digression. Keeping the notation of Item 113, we proved that m is an eigenvalue of A and the point $v \in S$ at which f attains this value is an eigenvector of A. The proof was a cocktail of linear algebra, analysis, calculus and geometry. Reference for this is my book on linear algebra.

Items 113–117 were done on 13 January 2010 (14:30 – 16:00).