Math 51 TA notes — Autumn 2007

Jonathan Lee

December 3, 2007

Minor revisions aside, these notes are now essentially final. Nevertheless, I do welcome comments!

Go to `http://math.stanford.edu/~jlee/math51/` to find these notes online.

Contents

1 Linear Algebra — Levandosky's book **2**
1.1 Vectors in $\mathbb{R}^n$ 2
1.2 Linear Combinations and Spans 2
1.3 Linear Independence 3
1.4 Dot Products and Cross Products 3
1.5 Systems of Linear Equations 4
1.6 Matrices 4
1.7 Matrix-Vector Products 5
1.8 Nullspace 5
1.9 Column space 5
1.10 Subspaces of $\mathbb{R}^n$ 6
1.11 Basis for a Subspace 6
1.12 Dimension of a Subspace 7
1.13 Linear Transformations 7
1.14 Examples of Linear Transformations 7
1.15 Composition and Matrix Multiplication 8
1.16 Inverses 8
1.17 Determinants 9
1.21 Systems of Coordinates 9
1.23 Eigenvectors 10
1.25 Symmetric matrices 11
1.26 Quadratic Forms 11

2 Vector Calculus — Colley's book **11**
2.2 Differentiation in Several Variables 11
2.2.1 Functions of Several Variables 11
2.2.2 Limits 12
2.2.3 The derivative 13
2.2.4 Properties of the Derivative; Higher Order Derivatives 15
2.2.5 The Chain Rule 15
2.2.6 Directional Derivatives and the Gradient 15
2.3 Vector-Valued Functions 16
2.3.1 Parametrized curves 16
2.4 Maxima and Minima in Several Variables 16
2.4.1 Differentials and Taylor's Theorem 16
2.4.2 Extrema of Functions 17
2.4.3 Lagrange Multipliers 18
2.4.4 Some Applications of Extrema 19

1 Linear Algebra — Levandosky's book

- useful (non-)Greek letters: $\alpha, \beta, \gamma, \delta, u, v, x, y, z$

1.1 Vectors in $\mathbb{R}^n$

- a *vector* in $\mathbb{R}^n$ is an ordered list of n real numbers; there are two basic vector operations in $\mathbb{R}^n$: *addition* and *scalar multiplication*
- examples of vector space axioms — commutativity, associativity — "can add in any order"
- *standard position* — vector's tail is at the origin

1.2 Linear Combinations and Spans

- a *linear combination* of vectors $\{v_1, \ldots, v_k\}$ in $\mathbb{R}^n$ is a sum of scalar multiples of the v_i
- the *span* is the set of all linear combinations
- a line L in $\mathbb{R}^n$ has a *parametric representation*

$$L = \{x_0 + \alpha v : \alpha \in \mathbb{R}\}$$

  with *parameter* α
- given *two distinct points* on a line, we can find its parametric representation; can also parametrize line segments

- two *non-zero* vectors in $\mathbb{R}^n$ will span either a line or a plane; the former happens if they're *collinear* (or one is *redundant*)
- a plane P in $\mathbb{R}^n$ has a *parametric representation*

$$P = \{x_0 + \alpha v_1 + \beta v_2 : \alpha, \beta \in \mathbb{R}\}$$

  with *parameters* α and β

- given *three non-collinear points* on a plane, we can find its parametric representation
- checking for redundancy — "row reduction"

1.3 Linear Independence

- a set of vectors $\{v_1, \ldots, v_k\}$ is called *linearly dependent* if at least one of the v_i is a linear combination of the others; otherwise, the set is *linearly independent*
- in case $k = 2$, this simply means one vector is a scalar multiple of the other
- *Proposition:* a set of vectors $\{v_1, \ldots, v_k\}$ is linearly dependent if and only if there exist scalars $\alpha_1, \ldots, \alpha_k$, not all zero, such that

$$\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$$

- to test for linear independence, set up a *system of linear equations* and solve it via *elimination*

1.4 Dot Products and Cross Products

- dot product measures orthogonality, can use to define *length* (or *norm* or *magnitude*) as

$$\|v\| = \sqrt{v \cdot v}$$

- *Length/angle formula:* for non-zero $v, w \in \mathbb{R}^n$ differing by angle θ,

$$v \cdot w = \|v\| \, \|w\| \, \cos\theta$$

- *Cauchy-Schwarz:* for non-zero $v, w \in \mathbb{R}^n$,

$$|v \cdot w| \leq \|v\| \, \|w\| \, ,$$

  with equality if $v = \alpha w$ for some $\alpha \in \mathbb{R}$

- *Triangle Inequality:* for non-zero $v, w \in \mathbb{R}^n$,

$$\|v + w\| \leq \|v\| + \|w\| \, ,$$

  with equality iff $v = \alpha w$ for some positive $\alpha \in \mathbb{R}$

- given $v, w \in \mathbb{R}^n$, say they're *orthogonal* if $v \cdot w = 0$; say they're *perpendicular* if they're non-zero as well

- *Proposition:* two non-zero vectors $v, w \in \mathbb{R}^n$ are perpendicular iff orthogonal

- *Proposition:* for orthogonal $v, w \in \mathbb{R}^n$,
$$\|v + w\|^2 = \|v\|^2 + \|w\|^2$$

- a vector is *perpendicular* to a plane if it is orthogonal to all vectors parallel to the plane; given such a perpendicular vector n and a point x_0 in the plane, this induces the description
$$P = \{x \in \mathbb{R}^3 \,:\, n \cdot (x - x_0) = 0\}$$

- given three points u, v, w in a plane, can compute vector n perpendicular to the plane by
  - setting up a system of equations to express that $v - u, v - w$ are orthogonal to n; *or*
  - using the *cross-product* $\times$, which anti-commutes à la *right-hand rule*

- *Length/angle formula:* for non-zero $v, w \in \mathbb{R}^3$ differing by angle θ,
$$\|v \times w\| = \|v\| \, \|w\| \, \sin\theta$$

- *Proposition:* vectors $v, w \in \mathbb{R}^3$ determine a parallelogram of area $\|v \times w\|$

1.5 Systems of Linear Equations

- an *inconsistent* system of linear equations is one without solutions

1.6 Matrices

- there are things in life such as *matrices, coefficient matrices, augmented matrices, row operations, reduced row echelon form* and *pivot entries*

- if a variable corresponds to a column with a pivot, call it a *pivot variable*; otherwise, call it a *free variable*

- to row reduce, iterate the following steps as necessary:
  - identify the left-most non-zero column that *doesn't* contain a pivot but *does* have a non-zero entry in a row without a pivot
  - make that column into a pivot column through row operations

- in a consistent system of linear equations, the free variables parametrize the set of solutions; thus, to solve a system of linear equations:
  - write down its augmented matrix
  - row reduce it
  - solve for the pivot variables
  - if the system turns out consistent, then express its solution as the span of the non-pivot columns translated by some vector

1.7 Matrix-Vector Products

- represent vectors in $\mathbb{R}^n$ by *column vectors*
- a matrix-vector product Ax can be thought of as a linear combination of the columns of A with coefficients given by x; equivalently, as the vector whose i-th coordinate is the dot product of x with the i-th row of A

1.8 Nullspace

- a linear system $Ax = b$ is *homogeneous* if $b = 0$; otherwise, it's *inhomogeneous*; the *null space* of an $m \times n$ matrix A is the set of solutions to $Ax = 0$, denoted by $N(A)$
- *Proposition:* the null space is a subspace of $\mathbb{R}^n$; furthermore, $N(A) = N(\mathrm{rref}(A))$
- *Proposition:* suppose that z is a particular solution to the linear system $Ax = b$; then the set of *all* solutions is given by

$$z + \{\text{solutions to } Ax = 0\}$$

- *Proposition:* for an $m \times n$ matrix A, the following are equivalent:
  - the columns of A are linearly independent
  - $N(A) = \{0\}$
  - $\mathrm{rref}(A)$ has a pivot in each column

1.9 Column space

- the column space $C(A)$ of a matrix A is defined to be the span of the columns of A
- *Proposition:* the system $Ax = b$ has a solution if and only if $b \in C(A)$
- to determine for which b the system $Ax = b$ is consistent when A and x are given:
  - row reduce the corresponding augmented matrix

- pick off the conditions

- for an $m \times n$ matrix A, the following are equivalent:
  - the columns of A span $\mathbb{R}^m$
  - $C(A) = \mathbb{R}^m$
  - $\text{rref}(A)$ has a pivot in each row

- for an $n \times n$ matrix A, the following are equivalent:
  - $C(A) = \mathbb{R}^n$
  - $N(A) = \{0\}$
  - $\text{rref}(A) = I_n$, where I_n is the identity matrix

1.10 Subspaces of $\mathbb{R}^n$

- *Proposition:* a *linear subspace* of a vector space is a subset of it closed under addition, scalar multiplication and containing the zero vector; an *affine linear subspace* is a translation of a subspace

- *Proposition:* the column and null spaces of an $m \times n$ matrix are subspaces of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively

1.11 Basis for a Subspace

- a *basis* for a vector space V is a linearly independent spanning set of vectors

- given a basis $\{v_i\}$ of a vector space V, every vector $v \in V$ is uniquely expressible as a linear combination of the v_i

- notice that row operations preserve linear dependence relations between the columns of a matrix; hence, a selection of columns to form a column space basis remains valid under row operations

- to find bases for the null and column spaces of a matrix:
  - row-reduce the matrix
  - to compute the nullspace, identify the free and pivot columns, solve for the pivot variables in terms of the free ones, write out a solution and then factor out variables to obtain a basis
  - to compute the column space, identify the pivot columns and take the corresponding columns in the original matrix as a basis

1.12 Dimension of a Subspace

- the *dimension* of a vector space V is the size of a basis for V; by the following, it is well-defined
- *Proposition:* if a set of m vectors spans a vector space V, then any set of $n > m$ vectors in V is linearly dependent
- given a vector space V of dimension d, any set of d vectors is linearly independent iff it spans V iff it's a basis for V
- given an $m \times n$ matrix A, define its rank and nullity to be

$$\operatorname{rank}(A) = \dim(C(A)) \quad \text{and} \quad \operatorname{nullity}(A) = \dim(N(A))$$

- *Rank-Nullity Theorem:* for an $m \times n$ matrix A,

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n$$

1.13 Linear Transformations

- terminology: *function, domain, codomain, real-valued, scalar-valued, vector-valued, image, pre-image*
- *Proposition:* $m \times n$ matrices correspond to linear transformations $\mathbb{R}^n \to \mathbb{R}^m$

1.14 Examples of Linear Transformations

- transformations represented by the obvious matrices: *identity transformation, scaling transformation, diagonal matrices*
- to rotate the plane $\mathbb{R}^2$ counter-clockwise by the angle θ, use the *rotation matrix*

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

- to rotate around an axis in $\mathbb{R}^3$, take a direct sum decomposition $\mathbb{R}^3 = \mathbb{R}^2 \oplus \mathbb{R}$ as appropriate and paste together the rotation matrix from above along with the identity matrix
- given a line L in $\mathbb{R}^n$ spanned by the unit vector $u = (u_1, u_2, \ldots, u_n)$, the *orthogonal projection* of $\mathbb{R}^n$ onto L is given by

$$\begin{bmatrix} u_1^2 & u_1u_2 & u_1u_3 & \cdots & u_1u_n \\ u_1u_2 & u_2^2 & u_2u_3 & \cdots & u_2u_n \\ u_1u_3 & u_2u_3 & u_3^2 & \cdots & u_3u_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_1u_n & u_2u_n & u_3u_n & \cdots & u_n^2 \end{bmatrix}$$

- given a line L in $\mathbb{R}^n$, reflection around L is given by the identity matrix subtracted from twice the orthogonal projection matrix above

1.15 Composition and Matrix Multiplication

- given functions $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$, their *composition* is $X \xrightarrow{f} Y \xrightarrow{g} Z$, denoted by $g \circ f$
- *Proposition:* given linear transformations $\mathbb{R}^k \xrightarrow{T} \mathbb{R}^n$ and $\mathbb{R}^n \xrightarrow{S} \mathbb{R}^m$, their composition is the linear transformation $\mathbb{R}^k \xrightarrow{S \circ T} \mathbb{R}^m$
- the *matrix product* of an $m \times n$ and an $n \times k$ matrix can be defined as the matrix representing the composition of the linear transformations represented by the given matrices; can also generalize the two interpretations of matrix-vector product
- matrix multiplication satisfies associativity and distributivity as linear transformations do

1.16 Inverses

- a set X, define the *identity function* $X \xrightarrow{I_X} X$ by $I_X(x) = x$ for all $x \in X$
- a function $X \xrightarrow{f} Y$ is called *invertible* if there exists an *inverse* $Y \xrightarrow{f^{-1}} X$ such that $f^{-1} \circ f$ and $f \circ f^{-1}$ are the identities on X and Y, respectively
- *Proposition:* inverses are unique
- *Proposition:* a function is invertible if and only if it's a bijection
- *Proposition:* a linear transformation $\mathbb{R}^n \xrightarrow{T} \mathbb{R}^m$ is a surjection if and only if $\operatorname{rank}(T) = m$
- *Proposition:* a linear transformation $\mathbb{R}^n \xrightarrow{T} \mathbb{R}^m$ is an injection if and only if $\operatorname{rank}(T) = n$
- *Proposition:* a linear transformation $\mathbb{R}^n \to \mathbb{R}^n$ represented by the matrix A is invertible if and only if $\operatorname{rref}(A) = I_n$
- to compute the inverse of a matrix, augment it by the identity matrix and then row-reduce; notice this works because solutions to $[A \mid I_n]$ correspond to those of $\operatorname{rref}([A \mid I_n]) = [I_n \mid A^{-1}]$
- given invertible functions $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$, then $(f^{-1})^{-1} = f$ and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$; there is a corresponding statement for matrices

1.17 Determinants

- one can define the *determinant* of a 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

  to be the value $ad - bc$; determinants of $n \times n$ matrices for $n \geq 3$ can be defined inductively via the alternating block-sum expansion formula

- a square matrix is *upper (lower) triangular* if the entries below (above) the diagonal are zero

- *Proposition:* the determinant is the unique alternating n-linear function that takes the value 1 on I_n

- determinants can thus be more easily computed by row-reducing first

- *Proposition:* an $n \times n$ matrix is invertible if and only if it has non-zero determinant

- *Proposition:* determinants are multiplicative: given $n \times n$ matrices A and B, then

$$\det(AB) = \det(A) \cdot \det(B)$$

- *Proposition:* given an $n \times n$ matrix A, the volume of the paralleliped generated by the columns of A is $|\det(A)|$

1.21 Systems of Coordinates

- as an application to finding parametrizations of subspaces of $\mathbb{R}^n$, we can introduce systems of coordinates — if $\mathcal{B} = \{v_1, v_2, \ldots, v_k\}$ is a basis for a vector space V, then any vector $v \in V$ is uniquely expressible as a linear combination of the v_i; define the *coordinates of v with respect to $\mathcal{B}$* to be the scalars $c_1, \ldots, c_k \in \mathbb{R}$ such that

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

- given a vector v in a vector space V with basis $\mathcal{B}$, write $[v]_{\mathcal{B}}$ to denote the vector whose entries are the coordinates of v with respect to $\mathcal{B}$

- if $\mathcal{B}$ is a basis for a subspace $V \subseteq \mathbb{R}^n$, form the *change of basis matrix C* whose columns are the elements of $\mathcal{B}$ expressed in standard coordinates; then given the coordinates $[v]_{\mathcal{B}}$ of a vector in V with respect to $\mathcal{B}$, we can calculate its standard coordinates from

$$v = C[v]_{\mathcal{B}}$$

- if $\mathcal{B}$ is a basis for $\mathbb{R}^n$, then given a vector $v \in \mathbb{R}^n$ expressed in standard coordinates, we can calculate its coordinates with respect to $\mathcal{B}$ from
$$[v]_{\mathcal{B}} = C^{-1}v$$
- given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ and a basis $\mathcal{B}$ with change of basis matrix C, then we have
$$[T]_{\mathcal{B}} = C^{-1}[T]C$$
where $[T]$ denotes the matrix representing T with respect to standard coordinates
- we say two $n \times n$ matrices A and B are *similar* if $A = CBC^{-1}$ for some invertible matrix C; that is, A and B represent the same linear transformation with respect to different bases
- *Proposition:*
  - similarity is an equivalence relation (satisfying symmetry, reflexivity, transitivity)
  - similar matrices have the same determinant
  - similar matrices have similar inverses
  - similar matrices have similar powers

1.23 Eigenvectors

- let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ a basis for $\mathbb{R}^n$; if $[T]_{\mathcal{B}}$ is a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$, then for each $1 \leq i \leq n$, we have that
$$T(v_i) = \lambda_i v_i$$
- if $Tv = \lambda v$ for some vector $v \neq 0$ and a scalar $\lambda \in \mathbb{R}$, then we define v to be an *eigenvector* with *eigenvalue* λ for the linear transformation T
- *Proposition:* given an $n \times n$ matrix A, then a scalar $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if $\lambda I_n - A$ has non-trivial nullspace if and only if $\det(\lambda I_n - A) = 0$
- *Proposition:* given an $n \times n$ matrix A with an eigenvalue λ, the set of λ-eigenvectors is the nullspace of $\lambda I_n - A$
- define the *characteristic polynomial* of an $n \times n$ matrix A to be the polynomial
$$p(\lambda) = \det(\lambda I_n - A)\,;$$
observe that its roots are the eigenvectors of A
- given a linear transformation T, say it is *diagonalizable* if there exists a basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is diagonal; such a basis is called an *eigenbasis* and consists of eigenvectors of A
- *Proposition:* if an $n \times n$ matrix A has n distinct eigenvalues, then it is diagonalizable

1.25 Symmetric matrices

- *Proposition:* all the eigenvalues of a real symmetric matrix are real numbers
- *Spectral Theorem:* let T be a self-adjoint linear transformation on a finite-dimensional vector space V; then V decomposes as the direct sum of the eigenspaces of T
- *Corollary:* if A is an $n \times n$ real symmetric matrix, then $\mathbb{R}^n$ has an orthonormal basis consisting of eigenvectors of A
- *Proposition:* given a linear transformation, two eigenvectors with distinct eigenvalues are necessarily orthogonal

1.26 Quadratic Forms

- given an $n \times n$ symmetric matrix, we can define a *quadratic form* $Q : \mathbb{R}^n \to \mathbb{R}$ such that $x \mapsto x^T Ax$
- a quadratic form is defined to be *positive definite, positive semi-definite, negative definite*, or *negative semi-definite* if the values it takes are positive, non-negative, negative or non-positive, respectively; if none of these holds, the form is defined to be *indefinite*
- *Remark:* by definition, if a quadratic form is positive definite, then it is also positive semi-definite
- *Proposition:* given a quadratic form Q defined by $x \mapsto x^T Ax$, we can recognize its type according to the following table:

type of quadratic form Q	eigenvalues of A	A's determinant	A's trace
positive definite	all positive	positive	positive
positive semi-definite	all non-negative	zero	positive
negative definite	all negative	positive	negative
negative semi-definite	all non-positive	zero	negative
indefinite	one positive, one negative	negative	irrelevant
degenerate	both zero	zero	zero

2 Vector Calculus — Colley's book

2.2 Differentiation in Several Variables

2.2.1 Functions of Several Variables

- given a function $f : \mathbb{R}^n \to \mathbb{R}$, define the *level set at height c of f* to be

$$\{v \in \mathbb{R}^n : f(v) = c\};$$

similarly, define the *contour set at height c of f* to be

$$\{(v, c) \in \mathbb{R}^{n+1} : f(v) = c\}$$

- given a function $f : \mathbb{R}^2 \to \mathbb{R}$, we can take *sections* (or *slices*) of its graph by fixing an appropriate coordinate of $\mathbb{R}^2$
- graphing surfaces in $\mathbb{R}^3$:
  - if one variable occurs as a linear term, such as in the examples:

    $$z = \frac{x^2}{4} - y^2\,, \quad x = \frac{y^2}{4} - \frac{z^2}{9}\,, \quad z = y^2 + 2\,;$$

    then express that variable as a function of the others (that is, solve for it) and then draw enough level sets to get a feel for the picture
  - if all variables appear as quadratic terms, such as in the examples:

    $$z^2 = \frac{x^2}{4} - y^2\,, \quad x^2 + \frac{y^2}{9} - \frac{z^2}{16} = 0\,, \quad \frac{x^2}{4} - \frac{y^2}{16} + \frac{z^2}{9} = 1\,, \quad \frac{x^2}{25} + \frac{y^2}{16} = z^2 - 1$$

    then rearrange the equation into the standard form for some quadric surface and graph it accordingly

2.2.2 Limits

- given a function $f : X \to \mathbb{R}^m$, where $X \subseteq \mathbb{R}^n$,

  $$\lim_{x \to a} f(x) = L$$

  means that for any $\epsilon > 0$, there always exists some $\delta > 0$ such that $0 < \|x - a\| < \delta$ implies that $\|f(x) - L\| < \epsilon$
- to prove that $f(x) - L$ can be made arbitrarily small, it helps to write it as the sum of smaller expressions such as

  $$f(x) - L = h_1(x) + h_2(x) + \cdots + h_k(x)\,;$$

  then by the triangle inequality, which states that

  $$\|f(x) - L\| \leq \|h_1(x)\| + \|h_2(x)\| + \cdots + \|h_k(x)\|\,,$$

  if the $h_i(x)$ can simultaneously be made arbitrarily small, then so can $f(x) - L$
- *Theorem:* limits are preserved under addition, multiplication, division (by a non-zero limit), and scaling

- *Theorem:* let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function with components $f_1, f_2, \ldots, f_m$; then if for some $a \in \mathbb{R}^n$ there exist $L_1, L_2, \ldots, L_m$ such that $\lim_{x\to a} f_i(x) = L_i$ for all i, then $\lim_{x\to a} f(x) = (L_1, L_2, \ldots, L_m)$

- since the limit of a function at a point is independent of its value at a point, especially if said value does not exist, it helps to simplify first, as in:

$$\lim_{(x,y)\to(0,0)} \frac{x^2 + 2xy + y^2}{x + y}, \quad \lim_{(x,y)\to(0,0)} \frac{x^4 - y^4}{x^2 + y^2}$$

- a function $f : X \to Y$ is *continuous* at a point $x \in X$ if the limit $\lim_{x\to a} f(x)$ exists and is equal to $f(x)$; if f is continuous at all its points, then say it is *continuous*

  - polynomials and linear transformations are continuous, as in:

$$\lim_{(x,y,z)\to(0,0,0)} x^2 + 2xy + yz + z^3 + 2$$

- the set of continuous functions is closed under addition, multiplication, division (by a non-zero function), scaling and composition; a function to $\mathbb{R}^m$ is continuous if and only if each of its components are

  - examples:

$$\lim_{(x,y)\to(2,0)} \frac{x^2 - y^2 - 4x + 4}{x^2 + y^2 - 4x + 4}$$

- if a limit exists, then it must equal the limit obtained by approaching the point along any direction; hence, to show a limit does *not* exist, show that approaching the point from different directions gives different limits, as in:

$$\lim_{(x,y,z)\to(0,0,0)} \frac{xy - xz + yz}{x^2 + y^2 + z^2}, \quad \lim_{(x,y)\to(0,0)} \frac{x^2}{x^2 + y^2}, \quad \lim_{(x,y)\to(0,0)} \frac{(x + y)^2}{x^2 + y^2}, \quad \lim_{(x,y)\to(0,0)} \frac{2x^2 + y^2}{x^2 + y^2}$$

- by introducing polar coordinates (r, θ) and the substitutions $x = r\cos\theta$, $y = r\sin\theta$, some limits become easier to compute, as in:

$$\lim_{(x,y)\to(0,0)} \frac{x^2 y}{x^2 + y^2}, \quad \lim_{(x,y)\to(0,0)} \frac{x^2}{x^2 + y^2}, \quad \lim_{(x,y)\to(0,0)} \frac{x^2 + xy + y^2}{x^2 + y^2}$$

2.2.3 The derivative

- given a function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$, for any $1 \leq i \leq n$, we can define the i-th *partial derivative* to be the function $\frac{\partial f}{\partial x_i}$ obtained by differentiating with respect to the variable x_i while treating all other variables as constants

- if the function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ (so that $f = (f_1, \dots, f_m)$ has all partial derivatives at the point $p = (p_1, \dots, p_n) \in \mathbb{R}^n$, then it has a linear approximation h at the point p given by
$$h(x) = f(p) + \frac{\partial f}{\partial x_1}(p) \cdot (x_1 - p_1) + \cdots + \frac{\partial f}{\partial x_n}(p) \cdot (x_n - p_n)\,;$$
more concisely, if we define the *Jacobian matrix* (that is, the *matrix of partial derivatives*) to be the $n \times m$ matrix Df whose (i,j)-th entry is $\frac{\partial f_i}{\partial x_j}$, we can write h as
$$h(x) = f(p) + (Df)(p)(x - p)$$
  - find the Jacobian and linear approximations for the function f at the point p:
    * $f(x, y) = x/y$ at $p = (3, 2)$
    * $f(x, y, z) = (xyz, \sqrt{x^2 + y^2 + z^2})$ at $p = (1, 0, -2)$
    * $f(t) = (t, \cos(2t), \sin(5t))$ at $a = 0$
    * $f(x, y, z, w) = (3x - 7y + z, 5x + 2z - 8w, y - 17z + 3w)$ at $a = (1, 2, 3, 4)$

- say f is *differentiable* at a point p if
$$\lim_{x \to p} \frac{f(x) - h(x)}{\|x - p\|} = 0$$

- if f is a map from $\mathbb{R}^2$ to $\mathbb{R}$ that is differentiable at $p = (a, b) \in \mathbb{R}^2$, then at the point $(p, f(a, b)) \in \mathbb{R}^3$, the graph has tangent vectors $(1, 0, \frac{\partial f}{\partial x}(a, b))$ and $(0, 1, \frac{\partial f}{\partial y}(a, b))$ and perpendicular vector $(-\frac{\partial f}{\partial x}(a, b), -\frac{\partial f}{\partial y}(a, b), 1)$ by the cross-product formula
  - find the tangent plane to the graph of $z = x^3 - 7xy + e^y$ at $(-1, 0, 0)$
  - find the tangent plane to the graph of $z = 4\cos(xy)$ at the point $(\pi/3, 1, 2)$
  - find the tangent plane to the graph of $z = e^{x+y}\cos(xy)$ at the point $(0, 1, e)$
  - find equations for the planes tangent to $z = x^2 - 6x + y^3$ that are parallel to the plane $4x - 12y + z = 7$

  (recall from linear algebra that there exist both standard and parametrized equations for a plane)

- *Theorem:* let X be a neighbourhood of a point $(a, b) \in \mathbb{R}^2$; if the partial derivatives of a function $f : X \to \mathbb{R}$ all exist and are continuous at (a, b), then f is differentiable at (a, b)

- *Theorem:* a differentiable function must be continuous

2.2.4 Properties of the Derivative; Higher Order Derivatives

- *Proposition:* taking derivatives is a linear operation
- *Proposition:* if f and g are scalar-valued differentiable functions on the same space, then the product and quotient rules hold
- by taking successive derivatives of a function k times, one obtains a *mixed partial derivative* of *order* k
- *Theorem:* given a function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ whose partial derivatives of order k all exist and are continuous, then taking mixed partial derivatives is independent of order

2.2.5 The Chain Rule

- given two differentiable functions $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^\ell$, the partial derivative

  $$\frac{\partial (g \circ f)_i}{\partial x_j}(p) \quad \text{is given by} \quad \sum_{k=1}^{m} \frac{\partial g_i}{\partial y_k}(f(p)) \cdot \frac{\partial f_k}{\partial x_j}(p)\,;$$

  (it may help to assign $\mathbb{R}^n$, $\mathbb{R}^m$ and $\mathbb{R}^\ell$ coordinates x_i, y_k and z_j, respectively)
- for f and g as above, we may compute the entire Jacobian matrix at once via

  $$D(g \circ f)(p) = (Dg)(f(p)) \cdot (Df)(p)$$

- polar/rectangular conversions: using the substitutions $x = r\cos(\theta)$ and $y = r\sin(\theta)$ between polar and Euclidean coordinates, we can find the partial derivatives of a function, defined with respect to Euclidean coordinates, with respect to its polar coordinates

2.2.6 Directional Derivatives and the Gradient

- given a function $f : \mathbb{R}^n \to \mathbb{R}$, a point $p \in \mathbb{R}^n$ and a unit vector $v \in \mathbb{R}^n$, we can define $D_v f(p)$, the *directional derivative of f at a in the direction of v,* to be $g'(0)$, where we have set $g(t) = f(p + tv)$
- define the *gradient* of a function $f : \mathbb{R}^n \to \mathbb{R}$ to be the vector of partial derivatives; that is,

  $$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

- directional derivatives can be expressed in terms of the gradient (which, as an overall derivative, encapsulates all the directional ones) by means of

  $$D_v f(p) = \nabla f(p) \cdot v$$

– find the directional derivatives $D_v f(p)$ for

function	point	direction
$f(x,y) = e^y \sin(x)$	$p = (\frac{\pi}{3}, 0)$	$v = \frac{1}{\sqrt{10}}(3,-1)$
$f(x,y) = \frac{1}{x^2+y^2}$	$p = (3,-2)$	$v = (1,-1)$
$f(x,y) = e^x - x^2 y$	$p = (1,2)$	$v = (2,1)$
$f(x,y,z) = xyz$	$p = (-1,0,2)$	$v = \frac{1}{\sqrt{5}}(0,2,-1)$
$f(x,y,z) = e^{-(x^2+y^2+z^2)}$	$p = (1,2,3)$	$v = (1,1,1)$

- the tangent plane of a function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $p \in \mathbb{R}^n$ is given by the parametric equation

$$h(x) = f(p) + \nabla f(p) \cdot (x - p)\,;$$

in the case $n = 2$, we can use this to find non-zero vectors perpendicular to the plane

- a surface in $\mathbb{R}^n$ defined by the equation $f(x,y,z) = c$ has a perpendicular vector given by ∇f; hence, if ∇f is non-zero, an equation for the tangent plane at a point p is given by

$$\nabla f(p) \cdot (v - p) = 0$$

– find the plane tangent to the following surfaces at the given points

surface	point
$x^3 + y^3 + z^3 = 7$	$p = (0,-1,2)$
$ze^y \cos(x) = 1$	$p = (\pi, 0, -1)$
$2xz + yz - x^2 y + 10 = 0$	$p = (1,-5,5)$
$2xy^2 = 2z^2 - xyz$	$p = (2,-3,3)$

2.3 Vector-Valued Functions

2.3.1 Parametrized curves

For sure.

2.4 Maxima and Minima in Several Variables

2.4.1 Differentials and Taylor's Theorem

- given a k-times differentiable function $f : \mathbb{R} \to \mathbb{R}$, we can define its k-th order *Taylor polynomial* at the point p to be

$$p_k(x) = \sum_{i=0}^{k} \frac{f^{(i)}(p)}{i!} \cdot (x - p)^i$$

- *Taylor's theorem* provides us an error estimate: if $f : \mathbb{R} \to \mathbb{R}$ is a $(k+1)$-times differentiable function, then there exists some number ξ between p and x such that

$$R_k(x,a) = f(x) - p_k(x) = \frac{f^{(k+1)}(\xi)}{(k+1)!} \cdot (x-a)^{k+1}$$

- given a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we can define its first-order *Taylor polynomial* at the point $p = (p_1, \ldots, p_n)$ to be

$$\begin{aligned} p_1(x) &= f(p) + Df(p) \cdot (x-p) \\ &= f(p) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot (x_i - p_i)\,; \end{aligned}$$

similarly, if f is twice-differentiable, we can define its second-order *Taylor polynomial* at the point p to be

$$\begin{aligned} p_2(x) &= f(p) + Df(p) \cdot (x-p) + \frac{1}{2}\left[(x-p)^T \cdot Hf(p) \cdot (x-p)\right] \\ &= f(p) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot (x_i - p_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i x_j}(p) \cdot (x_i - p_i)(x_j - p_j)\,, \end{aligned}$$

where for notational sanity, we define the *Hessian matrix* to be the $n \times n$ matrix whose (i,j)-th entry is $\frac{\partial^2 f}{\partial x_i x_j}$

  - find the first- and second-order Taylor polynomials for

the function	at the point
$f(x,y) = 1/(x^2+y^2+1)$	$a = (0,0)$
$f(x,y) = 1/(x^2+y^2+1)$	$a = (1,-1)$
$f(x,y) = e^{2x}\cos(3y)$	$a = (0,\pi)$

- as before, *Taylor's theorem* provides an error estimate; the same formula holds, with the change that ξ is taken to be a point on the line segment connecting p and x

2.4.2 Extrema of Functions

- given a function $f : \mathbb{R}^n \to \mathbb{R}$, know what its *extrema*, both *global* and *local*, are defined to be; this (and the following) works analogously to the single-variable case

- define a point p to be a *critical point* of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ if $Df(p) = 0$

- *Theorem:* local extrema of differentiable functions must be critical points

- *Theorem:* let p be a $f : \mathbb{R}^n \to \mathbb{R}$ be a twice-differentiable function; then

if $Hf(p)$ is	then p is a ... of f
positive definite	local minimum
negative definite	local maximum
neither of the above but still invertible	saddle point

  - find the point on the plane $3x - 4y - z = 24$ closest to the origin
  - determine the absolute extrema of
$$f(x, y) = x^2 + xy + y^2 - 6y$$
on the rectangle given by $x \in [-3, 3]$ and $y \in [0, 5]$
  - determine the absolute extrema of
$$f(x, y, z) = \exp(1 - x^2 - y^2 + 2y - z^2 - 4z)$$
on the ball
$$\{(x, y, z) \in \mathbb{R}^3 \, : \, x^2 + y^2 - 2y + z^2 + 4z \leq 0\}$$

- by Heine-Borel, say that a subset X of $\mathbb{R}^n$ is *compact* if it is closed and bounded

- *Extreme Value Theorem:* any continuous $\mathbb{R}$-valued function on a compact topological space attains global minima and maxima

2.4.3 Lagrange Multipliers

- supposing that we have a continuously differentiable function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$ is defined to be the set of solutions to $g_1 = g_2 = \cdots = g_k = 0$ for some continuously differentiable functions $g_1, \ldots, g_k : \mathbb{R}^n \to \mathbb{R}$, then we can determine the critical points of f by:

  - solving the system of linear equations (in the variables $x, \lambda_1, \ldots, \lambda_k$)
$$Df(x) = \lambda_1 Dg_1(x) + \cdots + \lambda_k Dg_k(x) \quad \text{and} \quad g_1(x) = g_2(x) = \cdots = g_k(x) = 0$$
using elimination, cross-multiplication or other convenient methods; each solution x will be a critical point of f
  - determining the points x where the functions $Dg_1, \ldots, Dg_k$ are linearly dependent, which in the case $k = 1$ (and $Dg = Dg_1$) amounts simply to finding those x such that $Dg(x) = 0$; only some of these points will be critical points and thus they all need to be inspected individually

- as a result, we now have three methods for determining the critical points of a continuously differentiable function $f : S \to \mathbb{R}$, for some specified set S:

  - if S is a curve (which has dimension 1), a surface (which has dimension 2), or in general, some subset of dimension n, then attempt to parametrize S by a function $g : \mathbb{R}^n \to S$ and subsequently compute the critical points of $f \circ g : \mathbb{R}^n \to \mathbb{R}$ — this method can quite often be unnecessarily brutal, requiring many error-prone calculations, which you may illustrate to yourself by finding the closest point on a line to a given point
  - find some geometric interpretation of the problem, draw some picture making it clear where the extrema should occur, and be creative computing the coordinates of such extrema — for example, in the case of finding the closest point on a given plane to a given point, this amounts to determining the intersection of the plane with the unique line perpendicular to the plane that crosses the specified point
  - use Lagrange multipliers

- good, wholesome, enriching entertainment:
  - find the largest possible sphere, centered around the origin, that can be inscribed inside the ellipsoid $3x^2 + 2y^2 + z^2 = 6$
  - the intersection of the planes $x - 2y + 3z = 8$ and $2z - y = 3$ forms a line; find the point on this line closest to the point $(2, 5, -1)$
  - the intersection of the paraboloid $z = x^2 + y^2$ with the plane $x + y + 2z = 2$ forms an ellipse; determine the highest and lowest (with respect to the z-coordinate) points on it

2.4.4 Some Applications of Extrema

- extrema naturally show up in any field of study utilizing quantitative data; consequently, it is probably more interesting to ask what the non-applications are