Advanced Problem Solving
Lecture Notes and Problem Sets

Peter Hästö
Torbjörn Helvik
Eugenia Malinnikova

Contents

1. Basic problem solving techniques

by Peter Hästö

Polya divides problem solving into four stages, Understanding the problem, Devising a plan, Carrying out the plan and Looking back. Of these stages I would here like to focus on the second and fourth.

In a standard mathematical proof you will see a neatly written presentation of how to solve the problem at hand. This is the result of Polya's stage three. However, to understand where the solution came from clearly stage two is more important. In the group discussions in this course we will undertake the difficult endeavor of discussing our solutions as they occurred in Stage 2. Some practical tricks to achieve this are an informed uninterest in the details (and results!) of carrying out the plan, and notes in the solution marking what led one to the ideas.

Another stage frequently neglected is the looking back. In this course this function will be naturally covered by the discussions in the tutor groups. However, to get the most out of these discussions every participant should be prepared to analyze the problems, which in particular involves having tried to solve them.

1.1. **Induction.** Formally induction is quite a simple technique: you verify the initial condition P_0 and the implication $P_n \Rightarrow P_{n+1}$, and remember to use the magic words "follows by induction". The difficulty is finding the right proposition and the index on which to work the induction. Of course, we can also do induction on several indices at the same time (in this case you should be especially careful about circular reasoning) or of more complex "indices", such as "n+m" or "nm".

Closely related to induction is recursion. In recursion we want to find some object, and reduce this task to finding one of a set of sub-object. For instance, consider a rectangular subset of a square lattice, say $R_{m,n} = \{(0,0), \dots, (0,n), (1,0), \dots, (m,n)\}$, and all paths going up and to the right connecting the corners of $R_{m,n}$. We want to count the number of such paths, say $r_{m,n}$. We can split the set of paths into two disjoint sets, those that pass through $(0,1)$ and those that pass through $(1,0)$. Thus we see that

$$r_{m,n} = r_{m,n-1} + r_{m-1,n}.$$

This is the same recursion as satisfied by the binomial coefficients, and since the initial conditions ($r_{i,1} = r_{1,i} = 1$) are also the same, we conclude that

$$r_{m,n} = \binom{m+n}{m}.$$

1.2. **Combinatorial proofs.** We continue with our rectangular lattice and give bijective or combinatorial proofs of the equation

$$r_{m,n} = \binom{m+n}{m} = \sum_{k=0}^{n} \binom{n}{k}\binom{m}{k}.$$

For simplicity of notation we assume that $n \le m$ throughout this section.

By a combinatorial or bijective proof we mean that both sides of the equation are interpreted as cardinalities of sets and the equality is shown by finding a bijection between the sets. In many cases the sets will actually be the same (and the bijection the identity) and the work goes into the interpretation.

For the above equality we already have the interpretation of the left-hand-side, it is the number of up-right paths connecting the lower-left corner to the upper-right corner. Fix $k \le n$ and choose a k-element subset M of $\{0, \dots, n-1\}$ and a k-element subset N of $\{1, \dots, m\}$.

We define a bijection from pairs (M,N) to paths as follows. We go right from $(0,0)$ to $(M_1,0)$, where M_1 is the smallest element in M. Then we go up from $(M_1,0)$ to (M_1,N_1), where N_1 is the smallest element in N. Then we got right to (M_2,N_1) etc. From (M_k,N_k) we go right all the way to (m,N_k) and then up to (m,n). Thus we have a path from $(0,0)$ to (m,n) for every pair (M,N). To comlpete the proof, we need to check that this correspondence defines a bijection, but this is left for the reader.

We can define another bijection on the same set to prove the same inequality as follows. Let D denote the diagonal at distance n from $(0,0)$, i.e. $D=\{(n,0),(n-1,1),\dots,(0,n)\}$. Now the number of paths from $(0,0)$ to $(n-k,k)$ equals $\binom{n}{k}$. The number of paths from $(n-k,k)$ to (m,n) equals $\binom{m}{m-k}=\binom{m}{k}$. So the number of paths through $(n-k,k)$ equals $\binom{n}{k}\binom{m}{k}$ and since every path passes through one and only one of the points in D, we see that the total number of paths is

$$\sum_{k=0}^{n}\binom{n}{k}\binom{m}{k}.$$

We leave it to the reader to give combinatorial proofs of the following identities, which are derived from the formula

$$(1+x)^n=\sum_{k=0}^{n}\binom{n}{k}x^k$$

by setting $x=1$, setting $x=-1$ and by differentiating and setting $x=1$, respectively:

$$\sum_{k=0}^{n}\binom{n}{k}=2^n,\quad \sum_{k=0}^{n}(-1)^k\binom{n}{k}=0,\quad \sum_{k=0}^{n}k\binom{n}{k}=n2^{n-1}. \tag{1.1}$$

1.3. **Inclusion-Exclusion.** Although not covered during the lecture, the principle of inclusion and exclusion is directly related to combinatorial proofs. The idea is that we constructed our bijection poorly, so that some elements were counted multiple times (inclusion). We correct the situation by removing the over-counted elements (exclusion). Typically, we will also over-do the removal, so we go for another inclusion, then exclusion etc.

As an example, let's consider the classical derangement problem: n people go to the restaurant and leave their coats in a coat-check. There is some mix-up with the tags, and they get back random coats. What is the probability that someone gets the right coat back?

Now we "estimate" the ways in which at least one person gets the right coat. For each $k=1,\dots,n$ we give person k his or her coat, and assign the other coats randomly in $(n-1)!$ ways. Obviously, in each case at least one person gets the right coat, and there are a total of $n!$ such assignments. However, several assignments are multiply counted, namely, an assignment in which exactly k people get the right coat is counted k times. So we generate assignments by choosing two-people subsets, giving these people the right coats, and distributing the rest randomly. There are $\binom{n}{2}(n-2)!$ such assignments, but again multiple counts. This time we count a set with k correct assignments $\binom{k}{2}$ times. When we subtract the latter set of assignments from the former, we see that those assignments where exactly 1 or 2 people get the right coat are correctly counted, whereas those where exactly k people get the right coat are counted $k-\binom{k}{2}$ times. Now we see that we can continue this process of generating assignments and alternatively adding and subtracting them. When we do this also for $=3,\dots n$, we see that the assignments for which exactly k people get the right coat are counted

$$\sum_{i=1}^{k}(-1)^{i+1}\binom{k}{i}=1$$

time, by the second formula in (1.1). So in the final tally, every assignment is counted the right number of times, which means that the total number of ways in which at least one person gets the right coat is

$$n! - \binom{n}{2}(n-2)! + \binom{n}{3}(n-3)! - \ldots + (-1)^i \binom{n}{1} + (-1)^{i+1}$$
$$= n!\Big(1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^i \frac{1}{(n-1)!} + (-1)^{i+1}\frac{1}{n!}\Big) = \Big[\frac{n!}{e}\Big].$$

where $[x]$ is the integer closest to x. The last expression follows since the sum is just the first terms in the Taylor expansion of e^{-1}. So the probability asked for is $\big[\frac{n!}{e}\big]\frac{1}{n!}$.

1.4. **The Pigeon-hole principle.** The pigeon-hole principle is very simple: it says that if $f\colon M \to N$ is a mapping of finite sets M, N, and M has more elements than N, then f is not an injection, i.e. there exists $m_1, m_2 \in M$ such that $f(m_1) = f(m_2)$. Stated another way, if you put $n+1$ letters in n pigeon-holes, then at least one pigeon-hole will contain at least two letters. Of course the difficulty in using this principle is in determining what the pigeon-holes and letters are in a particular problem.

2. Abstract algebra

typed by Peter Hästö
based on notes by Alexei Roudakov

2.1. **Groups.**

2.1.1. *Abstract groups.* We start our study of algebra by defining an operation which can serve for instance as a model of addition. Let G be a set and $\circ\colon G \times G \to G$. We say that $(G, \circ)$ is a *group* if

(1) there exists a neutral element $e \in G$, i.e. $a \circ e = e \circ a = a$ for all $a \in G$;
(2) every element has an inverse, i.e. for every $a \in G$ there exists $b \in G$ such that $a \circ b = e$;
(3) the operation $\circ$ is associative, i.e. $a \circ (b \circ c) = (a \circ b) \circ c$ for all $a, b, c \in G$.

The inverse of the element a is denoted by a^{-1}.

Note that in general we have $a \circ b \neq b \circ a$, i.e. the group is *non-commutative.* However, if $a \circ b = e$, then also $b \circ a = e$ (prove this). Often a group is written multiplicatively, which means that we denote the operation by $\cdot$ instead of $\circ$, and use the usual conventions of omitting this sign, so $a \cdot b$ is written ab. We say that a group is *commutative* or *abelian* if $ab = ba$ for every pair of elements.

Some examples of groups are: $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $\big((0, \infty), \cdot\big)$. For a natural number k we can define a group $\mathbb{Z}_k$ whose elements are the numbers 1, …, k with the operation $+$ defined as addition modulo k. These groups are all commutative

We can also construct new groups from old. For instance, let G be a group and S be a set. Consider the set $M = \{f \big| f\colon S \to G\}$. This set is made to a group by defining the element $f \circ_M g$ to be the function from S to G which maps $a \in S$ to $f(a) \circ_G g(a)$. We call this a point-wise evaluation.

Exercise: Let G be a finite group with the property that $g^2 = g \cdot g = e$ for every $g \in G$. Show that the number of elements of G is a power of 2.

2.1.2. *Permutation groups.* Let next S be a finite set. Then a bijection of S to itself is called a *permutation.* Note that the composition of two permutations is a permutation. We denote the group of permutations of $\{1, \dots, n\}$ by S_n. We use the notation $(a_1\, a_2\, \dots\, a_k)$ for the permutation π with $\pi(a_i) = a_{i+1}$, $\pi(a_k) = a_1$ and $\pi(a) = a$ for all other elements. Permutations of the type $(a\, b)$ are called *transpositions.*

Exercise: Let $\pi \in S_n$ and $\pi = \sigma_1 \cdots \sigma_k = \tau_1 \cdots \tau_m$, where the sigmas and taus are transpositions. Show that $(-1)^k = (-1)^m$.

We can represent a permutation graphically in $\mathbb{R}^2$ as follows: for $\pi \in S_n$ connect $(i, 1)$ with $(j, -1)$ if and only if $\pi(i) = j$. The connections should be by a smooth curve which stays between the lines $(x, 1)$ and $(x, -1)$. (The curves should be drawn so that two curves are never tangent, at most two curves meet in any one point, and a curve never intersects itself.) Let l be the number of intersection points in such a graph.

Exercise: Show that $(-1)^l$ depends only on π, not on how we draw the diagram. Thus we may define $\varepsilon(\pi) = (-1)^l$.

Mappings that preserve the natural structure in algebra are called *homomorphisms* (with appropriate prefix when necessary, e.g. group homomorphism). Explicitly, let G and H be groups and $f\colon G \to H$ be a mapping. We say that f is a homomorphism if

(1) $f(e_G) = e_H$;
(2) $f(a \circ_G b) = f(a) \circ_H f(b)$, for all $a, b \in G$.

EXERCISE: Show that if f is a homomorphism, then $f(a^{-1}) = f(a)^{-1}$.

The set $\{-1, 1\}$ is a group under normal multiplication, with 1 as neutral element.

EXERCISE: Let $\phi\colon S_n \to \{-1, 1\}$ be a homomorphism. Show that ϕ is the identity, or $\phi = \varepsilon$ defined in Exercise 2.1.2.

2.1.3. *Thoughts on problems with groups.* Since groups carry so little structure, they are quite elusive. However, certain special classes of groups, e.g. abelian groups which are finitely generated, are quite easy to understand.

One way of specifying some structure of a group, without giving away too much of its abstractness, is by specifying a relation. For instance, say we have $a^2 = e$ for every element of the group. These kind of problems have been quite popular in the competitions, since their solutions do not require (or benifit from) extensive knowledge of algebra.

Relations are particularly important when constructing groups from free groups. We say that G is a *free group* with generators A satisfying the following property: if F is a group and $\phi\colon A \to F$ is a mapping, there ϕ extends uniquely to a homomorphism of G. An example should clarify this – let $G = \mathbb{Z}$ and $A = \{1\}$. Now the mapping $\phi\colon A \to F$ extends to $\mathbb{Z}$ uniquely by $\phi(k) = k\phi(1)$ (we use $\phi(k) = \phi(k-1) + \phi(1) = k\phi(1)$ and proof by induction).

We can construct a free group as follows: we take a set $\{a_1, \dots, a_n\}$, which will be our generators. Our set is made up of finite sequences (usually called words) $(b_1, \dots, b_l)$, where each b_i is either a_j or a_j^{-1} for some j but the sequence should be without a_j and a_j^{-1} as consecutive elements. The operation on this set is that of concatenation, i.e.

$$(b_1, \dots, b_l)(c_1, \dots, c_l) = (b_1, \dots, b_l, c_1, \dots, c_l),$$

with the understanding that we cancel any a_j with a a_j^{-1} that comes next to it, e.g.

$$(a, b^{-1}, a, c)(c^{-1}, a, a) = (a, b^{-1}, a, a, a).$$

This defines a group, and the group is freely generated by $\{a_1, \dots, a_n\}$.

Once we have a free group G generated by a set $\{a_1, \dots, a_n\}$, we can construct other groups by requiring that a relation hold. A relation is just an element of G "set to" 1, the neutral element. For instance, in $\mathbb{Z}$ we can define the relation $k = 1$: Then we get the group $\mathbb{Z}_k$ defined previously. As another example consider the free group defined as in the previous paragraph by generators a and b, and require the relation $ab = ba$ (equivalently, $aba^{-1}b^{-1} = 1$). Then we can write the elements in our groups in a "normal form", with all the a's preceeding all the b's, so our elements are of the form $(a, \dots, a, b, \dots, b)$, which is more conveniantly written ar $a^k b^l$. In particular, this group is isomorphic to $\mathbb{Z}^2$.

2.2. **Other structures.** Groups are very flexible, but correspondingly capture only very little structure. In this section we have a look at two other algebraic strucutes, this time ones with two operations.

We say that $(R, +, \cdot)$ is a *ring* if

(1) $(R, +)$ is a commutative group;
(2) there exists a multiplicative identity 1, i.e. $1a = a1 = a$ for all $a \in R$;
(3) multiplication is associative, i.e. $a(bc) = (ab)c$ for all $a, b, c \in R$;
(4) the distributive law holds, i.e. $a(b+c) = ab+ac$ and $(a+b)c = ac+bc$ for all $a, b, c \in R$.

Notice that addition in a ring is commutative, but multiplication need not be.

The standard example of a (non-commutative) ring is the set of matrices. Of course $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ and $(\mathbb{C}, +, \cdot)$ are all rings. Another example is given by the set of polynomials over a ring, where addition and multiplication are defined in a point-wise sense. This means that we just calculate with polynomials as we always have done, only that the coefficients are now in some ring.

We can put more structure on our algebraic object by requiring that also multiplication is as close as possible to a commutative group. We say that $(F, +, \cdot)$ is a *field* if

(1) $(F, +, \cdot)$ is a ring;
(2) $(F \setminus \{0\}, \cdot)$ is a commutative group

The sets $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ and $(\mathbb{C}, +, \cdot)$ are also rings. For a prime number p, calculation modulo p in $\mathbb{Z}_p$ is also a field. If k is a composite number (i.e. k has non-trivial divisors), then $\mathbb{Z}_k$ is a ring, but not a field: suppose $k = mn$, where m and n are integers larger than one. Viewing m and n as elements in $\mathbb{Z}_k$, we see that $mn = 0$, so that the elements cannot have multiplicative inverses (required in a field). More generally, we see that $a \in \mathbb{Z}_k$ is invertible if and only if $(a, k) = 1$, that is, the numbers are relatively prime.

3. Polynomials over $\mathbb{C}$

by Eugenia Malinnikova

3.1. **Definition, Ring of polynomials.**

3.1.1. *Polynomials as functions.* By a polynomial we mean a function of the form

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + ...a_1 x + a_0.$$

If $a_n \neq$ then n is called the degree of the polynomial P. To define the class of functions we should specify:

- the domain of definition,
- what coefficients are allowed.

First, we consider polynomials with real coefficients that are defined on the real line. We say that a is a root of the polynomial P if $P(a) = 0$.

Clearly, a polynomial is a smooth function (i.e. it is continuous together with all its derivatives). We will use the standard notation $C^\infty(\mathbb{R})$ for the class of all smooth functions on real line $\mathbb{R}$. Polynomials of degree less than or equal to n can be characterized in the following way:

$$\mathcal{P}_n = \{f \in C^\infty(\mathbb{R}) : f^{(n+1)} \equiv 0\}.$$

Let P be a polynomial of degree n then P is equal to its nth Taylor's polynomial. From calculus we know that

$$a_k = \frac{P^{(k)}(0)}{k!}, \qquad P(x) = a_n x^n + o(x^n)\ (|x| \to \infty).$$

Remark 3.1. Sometimes it is convenient to rewrite a polynomial as a linear combination of the powers of $x - x_0$ for some particular point x_0

$$P(x) = \sum_{k=1}^{n} b_k (x - x_0)^k.$$

Clearly, $b_0 = P(x_0)$ and $P(x_0) = 0$ if and only if there exists a polynomial Q such that $P(x) = (x - x_0)Q(x)$.

Remark 3.2. Theorems from calculus can be applied for polynomials as they are smooth (in particular, continuous) functions.

Exercise: Let P be a polynomial.
a) Prove that P has a real root, provided that the degree of P is odd.
b) Suppose that P has k distinct real roots, show that its derivative has at least $k - 1$ distinct real roots.

Exercise: a) Prove that each polynomial can be written as the difference of two positive polynomials.
b) Prove that each polynomial can be written as the difference of two decreasing polynomials.

3.1.2. *Polynomial ring.* Let R be a ring. A polynomial $P(X)$ over R is defined as a formal expression of the form

$$P(X) = a_n X^n + a_{n-1} X^{n-1} + \ldots + a_1 X + a_0,$$

where the coefficients $a_n, \ldots, a_0$ are elements of the ring R and X is considered to be a formal symbol. Summands in the formula above, $a_k X^k$, are called monomials. Polynomials over R can be added by simply adding the corresponding coefficients as elements of R, and multiplied using the distributive law and the rules $aX = Xa$ for any $a \in R$ and $X^k X^l = X^{k+l}$.

EXERCISE: Prove that the set of polynomials over R with these operation is a ring. This ring is denoted by $R[X]$.

3.1.3. *Division of polynomials.* In this section we consider polynomials over a field F (one may assume that F is equal to $\mathbb{R}$ (real numbers) or $\mathbb{C}$ (complex numbers)).

Let S and T be two polynomials over F, $T \neq 0$. Then there exist unique polynomials Q and R such that $S = QT + R$ and $\deg R < \deg T$ (division with remainder).

If $R = 0$ we say that T divides S.

EXERCISE: Divide $P(X) = X^3 + 3X^2 - 2X + 1$ with remainder by
a) $X + 2$
b) $X^2 - X + 2$.

A polynomial is called irreducible if it is not divisible by any non-constant polynomial of lesser degree. We will describe irreducible polynomial over $\mathbb{R}$ and $\mathbb{C}$ in the next section.

3.2. **Polynomials over complex numbers.**

3.2.1. *Roots of quadratic polynomials and complex numbers.* Let $P(x) = ax^2 + bx + c$ be a quadratic polynomial with real coefficients. As we know from school, the number of its real roots is defined by the sign of the expression $b^2 - 4ac$. If it is non-negative then the real roots are given by the formula

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Let us consider the polynomial $x^2 + 1$. It has no real roots. We will add a formal symbol i to the field of real numbers and consider the set of formal expressions of the form $x + iy$, where $x, y \in \mathbb{R}$. This set with two operations

$$(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2)$$

$$(x_1 + iy_1)(x_2 + iy_2) = (x_1 y_1 - x_2 y_2) + i(y_1 x_2 + x_1 y_2)$$

is a field, we denote it by $\mathbb{C}$, elements of $\mathbb{C}$ are called complex numbers. If $z = x + iy$ is a complex number then x is called the real part of z, y is called the imaginary part of z, and $x - iy$ is called the complex conjugate of z. We will use the following notation $x = \Re(z), y = \Im(z), x - iy = \bar{z}$.

EXERCISE: Check that each quadratic polynomial with *complex* coefficients has two complex roots (that may coincide).

Another standard interpretation of complex numbers comes from identifying pairs (x, y) with points of the plane. If we then consider the polar coordinates, we get

$$x + iy = r\cos\phi + ir\sin\phi = r(\cos\phi + i\sin\phi) =: re^{i\phi}.$$

The advantage of the polar form is that the powers of a complex number are easily calculated in the polar form:

$$(re^{i\phi})^n = r^n e^{in\phi}$$

EXERCISE: Find all roots of the polynomial $z^n - 1$.

3.2.2. *Fundamental theorem of algebra and irreducible polynomials.*

Theorem 3.3. *Any polynomial with complex coefficients has a complex root.*

Corollary 3.4. *Any polynomial P of degree n over $\mathbb{C}$ has exactly n (not necessarily distinct) complex roots and can be written in the form*

$$P(z) = a_n(z-\zeta_1)\dots(z-\zeta_n). \tag{3.5}$$

It follows from the theorem that all polynomials over $\mathbb{C}$ of degree greater than one are reducible. So irreducible polynomials in $\mathbb{C}[x]$ are constants and polynomials of degree one.

Exercise: a) Let P be a polynomial with real coefficients. Prove that if ζ is a root of P, then so is $\bar{\zeta}$.
b) Describe all irreducible polynomials over $\mathbb{R}$.

It follows from the last exercise that any polynomial P with real coefficients can be written in the form:

$$P(x) = a_n \prod_{k=1}^{m}(x-a_k)\prod_{l=1}^{s}(x^2+p_s x+q_s),$$

where a_k are the real roots of the polynomial and quadratic polynomials $x^2+p_s x+q_s$ have no real roots.

Exercise: Let P be a polynomial over $\mathbb{R}$ such that $P(x)\geq 0$ for any $x\in\mathbb{R}$. Prove that there exist two polynomials Q and R over $\mathbb{R}$ such that $P=Q^2+R^2$.
(Hint: Use the statement above and the following identity $(a^2+b^2)(c^2+d^2)=(ac+bd)^2+(ad-bc)^2$.)

3.2.3. *Roots of the derivative.* Let A be a set of complex numbers, we consider elements of A as points on the plane. The *convex hull* of A is the intersection of all half-planes that contain A. In other words, the convex hull of A is the smallest convex set that contains A. If A is a finite set, then its convex hull is a polygon.

The following result on the roots of the derivative of a polynomial is both beautiful and useful.

Theorem 3.6. *Let $\zeta_1,\dots,\zeta_n$ be all roots of the polynomial $P\in\mathbb{C}[z]$. Then all roots of P' lie in the convex hull of $\{\zeta_1,\dots,\zeta_n\}$.*

Proof. Using the definition of the convex hull, we see that it suffices to prove that if all roots of P lie in a half-plane E, then all roots of P' lie in E. By a linear change of variables, $Q(z)=P(\alpha z+\beta)$, we see that it is enough to prove this when E is the upper half-plane, i.e. $E=\{z=x+iy\colon y\geq 0\}$.

Suppose that $\Im(\zeta_k)\geq 0$ for $k=1,\dots,n$. As we know from the previous subsubsection, $P(z)=a(z-\zeta_1)\dots(z-\zeta_n)$. We take the derivative of the product of n functions

$$P'(z) = a\sum_{k=1}^{n}(z-\zeta_1)\dots\widehat{(z-\zeta_k)}\dots(z-\zeta_n),$$

where the hat over a factor means that the factor is omitted. Now the ratio P'/P can be written in a simple way

$$\frac{P'(z)}{P(z)} = \sum_{k=1}^{n}\frac{1}{z-\zeta_k}. \tag{3.7}$$

Assuming that $\Im(\eta) < 0$, we will show that $P'(\eta) \neq 0$. We apply the above formula

$$\frac{P'(\eta)}{P(\eta)} = \sum_{k=1}^{n} \frac{1}{\eta - \zeta_k} = \sum_{k=1}^{n} \frac{\overline{\eta - \zeta_k}}{|\eta - \zeta_k|^2},$$

and take the imaginary parts

$$\Im\left(\frac{P'(\eta)}{P(\eta)}\right) = \sum_{k=1}^{n} \Im\left(\frac{\bar{\eta} - \bar{\zeta}_k}{|\eta - \zeta_k|^2}\right) = \sum_{k=1}^{n} \frac{-\Im(\eta) + \Im(\zeta_k)}{|\eta - \zeta_k|^2} > 0.$$

It follows that $P'(\eta) \neq 0$ and the theorem is proved. □

One can get formula (3.7) by taking the derivative of $\log P(z)$. This formula has many applications, in particular, it is useful in the next exercise.

EXERCISE: Let P be a polynomial of degree n that has n distinct real roots $x_1, \dots, x_n$. Prove that
a) $\sum_{k=1}^{n} \frac{P''(x_k)}{P'(x_k)} = 0$,
b) for any $c > 0$ the set $\{x \in \mathbb{R} : \frac{P'(x)}{P(x)} > c\}$ is a union of finitely many disjoint intervals of total combined lengths $\frac{n}{c}$.

3.3. **Different ways to determine polynomial.**

3.3.1. *Roots and Coefficients.* Any polynomial over $\mathbb{C}$ is determined by its leading coefficient and its roots. (We use the convention here that every polynomial of degree n has n complex roots and formula (3.5) is valid; some of the roots may coincide.) On the other hand each polynomial is determined by its coefficients. The relations between the roots and the coefficients are described below.

Theorem 3.8. (Viet) *Suppose that polynomial $P(z) = \sum_{k=0}^{n} a_k z^k$ has roots $\zeta_1, \dots, \zeta_n$ then*

$$a_0 = (-1)^n a_n \prod_{j=1}^{n} \zeta_j, \quad a_1 = (-1)^{n-1} a_n \sum_{k=1}^{n} \zeta_1 \dots \hat{\zeta_k} \dots \zeta_n,$$

$$\vdots$$

$$a_{n-2} = a_n \sum_{j \neq l} \zeta_j \zeta_l, \quad a_{n-1} = -a_n \sum_{k=1}^{n} \zeta_k.$$

In other words, a_k is equal to $(-1)^{n-k} a_n$ times the sum of all products of k elements of $\{\zeta_1, \dots \zeta_n\}$.

EXERCISE: Let $P(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1 x + a_0$ be a polynomial with real coefficients. Suppose that P has n distinct positive roots. Prove that

$$(-a_{n-1})^n > n^n |a_0|.$$

3.3.2. *Polynomial interpolation.* Each polynomial of degree less than $n+1$ is determined by its values at $n+1$ distinct points. In fact, suppose that two polynomials P_1 and P_2 coincide at $n+1$ points and $\deg P_1 \le n, \deg P_2 \le n$. Then the difference $P_1 - P_2$ is a polynomial of degree less than $n+1$ that has $n+1$ distinct roots, and so $P_1 - P_2 = 0$.

Theorem 3.9 (Lagrange interpolation formula)**.** *Let P be a polynomial of degree less than $n+1$ and $P(z_j) = a_j, j = 1, \ldots, n+1$. Then*

$$P(z) = \sum_{j=1}^{n+1} a_j \frac{(z-z_1)\ldots\widehat{(z-z_j)}\ldots(z-z_{n+1})}{(z_j-z_1)\ldots\widehat{(z_j-z_j)}\ldots(z_j-z_{n+1})}.$$

It is easy to see that the above polynomial satisfies $P(z_j) = a_j$ and we noted before that there exists unique polynomial of degree less than $n+1$ that solves the interpolation problem.

Exercise: (Newton interpolation) There is another way to write down the same polynomial:

$$P(z) = b_1 + (z - z_1)\left(b_2 + (z - z_2)\left(\ldots(b_{n-1} + (z - z_n)b_n)\ldots\right)\right),$$

where coefficient b_k depend on $a_1, \ldots a_k$ and $x_1, \ldots, x_k$.

a) Write down formulas for b_1 and b_2.

b) Find Newton and Lagrange interpolation polynomials for the data $P(0) = 3, P(1/2) = 4, P(1) = 6$.

4. Linear algebra

by Peter Hästö

4.1. **Linear, or vector, spaces.** We start by recalling the basic definitions of vector spaces. Algebraically, a vector space is made up of a commutative group $(V,\cdot)$, a field $(K,+,\cdot)$ and a scalar multiplication $\cdot\colon K\times V\to V$ which satisfy the following properties:

(1) $a(u+v)=au+av$;
(2) $(a+b)v=av+bv$;
(3) $a(bv)=(ab)v$; and
(4) $1v=v$.

Here $a,b\in K$, $u,v\in V$, 1 is the unit element of K and 0 is the zero element of V and K. In this case we say that V *is a vector space over* K.

Some examples of groups: any field is a vector space over itself; K^n is a vector space over K, for any field; n-by-n matrices over a field K, and more generally linear transformations of a field K; $\mathbb{R}$ is a vector space over $\mathbb{Q}$.

Some further important terms (V is a vector space over K):

- a vector subspace of V is a subset V' of V such that (V',K) is a vector space in the inherited operations;
- v is a *linear combination* of $\{v_i\}_i$ if $v=\sum a_iv_i$, where $a_i\in K$;
- the *span* of a set $\{v_i\}_i$ is the smallest subspace containing all the vectors, or, equivalently, the set of all finite linear combinations of vectors in $\{v_i\}_i$;
- $\{v_i\}_i$ is *linearly independent* if any vector in the span of $\{v_i\}_i$ can be written as a linear combination in exactly one way, or, equivalently, if the equation $\sum a_iv_i=0$ has only the solutions $a_1=\ldots=a_k=0$;
- $\{v_i\}_i$ is a *basis* for V if the set spans V and is linearly independent;
- the dimension of V is the number of elements in a basis of V.

4.2. **Linear operators.** Another very important concept related to vector spaces, emphasized by the alternative name linear spaces, is that of linear operators. If V and V' are vector spaces on K, then we say that the operator $T\colon V\to V'$ is linear if

$$T(au+bv)=aT(u)+bT(v)$$

for all $a,b\in K$ and $u,v\in V$. Recall that if we have chosen bases for finite vector spaces V and V', then we may represent a linear transformation by a matrix.

Some important words here are (V,V' are vector spaces over K, T is a linear operator from V to V'):

- the *image* of T is $T(V)$;
- the *kernel* of T is $T^{-1}(0)$, i.e. all those elements of V that map to 0;
- for finite dimensional V', the rank of T is the dimension of the image of T.

To understand the action of a linear operator, it is useful to single out some vectors for this the operation is especially simple. This leads us to Eigenvectors. We say that $v\in V\setminus\{0\}$ is an *Eigenvector* corresponding to the *Eigenvalue* $\lambda\in K$ if $T(v)=\lambda v$.

A "Euclidean style" descriptions of the procedure for determining the Eigenvalues and vectors follows. Recall the following procedure for calculating the eigenvectors of the matrix

$$A=\begin{pmatrix}7&2\\-2&2\end{pmatrix}:$$

First, calculate

$$\det(A-\lambda I)=\det\begin{pmatrix} 7-\lambda & 2 \\ -2 & 2-\lambda \end{pmatrix}=(7-\lambda)(2-\lambda)+4=(\lambda-3)(\lambda-6).$$

This gives the eigenvalues $\lambda_1=3$ and $\lambda_2=6$.

Second, look at the first eigenvalue, 3. Solve the equation $(A-3I)v_1=0$:

$$\begin{pmatrix} 4 & 2 \\ -2 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 4x+2y \\ -2x-y \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Hence $v_1=\begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

Then, look at the second eigenvalue, 6. Solve the equation $(A-6I)v_2=0$:

$$\begin{pmatrix} 1 & 2 \\ -2 & -4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} x+2y \\ -2x-4y \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Hence $v_2=\begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

Why does this procedure give all the eigenvectors of the matrix A, i.e. all non-zero vectors such that $Av=\lambda v$?

Problems.

(1) Find a basis for the image and the kernel of

$$\begin{pmatrix} 1 & 3 \\ 3 & 4 \end{pmatrix}$$

considered as a linear transformation of a two-dimensional vector space
  a) over $\mathbb{R}$,
  b) over $\mathbb{Z}_5$.

(2) Let V be the vector space whose elements are arithmetic sequences, $(a, a+k, a+2k, a+3k, \ldots)$, $a, k \in \mathbb{R}$. Find a basis of this space and thus determine its dimension. Give an example of a one-dimensional subspace.

(3) Let $L^1(\mathbb{R}^n)$ be the set of all continuous functions $f\colon \mathbb{R}^n \to \mathbb{R}$ for which

$$\int_{\mathbb{R}^n} |f(x)|\,dx<\infty.$$

Show that $L^1(\mathbb{R}^n)$ is infinite dimensional by finding a subspace with infinitely many basis vectors.

(4) Define $\mathbb{R}^+=\{x>0\}$. Show that $\left(\mathbb{R}^+\right)^2$ is not a subspace of the vector space $\mathbb{R}^2$. Let V be the set of 2-vectors over $\mathbb{R}^+$, i.e. the elements of V are vectors of the form $\begin{pmatrix} a \\ b \end{pmatrix}$, $a, b \in \mathbb{R}^+$. We define an addition on V by

$$\begin{pmatrix} a \\ b \end{pmatrix}\oplus\begin{pmatrix} c \\ d \end{pmatrix}=\begin{pmatrix} ac \\ bd \end{pmatrix}$$

and a scalar multiplication by

$$k\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} a^{\log k} \\ b^{\log k} \end{pmatrix}.$$

We define addition and multiplication in $\mathbb{R}^+$ by $l\oplus k=lk$ (i.e. addition corresponds to ordinary multiplication) and $l\odot k=\exp\{\log(l)\log(k)\}$. Show that V is a vector space over $\mathbb{R}^+$.

(5) Let $P_0 = \{(0, y, z)\}$ and $P_1 = \{(x, y, z) \colon x + y + z = 0\}$. Let $A \colon \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation such that $A(P_k) = P_k$. Show that the dimension of the kernel of A is 0 or 1. Give examples which show that both cases can occur.

(6) Let V and W be vector spaces. Suppose that $T \colon V \to W$ is a linear transformation with the following properties:

a) if $T(x) = T(y)$ then $x = y$ (i.e. T is one-to-one, or injective)

b) for any $w \in W$ there exists $v \in V$ such that $T(v) = w$ (i.e. T is onto, or surjective).

Show that T is an isomorphism.

(7) Consider the field $\mathbb{Q}$ of rational numbers (i.e. fractions of integers). We consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$.

(a) Find a 2-dimensional subspace of $\mathbb{R}$.

(b) Show that the equation $f(x + y) = f(x) + f(y)$ has a solution **not** of the form $f(x) = cx$ on this subspace.

4.3. **Determinants.** In this section we recall one abstract way of defining the determinant of a matrix.

Let $v_1, \dots v_n$, x and y be row vectors with n elements (in some field K, e.g. $\mathbb{R}$). Then det is defined as the unique function which takes n-tuples of vectors with n components, that is elements of K^n, and satisfies the following conditions:

- det is *multilinear*, i.e.
$$\det(ax + by, v_2, \dots, v_n) = a \det(x, v_2, \dots, v_n) + b \det(y, v_2, \dots, v_n),$$
where $a, b \in K$ and the same holds for all other vectors v_i.
- det is antisymmetric, i.e.
$$\det(v_1, v_2, \dots, v_n) = -\det(v_2, v_1, \dots, v_n),$$
and the same for all other swaps of vectors (not necessarily adjacent).
- $\det(I_n) = 1$.

Let us start by checking that det is uniquely determined by these conditions, i.e. that there are not two different functions satisfying them. We make the following observations:

Lemma 4.1. *If $v_i = v_j$ for $i \neq j$, then* $\det(v_1, \dots, v_n) = 0$.

Proof. We use property (2) of the determinant to swap v_i and v_j:
$$\det(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -\det(v_1, \dots, v_j, \dots, v_i, \dots, v_n).$$
But since $v_i = v_j$ we obviously have
$$\det(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = \det(v_1, \dots, v_j, \dots, v_i, \dots, v_n).$$
From this the claim directly follows. □

Corollary 4.2. *Let $a \in K$. Then*
$$\det(v_1, \dots, v_i, \dots, v_n) = \det(v_1, \dots, v_i + av_j, \dots, v_n).$$

Proof. By Property (1) of the determinant,
$$\det(v_1, \dots, v_i + av_j, \dots, v_n) = \det(v_1, \dots, v_i, \dots, v_n) + a \det(v_1, \dots, v_j, \dots, v_n).$$
But by the previous lemma, the latter term is zero. □

Lemma 4.3. *If $v_i = 0$, then* $\det(v_1, \dots, v_n) = 0$.

Proof. Let $a \in \mathbb{R}$, $a \neq 1$. By Property (1),

$$\det(v_1, \dots, av_i, \dots, v_n) = a \det(v_1, \dots, v_i, \dots, v_n).$$

Since $v_i = 0$ it follows trivially that $av_i = v_i$. Therefore the two determinants in the previous equation are the same. Thus the equations is

$$(a-1)\det(v_1, \dots, v_i, \dots, v_n) = 0.$$

Since $a - 1 \neq 0$, this implies the claim. □

We consider the matrix A whose i^{th} row is given by v_i. We denote $\det(v_1, \dots, v_n)$ also by $\det(A)$.

Theorem 4.4. *The function* $\det \colon (V^n)^n \to \mathbb{R}$ *is uniquely determined by the conditions (1)–(3).*

Proof. We consider the matrix A whose i^{th} row is given by v_i. Let us say that A is brought to row-reduced echelon form by performing k row swaps, and by dividing the rows by the constants k_1, …, k_n (look up the description of the algorithm to produce the row-reduced echelon form in the book, if necessary). Then each of the row-swaps produces a factor -1 in front of the determinant (by Property (2)) and division of a row by k_i produces a factor of k_i in front (by Property (1)). Subtracting a multiple of the row containing the pivotal element has no effect of the determinant, by Corollary 4.2.

There are then two possibilities: if the rref is I_n, then

$$\det(A) = (-1)^k k_1 \cdots k_n \det(I_n) = (-1)^k k_1 \cdots k_n.$$

If, on the other hand, rref is not I_n, then it contains at least one zero-row (since A is n-by-n). Then, by Lemma 4.3, $\det(A) = 0$. We have thus put forth an algorithm for computing the determinant of any matrix A – clearly this implies that the determinant function is uniquely defined. □

In the proof of the previous theorem we saw one of the fundamental properties of the determinant:

Corollary 4.5. *The determinant of A is* 0 *if and only if A is not invertible.*

Proof. The square matrix A is invertible if and only if its rref is I_n. In the previous proof we saw that this is precisely the case when A is invertible. □

Also of great importance is the following product formula for the determinant:

Theorem 4.6. *Let A and B be n-by-n square matrices. Then* $\det(AB) = \det(A)\det(B)$.

Proof. Suppose first that B is not invertible. Then $\det(A)\det(B) = 0$, so we must show that $\det(AB) = 0$, as well. Let x be a non-zero vector such that $Bx = 0$ (exists, since B is not invertible). Then $ABx = A(Bx) = A0 = 0$, and so AB is not invertible (since an invertible matrix does not take a non-zero vector to zero).

If B is invertible, we can consider the function

$$D(A) = \frac{\det(AB)}{\det(B)}.$$

We can check that this function satisfies the conditions (1)–(3) above. Since the determinant was the unique function satisfying these conditions, this implies that $D(A) = \det(A)$. □

The following results follow directly by applying the formula in the previous theorem.

Corollary 4.7. *Let A be an invertible matrix. Then* $\det(A^{-1}) = (\det(A))^{-1}$.

Corollary 4.8. *Let A and B be similar n-by-n matrices. Then* $\det(B) = \det(A)$.

Proof. There exists an invertible matrix S such that $B = S^{-1}AS$ (by the definition of begin similar). Thus

$$\det(B) = \det(S^{-1}AS) = \det(S^{-1})\det(A)\det(S) = \det(S)^{-1}\det(S)\det(A) = \det(A). \quad \square$$

If T is a linear operator, we know that its matrix in two different bases are similar. In view of the previous corollary, it makes sense to define the determinant of this operator to be the determinant of any of these matrices.

Using the multilinearity of the determinant (Property (1)), one can derive the so-called Laplace expansion for the determinant of a matrix.

4.4. **Representation of a matrix.** We view a matrix as representing a linear transformation from a (finite dimensional) vector space to itself when we have chosen a bases for the spaces. If we choose different bases, then the matrix will be different. However, the matrices for the transformation are similar, i.e. $A' = SAS^{-1}$ for some invertible matrix S. This leads to the question of choosing S so that A' is as simple as possible for a given A. In many senses the *Jordan normal form* is this simplest representation. We know that not every matrix can be diagonalized. The Jordan normal form is in some sense the closest you can get to a diagonal matrix. In order to present the proper context for the JNF, we need to introduce some concepts.

4.4.1. *Decomposition.* Let V and W be vector spaces. Their *direct sum*, $V \oplus W$, is defined as pairs of vectors (v, w), $v \in V$, $w \in W$ with the following addition:

$$(v_1, w_1) \oplus (v_2, w_2) := (v_1 + v_2, w_1 + w_2).$$

(Note that the additions on the right-hand-side are in the vector spaces V and W.)

Exercise: Show that $V \oplus W$ is a vector space.

Let S and T be linear operators on V and W, respectively. Then we define $S \oplus T$ by

$$S \oplus T(v \oplus w) := (S(v), T(w)).$$

Exercise: Show that $S \oplus T$ is linear operator.

Let T be a linear operator on the finite vector space V. We say that the subspace $V' \subset V$ is *invariant under* T if $T(V') \subset V'$. If $\dim(V) = n$ and $\dim(V') = m$ then T is represented by a block matrix of the form

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix},$$

where A is m-by-m and C is $(n-m)$-by-$(n-m)$. In other words the fact that we know an invariant subspace gives us significant structural information about the matrix of the transformation. Suppose next that T happens to be such a linear operator as to have V' and V'' as invariant subspaces. Suppose further that $V' \cap V'' = \emptyset$ and $V = V' \oplus V''$. Then the matrix of T in a suitably chosen basis is

$$\begin{pmatrix} A & O \\ 0 & C \end{pmatrix},$$

where A and C have the same dimensions as before.

4.4.2. *Jordan normal form.* We next define some sort of an iterated Eigenvector: we say that v lies in the *root space* $R(\lambda)$ of the linear operator T if $(V - \lambda I)^m v = 0$ for some m. Clearly, if v is an Eigenvector corresponding to the Eigenvalue λ then $v \in R(\lambda)$. It turns out that V can be decomposed into its root-spaces:

Theorem 4.9. *Let T be a linear operator on the finite dimensional vector spaces V. Suppose T has Eigenvalues $\lambda_1, \dots, \lambda_k$. Then*

$$V = R(\lambda_1) \oplus \dots \oplus R(\lambda_k).$$

In view of what was said in the previous section, this implies that T is represented by a matrix of block-diagonal type, with one block for each distinct Eigenvalue. It turns out that we can even do much better than this.

Theorem 4.10 (Jordan Normal Form)**.** *Let T be a linear operator on the finite dimensional vector spaces V. Then T has a representation in block diagonal form where every block has the form*

$$\begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ & & \dots & & \\ 0 & 0 & \dots & 0 & \lambda \end{pmatrix},$$

where the diagonal entries are Eigenvalues of T and the entries immediately above the diagonal are 1's.

Note that the matrix given in the previous theorem is as a whole of the form with Eigenvalues on the diagonal and 1's and 0's immediately above the diagonal. Thus T is represented as the sum of a diagonal operator and a nilpotent operator. (Recall that a linear operator T on the vector space V is said to be *nilpotent* if there exists $k > 0$ such that $T^k(v) = 0$ for every $v \in V$.)

Exercise: Let V be the space of functions of the form $f(x) = e^x p(x)$, where $p(x)$ is polynomial of degree at most $n-1$. Find a basis of V in which the matrix of the operator d/dx is the $n \times n$ Jordan block with λ on the diagonal.

Problems.

(1) Let $A \in M_{\mathbb{R}}(n,n)$. What is the smallest value of n if $A^{m+1} = 0$, but $A^m \neq 0$. Give an example when $n = 3$.

(2) Let $A \in M_{\mathbb{R}}(n,m)$ and $B \in M_{\mathbb{R}}(m,n)$. If AB is invertible, what can you say about the relationship between n and m?

(3) Let A and B be square matrices. If $AB + A + B = 0$, show that A and B commute, i.e. that $AB = BA$.

(4) The following are some properties of matrix multiplication:
- $AI = IA = I$,
- $(AB)C = A(BC)$, and
- $AB \neq BA$,

where A, B, C, I are n by n matrices, and I denotes the identity. Describe how matrix multiplication is defined (i.e. the logic of the definition), and why it implies these properties.

(5) Among all unit vectors

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$$

find the one for which $ax + by + cz$ is maximal, where a, b and c are nonzero constants. Can you interpret this result in terms of the dot product?

(6) Consider the vector space V of all symmetric 2-by-2 real matrices which have eigenvector $v_1 = \binom{1}{1}$ (the matrix may or may not have other eigenvectors). We make it an inner product space by defining

$$\langle A, B \rangle = \mathrm{tr}(A^T B),$$

where tr denotes the trace, i.e. the sum of the diagonal entries. Find an orthogonal basis of V.

(7) Find all inner products on the vector space $\mathbb{R}^2$. (Hint: Recall that every inner product is represented by a matrix, but not the other way around.)

Find all complex inner products on $\mathbb{C}^2$. (Hint: Recall that every complex inner product is represented by a complex matrix, but not the other way around.)

(8) In constructing a model of a certain kind of fluid a complicated equation arises involving traces and determinants of symmetric matrices. In order to show that one of the terms is negligible compared to the others, the inequality

$$\operatorname{tr}(A) \geq n \det(A)^{1/n}$$

is needed. Show that it holds for **all** symmetric n-by-n real matrices A. Recall that the trace of A, $\operatorname{tr}(A)$, is the sum of the elements on the diagonal.

(9) Consider vectors in $\mathbb{R}^2$ with vectors $v_1 = (x_1, y_1)$ and $v_2 = (x_2, y_2)$. Define a bilinear form by

$$\rho(v_1, v_2) = x_1x_2 + x_1y_2 + x_2y_1 + kx_2y_1.$$

For which values of k is this (a) symmetric, (b) bilinear and (c) an inner product.

(10) Define inner products spaces such that the Cauchy-Schwarz inequality allows you to affirm the following inequalities:

$$\left(\sum_{i=1}^{n} v_i w_i\right)^2 \leq \left(\sum_{i=1}^{n} v_i^2\right)\left(\sum_{i=1}^{n} w_i^2\right).$$

$$\left(\int_a^b f(x)g(x)\,dx\right)^2 \leq \left(\int_a^b f(x)^2 dx\right)\left(\int_a^b g(x)^2 dx\right).$$

(11) Consider the space of second degree polynomials. Define an inner product by the formula

$$\langle p, q\rangle = \int_a^b p(x)q(x)\,dx.$$

For which choises, if any, of $a, b \in \mathbb{R}$ is the standard basis $\{1, x, x^2\}$ orthogonal?

(12) Let V be an inner product space. Suppose v is orthogonal to each of $w_1, \ldots, w_k$. Show that v is orthogonal to $\operatorname{span}\{w_1, \ldots, w_k\}$. Show also that any set of orthogonal vectors is linearly independent.

(13) Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of a Euclidean space. Prove Bessel's inequality

$$\sum_{i=1}^{k} \langle v, e_i\rangle^2 \leq \|v\|^2,$$

and Parseval's identity

$$\langle v, w\rangle = \sum_{i=1}^{n} \langle v, e_i\rangle\langle e_i, w\rangle.$$

5. Matrix calculus

by Torbjørn Helvik

5.1. **A matrix is a linear map.** A $m \times n$ matrix A defines a linear map $\mathbb{R}^n \to \mathbb{R}^m$. The vector x is mapped to the vector Ax. Conversely, every linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by a $m \times n$ matrix A such that $f(x) = Ax$. Note that A depends on the choice of basis for $\mathbb{R}^n$. When not stating otherwise we use the standard euclidian basis. Composition of linear maps are given by matrix products: If $g : \mathbb{R}^m \to \mathbb{R}^k$ is represented by B then

$$(g \circ f)(x) = BAx$$

There is a simple way of finding the matrix that represents a map f in a given basis. Assume that $f : \mathbb{R}^n \to \mathbb{R}^m$ and let $(e_i)_{i=1}^n$ constitute a basis for $\mathbb{R}^n$. Write $x = \sum_i x_i e_i$. Then

$$f(x) = f(\sum_i x_i e_i) = \sum_i x_i f(e_i) = Ax$$

with the i'th column of A being $f(e_i)$.

Exercise: Find the matrices in the standard basis representing the following operations on $\mathbb{R}^3$:

(1) Reflection through the $x-y$ plane
(2) Rotation by $\pi/2$ in the $x-y$ plane, followed by a rotation by $-\pi/2$ in the $y-z$ plane
(3) Projection onto the plane defined by $x = y$.

Let A be a $m \times n$ matrix. The rank of A is the number of linearly independent column vectors of A. This is equal to the number of linearly independent row vectors of A. For the rest of this section, define $k = \text{rank}(A)$

We now look at some important subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ related to A. Recall that for any map $f : X \to Y$

$$\text{Im } f = \{y \in Y : \exists x \in X \text{ with } f(x) = y\} = f(X)$$
$$\ker f = \{x \in X : f(x) = 0\} = f^{-1}(0)$$

We first look at the image of A. Taking the product Ax is the same as taking a linear combination of the columns of A. In fact if we write the columns of A as $\{v_1, v_2, \dots, v_n\}$, then $Ax = \sum_i x_i v_i$.

Therefore,

$$\text{Im } A = \text{span}\{v_1, \dots, v_n\} \subseteq \mathbb{R}^m$$

This is a k-dimensional linear subspace of $\mathbb{R}^m$. Let us denote the orthogonal component of Im A by X, such that $\mathbb{R}^m = \text{Im } A \oplus X$. We claim that $X = \ker A^T$ (recall that the kernel is the subspace consisting of all vectors that is mapped to 0 by the matrix). Indeed, if $y \in \text{Im } A$ and $x \in \ker A^T$, then

$$z^T y = z^T (Ax) = (A^T z) x = 0$$

One can also show that $\dim(\ker A^T) = m - k$. The same relation is of course valid for Im A^T and $\ker A$.

To sum up, the four fundamental subspaces of the matrix A are connected in the following way

$$\mathbb{R}^n = \text{Im} A^T \oplus \ker A \qquad \text{(dimensions } k \text{ and } n-k)$$
$$\mathbb{R}^m = \text{Im} A \oplus \ker A^T \qquad \text{(dimensions } k \text{ and } m-k)$$

The first pair is invariant under row operations on A and the second pair under column operations.

Let A be a square matrix. If there exists a matrix B such that $AB = I$ we call B the inverse of A and write $B = A^{-1}$. In this case, we say that A is invertible (or non-singular). It is also the case that $BA = I$.

The following are equivalent

- A is invertible
- $\det(A) \neq 0$
- 0 is not an eigenvalue of A
- $\mathrm{rank}(A) = n$

The inverse of a 2×2 matrix is given by the formula

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - cb} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Inverses of larger matrices can be found by Gauss-Jordan elimination.

Exercise: Let A and B be square matrices. If $AB + A + B = 0$, show that $AB = BA$.

5.2. **The characteristic polynomial.** The characteristic polynomial $c_A(x)$ of the square $n \times n$ matrix A is defined as

$$c_A(x) = \det(A - xI)$$

The roots of c_A are the eigenvalues of A. The multiplicity of λ as a root in $c_A(x)$ is the algebraic multiplicity of λ. The geometric multiplicity of λ is the dimension of the eigenspace corresponding to λ (that is, the dimension of $\ker(A - \lambda I)$). The geometric multiplicity is never larger than the algebraic multiplicity.

Lemma 5.1. *The geometric multiplicity of $\lambda = 0$ is $n - k$ iff rank$(A) = k$.*

Exercise: Prove this

Write

$$c_A(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$$

The coefficients a_{n-1} and a_0 of c_T are particularly useful:

$$a_{n-1} = \mathrm{Trace}(A) = \sum \lambda_i a_0 = \mathrm{Det}(A) = \prod \lambda_i$$

Similar matrices have the same characteristic polynomial:

Lemma 5.2. *If $B = S^{-1}AS$, with S invertible, then $c_B(x) = c_A(x)$*

Also:

Lemma 5.3. *If A and B are square, $c_{AB}(x) = c_{BA}(x)$.*

One can look at polynomials of matrices. The following result can be useful:

Lemma 5.4. *If p is a polynomial such that $p(A) = 0$, then $p(\lambda) = 0$ for all eigenvalues λ of A.*

Exercise: Prove this.

The Cayley-Hamilton theorem says that any matrix satisfies its own characteristic equation:

Theorem 5.5 (Cayley-Hamilton)**.**

$$c_A(A) = 0$$

Exercise: a) Show that for any $m \in \mathbb{N}$ there exists a real $m \times m$ matrix A such that $A^3 = A + I$.

b) Show that $\det A > 0$ for every real $m \times m$ matrix satisfying $A^3 = A + I$.

5.3. **Diagonalization.** Let A be a $n \times n$ matrix with n independent eigenvectors $\{x_1, \ldots, x_n\}$. Let S be the matrix with these eigenvectors as column vectors, and let Λ be a diagonal matrix with the eigenvalues $\lambda_1, \ldots, \lambda_n$ on the diagonal. Then

$$A = S\Lambda S^{-1}$$

This is called the diagonal form of A. A is not diagonalizable if it has less than n independent eigenvectors. But note that repeated eigenvalues is no problem as long as their algebraic multiplicity equals their geometric multiplicity. Also, there is no connection between invertibility and diagonalizability.

Lemma 5.6. *A real symmetric matrix can be diagonalized by a orthogonal matrix $S^T S = I$, and a Hermitian matrix can be diagonalized by a unitary matrix ($S^H S = I$).*

Lemma 5.7. *If A and B are diagonalizable, they share the same eigenvector matrix S iff $AB = BA$.*

Note that $A^n = S\Lambda^n S^{-1}$, and Λ^n is found simply by squaring all elements since Λ is diagonal. This is an important use of the diagonal form.

If some eigenvalues has lower geometric multiplicity than algebraic, A is not diagonalizable. The best we can do in this case is the Jordan form. If A has s independent eigenvectors, it is similar to a matrix with s blocks:

$$M^{-1}AM = J = \begin{bmatrix} J_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & J_s \end{bmatrix}$$

Each block is a triangular matrix with a single eigenvalue λ_i and a single eigenvector:

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \cdot & \cdot & \\ & & \cdot & 1 \\ & & & \lambda_i \end{bmatrix}$$

The same eigenvalue λ_i may appear in several blocks if it corresponds to several independent eigenvectors.

Exercise:

Let A and B be square matrices of the same size. If

$$\mathrm{rank}(AB - BA) = 1,$$

show that $(AB - BA)^2 = 0$.

5.4. **A few more tricks.** The following is general trick that can be helpful. In a matrix multiplication AB, one can divide A into submatrices of the size $n \times m$ and B into submatrices of size $m \times n$, and do the multiplication as if these were the elements of the matrix. This is perhaps best communicated by an example. Let A to F be $n \times n$ matrices. Then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} E \\ F \end{bmatrix} = \begin{bmatrix} AE + BF \\ CE + DF \end{bmatrix}$$

Also, note that if A is a matrix with row vectors $(a_1^T, a_2^T, \ldots, a_n^T)$ and B is a matrix with column vectors $(b_1, b_2, \ldots, b_n)$, then the rows of AB are $(a_1^T B, \ldots, a_n^T B)$ and the columns of AB are $(Ab_1, \ldots, Ab_n)$. This can be utilized to construct matrixes that perform operations such as permutations on the rows or columns of a matrix C.

5.5. **Non-negative matrices.** NB: This section is not important for the competition, but the Perron-Frobenius theorem is a strong result that one should know about.

Let A be a $k \times k$ matrix. We say A is non-negative if $a_{ij} \geq 0 \, \forall \, i, j$.

Definition 5.8. Let A be a non-negative matrix, and let $a_{ij}^{(n)}$ denote the (i, j)-elements of A^n. We say A is

- *Irreducible* if for any pair i, j there is some $n > 0$, such that $a_{ij}^{(n)} > 0$.
- Irreducible and *aperiodic* if there is some $n > 0$, such that $a_{ij}^{(n)} > 0$ for all pairs i, j.

Theorem 5.9 (Perron-Frobenius Theorem)**.** *Let A be a non-negative $k \times k$ matrix.*

(1) A has a real eigenvalue $\lambda \geq 0$, and $\lambda \geq |\mu|$ where μ is any other eigenvalue. If A is irreducible then $\lambda > 0$ and if A also is aperiodic, then $\lambda > |\mu|$.
(2) We have $\min_i(\sum_{j=1}^k a_{ij}) \leq \lambda \leq \max_i(\sum_{j=1}^k a_{ij})$
(3) λ has a non-negative eigenvector u, and a non-negative left eigenvector v.
(4) If A is irreducible, then λ is a simple root of the characteristic polynomial, and the corresponding eigenvectors are strictly positive.
(5) If A is irreducible, then λ is the only eigenvalue of A with a non-negative eigenvector.
(6) If u and v are the right and left eigenvector, normalized such that $v \cdot u = 1$, then

$$\lim_{n\to\infty} \frac{1}{\lambda^n} A^n = uv$$

Note that his implies that any stochastic matrix has 1 as the largest eigenvalue λ (a matrix is stochastic if all row sums are 1).

by EUGENIA MALINNIKOVA

6.1. Sequences and limits.

6.1.1. *Basic facts and examples.* We begin with some simple rules and theorems that are useful when calculating limits:

(1) Each monotone sequence has a limit (either finite or infinite).
(2) Let $\{x_n\}_n, \{a_n\}_n, \{b_n\}_n$ be sequences of real numbers such that $a_n \le x_n \le b_n$ for each n. If the limits $\lim_{n\to\infty} a_n$ and $\lim_{n\to\infty} b_n$ exist and are equal, then $\lim_{n\to\infty} x_n$ exists and

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

(3) A sequence $\{x_n\}_n$ has a limit if and only if $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n$, where

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup_{k\ge n} x_k \quad \text{and} \quad \liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf_{k\ge n} x_k.$$

(4) If the sequence $\{x_n\}_n$ is defined by $x_n = f(n)$ for some function f and $\lim_{x\to\infty} f(x)$ exists, then $\lim_{n\to\infty} x_n$ exists and is equal to $\lim_{x\to\infty} f(x)$.

EXERCISE: Calculate $\lim_{n\to\infty}(\sqrt{n^2+n}-n)$.

As usual in such kind of problems we should prove that the limit exists and find it. Here are several solutions that use a number of simple rules listed above.
SOLUTION 1: We have

$$\sqrt{n^2+n}-n = \frac{n}{\sqrt{n^2+n}+n} = \frac{1}{\sqrt{1+1/n}+1}.$$

Then clearly, $\lim_{n\to\infty}(\sqrt{n^2+n}-n) = \frac{1}{2}$.
SOLUTION 2: It is easy to check that $n + \frac{n-1}{2n} < \sqrt{n^2+n} < n + \frac{1}{2}$. Thus we have $\frac{n-1}{2n} < \sqrt{n^2+n} - n < \frac{1}{2}$. Now, clearly, the first and the last terms converge to $\frac{1}{2}$ when n goes to ∞. Applying the "squeezing" rule we get the limit exists and $\lim_{n\to\infty}(\sqrt{n^2+n}-n) = \frac{1}{2}$

EXERCISE: Let $\{a_n\}$ be a seguence such that $a_1 = 1, a_{n+1} > \frac{3}{2}a_n$.
a) Prove that the sequence $\{\left(\frac{2}{3}\right)^{n-1} a_n\}$ has a finite limit or tends to infinity.
b) For any $\alpha > 1$ there exists a sequence with these properties such that

$$\lim_{n\to\infty} \left(\frac{2}{3}\right)^{n-1} a_n = \alpha.$$

SOLUTION: a) The sequence $\{\left(\frac{2}{3}\right)^{n-1} a_n\}$ is increasing since $a_{n+1} > \frac{3}{2}a_n$. Thus it either has a finite limit or tends to infinity.
b) For any $\alpha > 1$ there exists a sequance $\{b_n\}$ such that $b_1 = 1, b_{n+1} > b_n$ and b_n tends to α when n goes to infinity. (Can you write down a formula that defines such a sequence?) Now let $a_n = \left(\frac{3}{2}\right)^{n-1} b_n$. $\square$

Sometimes it is easier to prove that the limit exists first, and then use that information to calculate the limit. Moreover we may assume that the limit exists, then calculate it and finally prove that the assuption is true.

EXERCISE: We consider the Fibonacci sequence $F_0 = 1,\ F_1 = 1,\ F_{n+1} = F_n + F_{n-1}$. Calculate $\lim_{n\to\infty} \frac{F_{n+1}}{F_n}$.

Solution: It is clear that the sequence is bounded, $1 \leq \frac{F_{n+1}}{F_n} \leq 2$, if we calculate the first few terms we will see that it is not monotone. Suppose that this sequence has a limit ϕ. Then

$$\phi = \lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \lim_{n\to\infty} \frac{F_n + F_{n-1}}{F_n} = \lim_{n\to\infty} (1 + \frac{F_{n-1}}{F_n}) = 1 + \frac{1}{\phi}.$$

Thus ϕ is equal to the positive solution of the quadratic equation $x^2 - x - 1 = 0$, i.e. $\phi = \frac{1}{2}(1+\sqrt{5})$.

Now we want to estimate $|\frac{F_{n+1}}{F_n} - \phi| = \left|\left(1 + \frac{F_{n-1}}{F_n}\right) - \left(1 + \frac{1}{\phi}\right)\right|$. We denote $a_n = \frac{F_{n+1}}{F_n}$, then we have

$$|a_n - \phi| = \left|\frac{1}{a_{n-1}} - \frac{1}{\phi}\right| = \frac{|\phi - a_{n-1}|}{a_{n-1}\phi} < \frac{|\phi - a_{n-1}|}{\phi}.$$

Thus $|a_n - \phi|$ tends to zero as n goes to infinity and we get $\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \phi$. □

Problems.

(1) Calculate $\lim_{n\to\infty} \frac{1^n + 2^n + \ldots + n^n}{n^n}$.

(2) Let $x_1 = \sqrt{a},\ x_{n+1} = \sqrt{a + x_n}$ for $n \geq 1$, where $a \geq 1$. Find $\lim_{n\to\infty} x_n$.

(3) Let $x_1 = 2005,\ x_{n+1} = \frac{1}{4 - 3x_n}$ for $n \geq 1$. Find $\lim_{n\to\infty} x_n$.

(4) Let $a_1 = 1,\ a_n = \frac{1}{n}\sum_{k=1}^{n-1} a_k a_{n-k}$ for $n \geq 2$. Show that
(i) $\limsup_{n\to\infty} |a_n|^{1/n} < 2^{-1/2}$,
(ii) $\limsup_{n\to\infty} |a_n|^{1/n} \geq \frac{2}{3}$.

(5) Find $\lim_{N\to\infty} \frac{\ln^2 N}{N} \sum_{k=2}^{N-2} \frac{1}{\ln k \ln(N-k)}$.

6.2. **Series.**

6.2.1. *Basic facts on convergence.*

(1) The series $\sum_{n=1}^{\infty} a_n$ converges if and only if for any $\varepsilon > 0$ there exists N such that

$$\left|\sum_{k=n}^{m} a_k\right| \leq \varepsilon,$$

whenever $m \geq n \geq N$.

(2) Let $\alpha = \limsup_{n\to\infty} |a_n|^{\frac{1}{n}}$. If $\alpha < 1$ then the series $\sum_{n=1}^{\infty} a_n$ conveges, if $\alpha > 1$ then the series $\sum_{n=1}^{\infty} a_n$ diverges.

(3) If $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$, then $\sum_{n=1}^{\infty} a_n$ converges.

(4) If $a_n > a_{n+1} > 0$ and $\lim_{n\to\infty} a_n = 0$, then the series $\sum_{n=1}^{\infty}(-1)^n a_n$ converges.

Exercise: Does there exist a bijective map $\pi : \mathbb{N} \to \mathbb{N}$ such that $\sum_{n=1}^{\infty} \frac{\pi(n)}{n^2} < \infty$?

Hint: Consider $t_k = \sum_{n=3^k+1}^{3^{k+1}} \frac{\pi(n)}{n^2}$, take terms with $\pi(n) \geq 3^k$ and prove that $t_k \geq \frac{1}{9}$ for each k.

Exercise: Find out if the series $\sum_{n=1}^{\infty}(-1)^n a_n$ converges or diverges, where
a) $a_n = \frac{10^n}{n!}$, b) $a_n = \frac{2^n n!}{n^n}$, c) $a_n = \frac{\sin(n\pi/4)}{\ln n}$.

6.2.2. *Integral test, Series as Riemann sums.* The simplest version of the integral test, that is used very often to determine if a series is convergent or divergent, is the following:
If $f : [1,\infty) \to [0,\infty)$ is a non-increasing function then $\sum_{n=1}^{\infty} f(n) < \infty$ if and only if $\int_1^{\infty} f(x)dx < \infty$.

In particular, we use that test to determine for which values of p series $\sum \frac{1}{n^p}$, $\sum \frac{1}{n \ln^p n}$ converge. Sometimes this test is convinient to combine with some prior estimates of the terms of a series.

EXERCISE: Let $\nu(n)$ be the number of digits of (the decimal representation) of n. Find out if the series $\sum \frac{\nu(n)}{n^2}$ converges.

SOLUTION: We note that $\nu(n) = k$ means $10^{k-1} \leq n < 10^k$. Therefor $\nu(n) \leq (\ln 10)^{-10} \ln n + 1$ and $\sum \frac{\nu(n)}{n^2} < \infty$. □

In general, it might be useful to write down a series as a Riemann some for an integral. (Unformal rule, if you see $\lim \sum a_n$, try to compaire $\sum a_n$ to some integral.)

EXERCISE: Find $\lim_{t \nearrow 1}(1-t)\sum \frac{t^n}{1+t^n}$.

HINT: $\sum \frac{t^n}{1+t^n} = \sum \frac{e^{n \ln t}}{1+e^{n \ln t}} = (-\ln t)^{-1} \sum (e^{-n \ln t} + 1)^{-1}$.

6.2.3. *Summation by parts.* Let $A_n = \sum_{k=1}^{n} a_k$, then

$$\sum_{n=1}^{m} a_n b_n = \sum_{1}^{m-1} A_n(b_n - b_{n+1}) + A_m b_m.$$

To prove the above formula write $a_n = A_n - A_{n-1}$, rewrite the sum as two sums as change the summation index in the second sum to combine terms with A_n. This calculation explains two following tests:

The series $\sum_{n=1}^{\infty} a_n b_n$ converges if
(i) $\sum_{n=1}^{\infty} a_n < \infty$ and $\{b_n\}$ is a monotone bounded sequence,
or
(ii) $A_n = \sum_{k=1}^{n} a_k$ are bounded and the sequence $\{b_n\}$ is monotone and tends to zero.

EXERCISE: Prove that $\sum_{n=1}^{\infty} \frac{\sin n}{n} < \infty$.

HINT: Take $a_n = \sin n,\ b_n = 1/n$ and prove that $s_n = a_1 + ... + a_n$ are bounded.

EXERCISE: Suppose that $b_1 \geq b_2 \geq b_3 \geq ...$ and $\lim_{n\to\infty} b_n = 0$. Show that $\sum_{n=1}^{\infty} 2^n b_n < \infty$ if and only if $\sum_{n=1}^{\infty} 2^n(b_n - b_{n+1}) < \infty$.

HINT: $\sum_{n=1}^{N} 2^n b_n = \sum_{n=1}^{N-1}(2^{n+1}-2)(b_n - b_{n+1}) + (2^{N+1}-2)b_N$, prove first that $\sum_{n=1}^{\infty} 2^n(b_n - b_{n+1}) < \infty$ implies $\lim_{N\to\infty} 2^N b_N = 0$.

6.2.4. *Series summation.* General methods

(1) If $a_n = b_{n+1} - b_n$ and $b = \lim_{n\to\infty} b_n$ then $\sum_{n=1}^{\infty} a_n = b - b_1$.
(2) If $\sum_{n=1}^{\infty} a_n < \infty$ then $\sum_{n=1}^{\infty} a_n = \lim_{x\nearrow 1} \sum_{n=1}^{\infty} a_n x^n$.
(3) To find a sum of the form $\sum_{n=1}^{\infty} a_n \cos nx$ or $\sum_{n=1}^{\infty} a_n \sin nx$ it is useful to write this sum as a real or imaginary part of the sum $\sum_{n=1}^{\infty} a_n z^n$.

EXERCISE: Find $\sum_{n=1}^{\infty} a_n$, where
a) $a_n = \frac{1}{n(n+m)}$, b) $a_n = \frac{1}{n2^n}$, c) $a_n = \frac{n^2}{n!}$, d) $a_n = \frac{\sin nx}{n}$.

SOLUTION: a) $a_n = \frac{1}{m(n+m)} - \frac{1}{mn}$, $\sum_{n=1}^{\infty} a_n = \frac{1}{m}(1 + \frac{1}{2} + \frac{1}{3} + ... \frac{1}{m})$.
b)

$$\sum_{n=1}^{\infty} a_n = \lim_{x \nearrow 1} \sum_{n=1}^{\infty} \frac{x^n}{n 2^n} = \lim_{x \nearrow 1} \sum_{n=1}^{\infty} \int_0^x \frac{t^{n-1}}{2^n} dt = \lim_{x \nearrow 1} \int_0^x \frac{1}{2-t} dt = \ln 2.$$

Justify the calculation!
c) $a_n = \frac{n}{(n-1)!}$, we write the series as a power series once again, but now nx^{n-1} is the derivative of x^n:

$$\sum_{n=1}^{\infty} a_n = \lim_{x \nearrow 1} \sum_{n=1}^{\infty} \frac{nx^{n-1}}{(n-1)!} = \lim_{x \nearrow 1} (\sum_{n=1}^{\infty} \frac{x^n}{(n-1)!})' = \lim_{x \nearrow 1} (xe^x)' = 2e.$$

d) Let $z = e^{ix}$, then $\sin nx = \Im z^n$ and

$$\sum_{n=1}^{\infty} a_n = \Im \left(\sum_{n=1}^{\infty} \frac{z^n}{n} \right) = \Im(-\ln(1-z)) = -\Im(\ln(1-e^{ix}).$$

Further, we have $1 - e^{ix} = e^{ix/2}(e^{-ix/2} - e^{ix/2}) = -ie^{ix/2} \sin x/2$ and

$$-\Im(\ln(1-e^{ix}) = -\ln(-ie^{ix/2}) = \frac{\pi - x}{2}.$$

We get

$$\sum_{n=1}^{\infty} \frac{\sin nx}{n} = \frac{\pi - x}{2}.$$

Check that all calculations may be justified for $x \in (0, \pi]$, find out why this argument does not work for $x = 0$. $\square$

Problems.

(1) Suppose that $a_n \geq 0$ and $\sum_{n=1}^{\infty} a_n < \infty$. Prove that $\sum_{n=1}^{\infty} \frac{\sqrt{a_n}}{n} < \infty$.

(2) Suppose $\sum_{n=1}^{\infty} a_n < \infty$. Do the following sums have to converge as well?
a)$a_1 + a_2 + a_4 + a_3 + a_8 + a_7 + a_6 + a_5 + a_{16} + a_{15} + ... + a_9 + a_{32} + ...$,
b)$a_1 + a_2 + a_3 + a_4 + a_5 + a_7 + a_6 + a_8 + a_9 + a_{11} + a_{13} + a_{15} + a_{10} + a_{12} + a_{14} + a_{16} + ...$

(3) Suppose that $a_n > 0$, $\sum_{n=1}^{\infty} a_n = \infty$ and $s_n = a_1 + ... + a_n$. Prove that
a) $\sum_{n=1}^{\infty} \frac{a_n}{a_n + 1} = \infty$,
b) $\sum_{n=1}^{\infty} \frac{a_n}{s_n} = \infty$,
c) $\sum_{n=1}^{\infty} \frac{a_n}{s_n^2} < \infty$.

(4) Let $\{\varepsilon_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers such that $\lim_{n \to \infty} \varepsilon_n = 0$. Find $\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \ln \left(\frac{k}{n} + \varepsilon_n \right)$.

(5) (i) Prove that $\lim_{x \to \infty} \sum_{n=1}^{\infty} \frac{nx}{(n^2+x)^2} = \frac{1}{2}$.
(ii)Prove that there is a positive constant c such that for every $x \in [1, \infty)$ we have

$$\left| \sum_{n=1}^{\infty} \frac{nx}{(n^2+x)^2} - \frac{1}{2} \right| \leq \frac{c}{x}.$$

(6) For which x the series $\sum_{n=1}^{\infty} \ln \left(1 + \frac{(-1)^n}{n^x} \right)$ converges?

(7) Show that $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}\sin(\ln n)}{n^{\alpha}}$ converges if and only if $\alpha > 0$

(8) Find out if the series $\sum_{n=1}^{\infty} \frac{\sin n \sin n^2}{n}$ converges or not.

(9) Let $b_0 = 1$,

$$b_n = 2 + \sqrt{b_{n-1}} - 2\sqrt{1 + \sqrt{b_{n-1}}}.$$

Calculate $\sum_{n=1}^{\infty} 2^n b_n$.

(10) Calculate the sums of the following series $\sum_{n=1}^{\infty} \frac{cosnx}{n}$.

7. Problem sets

NOTE: Many of the following problems were taken from sources such as the International Mathematics Competition for University Students, or the Putnam Competition. No claim of originality is made on the parts of the authors, and the problems are included for educational purposes only.

7.1. Basics.

7.1.1 Let n be a fixed positive integer. How many ways are there to write n as a sum of positive integers, $n = a_1 + a_2 + \ldots + a_k$, with k an arbitrary positive integer and $a_1 \leq a_2 \leq \ldots \leq a_k \leq a_1 + 1$ For example, with 4 there are four ways: 4, $2+2$, $1+1+2$, $1+1+1+1$.

7.1.2 Let S be an infinite set of real numbers such that $|s_1 + s_2 + \ldots + s_k| < 1$ for every finite subset $\{s_1, s_2, \ldots, s_k\} \subset S$. Show that S is countable.

7.1.3 a) Show that the unit square can be partitioned into n smaller squares if n is large enough.
b) Let $d > 2$. Show that there is a constant $N(d)$ such that, whenever $n \geq N(d)$, a d-dimensional unit cube can be partitioned into n smaller cubes.

7.1.4 Let $A_n = \{1, 2, \ldots n\}$, where $n \geq 3$. Let $\mathcal{F}$ be the family of all non-constant functions $f \colon A_n \to A_n$ satisfying the following conditions:
(1) $f(k) \leq f(k+1)$ for $k = 1, 2, \ldots n-1$;
(2) $f(k) = f(f(k+1))$ for $k = 1, 2, \ldots n-1$.
Find the number of functions in $\mathcal{F}$.

7.1.5 Let r, s, t be positive integers which are pairwise relative prime. Suppose that a and b are elements of a commutative multiplicative group with unit e such that $a^r = b^s = (ab)^t = e$. Show that $a = b = e$.

7.1.6 Let X be a set of $\binom{2k-4}{k-2} + 1$ real numbers, $k \geq 2$. Prove that there exists a monotone sequence $(x_i)_{i=1}^k \subset X$ such that

$$|x_{i+1} - x_1| \geq 2|x_i - x_1|$$

for all $i = 1, \ldots, k-1$.

7.2. Abstract algebra.

7.2.1 Let G be a group and a be an element in G. The minimal positive integer n such that $a^n = e$ is called the order of element a and is denoted $\operatorname{ord} a$. Show that for any $a \in G$
a: if $\gcd(k, \operatorname{ord} a) = 1$, then $\operatorname{ord}(a^k) = \operatorname{ord} a$;
b: if $\operatorname{ord}(a) = 2$, then $\operatorname{ord}(aba) = \operatorname{ord} b$ for any $b \in G$;
c: $\operatorname{ord}(ab) = \operatorname{ord}(ba)$ for any $b \in G$.

7.2.2 Consider permutations of $\{1, ..., n\}$. They form a group with composition as the operation, which is denoted S_n. As usual $\pi \in S_n$ and $\pi = (a_1, ..., a_k)$ means $\pi(a_i) = a_{i+1}, i = 1, ..., k-1, \pi(a_k) = a_1$ and $\pi(b) = b$ for $b \neq a_1, ..., a_k$.
Let k be odd and $a_1, ...a_{k-1}, b_1, ..., b_k, ..., c_1, ..., c_{k+1}$ be different elements of $\{1, ..., n\}$. Show that

$$\operatorname{ord}\left((a_1, ...a_{k-1})(b_1, ..., b_k)(c_1, ..., c_{k+1})\right) = \frac{(k-1)k(k+1)}{2}.$$

7.2.3 Let G be a group with a fixed element a. We consider a graph with the vertex set G and for each $x \in G$ we put an edge between x and ax. Show that this graph is a union of cycles of size $n = \operatorname{ord} a$.

7.2.4 Let G be the subgroup of $GL_2(\mathbb{R})$ generated by A and B, where

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Let H consist of all matrices $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ in G for which $a_{11} = a_{22} = 1$.

a) Show that H is an abelian subgroup of G.

b) Show that H is not finitely generated.

7.2.5 Suppose that in a not necessarily commutative ring R the square of any element is 0. Prove that $abc + abc = 0$ for any three elements a, b, c.

7.2.6 Let $a_1, a_2, .., a_{51}$ be non-zero elements of a field. We simultaneously replace each element with the sum of the 50 remainings ones. In this way we get a sequence $b_1, ..., b_{51}$. If the new sequence is a permutation of the original one, what can the characteristic of the field be?

7.3. **Polynomials over $\mathbb{C}$.**

7.3.1 Let $P_0, P_1, ..., P_{n-1}$ be the vertices of a regular n-gon inscribed in the unit circle. Prove that $|P_1P_0||P_2P_0|...|P_{n-1}P_0| = n$.

7.3.2 Let all roots of an nth degree polynomial $P(z)$ with complex coefficients lie on the unit circle in the complex plane. Prove that all roots of the polynomial $2zP'(z) - nP(z)$ lie on the same circle.

7.3.3 Let $P(z)$ be an algebraic polynomial of degree n having only real zeros and real coefficients. **(a)** Prove that $(n-1)\left(P'(x)\right)^2 \geq nP(x)P''(x)$ for every real x. **(b)** Examine the case of equality.

7.3.4 Consider the following set of polynomials:

$$\mathcal{P} = \left\{ f : f = \sum_{k=0}^{3} a_k x^k, a_k \in R, |f(\pm 1)| \leq 1, |f(\pm 1/2)| \leq 1 \right\}.$$

Evaluate $\sup_{f \in \mathcal{P}} \max_{-1 \leq x \leq 1} |f''(x)|$ and find all polynomials $f \in \mathcal{P}$ for which the above supremum is attainded.

7.3.5 Let $p(x) = x^5 + x$ and $q(x) = x^5 + x^2$ find all pairs (w, z) of complex numbers with $w \neq z$ for which $p(w) = p(z)$ and $q(w) = q(z)$.

7.3.6 Let $p(z)$ be a polynomial of degree $n \geq 1$ with complex coefficients. Prove that there exist at least $n + 1$ complex numbers z for which $p(z)$ is 0 or 1.

7.4. **Linear algebra.**

7.4.1 Find the maximum of $x^3 - 3x$ on the set $\{x \colon x^4 + 36 \leq 13x^2\}$.

7.4.2 Let S be a set of real numbers which is closed under multiplication, i.e. $a, b \in S$ implies $ab \in S$. Let T and U be disjoint subsets of S whose union is S. Given that the product of any three (not necessarily distinct) elements of T is in T and that the product of any three elements of U is in U, show that at least one of the two subsets T, U is closed under multiplication.

7.4.3 Let V be a real vector space, and let $f, f_1, f_2 \dots f_k$ be linear maps from V to $\mathbb{R}$. Suppose that $f(x) = 0$ whenever $f_1(x) = f_2(x) = \dots = f_k(x) = 0$. Prove that f is a linear combination of $f_1, f_2, \dots, f_k$.

7.4.4 Consider the field $\mathbb{Q}$ of rational numbers (i.e. fractions of integers). We consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$.

(a) Find a 2-dimensional subspace of $\mathbb{R}$.

(b) Show that the equation $f(x + y) = f(x) + f(y)$ has a solution **not** of the form $f(x) = cx$ on this subspace.

7.4.5 Let $A \colon \mathbb{R}^3 \to \mathbb{R}^3$ be linear. Suppose that Av and v are orthogonal for every $v \in V$.

(a) Show that $A^T = -A$.

(b) Show that there exists $u \in \mathbb{R}^3$ such that $Av = u \times v$.

7.4.6 Recall that a function A is an involution if A^2 is the identity. Let V be a finite dimensional real vector space.

(a) Let $A: V \to V$ be an involution. Show that there exists a basis of V consisting of Eigenvectors of V.

(b) Find the largest number of pairwise commutative involutions on V.

7.4.7 Let A and B be complex matrices which satisfy the equation $AB + A + B = 0$, where 0 is the additive neutral element. Show that $AB = BA$.

7.4.8 Let $\alpha \in \mathbb{R} \setminus \{0\}$. Suppose that F and G are linear maps (operators) from $\mathbb{R}^n$ into $\mathbb{R}^n$ satisfying $F \circ G - G \circ F = \alpha F$.

a) Show that $F^k \circ G - G \circ F^k = \alpha k F^k$ for all $k \in \mathbb{N}$.

b) Show that $F^k = 0$ for some k.

7.4.9 Let A be an $n \times n$ diagonal matrix with characteristic polynomial

$$(x - c_1)^{d_1}(x - c_2)^{d_2} \cdots (x - c_k)^{d_k},$$

where c_1, c_2, $\dots$, c_k are distinct (which means that c_1 appears d_1 times on the diagonal, c_2 appears d_2 times on the diagonal, etc., and $d_1 + d_2 + \ldots + d_k = n$). Let V be the space of all $n \times n$ matrices B such that $AB = BA$. Prove that the dimension of V is

$$d_1^2 + d_2^2 + \ldots + d_k^2.$$

7.4.10 For $n \geq 1$ let M be an $n \times n$ complex matrix with distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$, with multiplicities $m_1, m_2, \dots, m_k$, respectively. Consider the linear operator L_M defined by $L_M(X) = MX + XM^T$, for any complex $n \times n$ matrix X. Find its eigenvalues and their multiplicities. (M^T denotes the transpose of M; that is, if $M = (m_{k,l})$, then $M^T = (m_{l,k})$.)

7.4.11 Let A be an $n \times n$ matrix with complex entries and suppose that $n > 1$. Prove that

$$A\overline{A} = I_n \iff \exists S \in GL_n(\mathbb{C}) \quad \text{such that} \quad A = S\overline{S}^{-1}.$$

(If $A = [a_{ij}]$ then $\overline{A} = [\overline{a_{ij}}]$, where $\overline{a_{ij}}$ is the complex conjugate of a_{ij}; $GL_n(\mathbb{C})$ denotes the set of all $n \times n$ invertible matrices with complex entries, and I_n is the identity matrix.)

7.5. **Matrix calculus.**

7.5.1 Compute the determinant of the $n \times n$ matrix $A = [a_{ij}]$, where

$$a_{ij} = \begin{cases} (-1)^{|i-j|} & \text{if } i \neq j \\ 2 & \text{if } i = j \end{cases}$$

7.5.2 Let A be a real 4×2 matrix and B be a real 2×4 matrix such that

$$AB = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

Find BA.

7.5.3 Let M be an invertible $2n \times 2n$ matrix, represented in block form as

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad \text{and} \quad M^{-1} = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

Show that $\det M \cdot \det H = \det A$.

7.5.4 Let X be a non-singular matrix with columns $(v_1, v_2, \dots, v_n)$. Let Y be the matrix with columns $(v_2, v_3, \dots, v_n, 0)$. Show that the matrixes $A = YX^{-1}$ and $B = X^{-1}Y$ have rank $n - 1$ and have only 0's for eigenvalues.

7.5.5 a) Let the mapping $f : M_n \to \mathbf{R}$ from the space of all real $n \times n$ matrices to the reals be linear (ie. $f(A+B) = f(A) + f(B)$ and $f(cA) = cf(A)$). Prove that there exists a unique matrix $C \in M_n$ such that $f(A) = \text{Trace}(AC)$.

b) If in addition $f(AB) = f(BA)$ for any $A, B \in M_n$, prove that there exists $\alpha \in \mathbb{R}$ such that $f(A) = \alpha \cdot \text{Trace}(A)$.

7.5.6 For a $n \times n$ real matrix A, e^A is defined as $\sum_{n=0}^{\infty} \frac{1}{n!} A^n$. Prove or disprove that for all real polynomials p and square matrices A and B, $p(e^{AB})$ is nilpotent if and only if $p(e^{BA})$ is nilpotent. (The matrix A is nilpotent if $A^k = 0$ for some $k \in \mathbb{N}$.)

7.6. **Sequences and series.**

7.6.1 Calculate $\lim\limits_{n\to\infty} \dfrac{1^n + 2^n + ... + n^n}{n^n}$.

7.6.2 Let $x_1 = 2005$, $x_{n+1} = \frac{1}{4-3x_n}$ for $n \geq 1$. Find $\lim_{n\to\infty} x_n$.

7.6.3 Let $a_0 = \sqrt{2}$, $b_0 = 2$,

$$a_{n+1} = \sqrt{2 - \sqrt{4 - a_n^2}}, \quad b_{n+1} = \frac{2b_n}{2 + \sqrt{4 + b_n^2}}.$$

a) Prove that the sequences $\{a_n\}$ and $\{b_n\}$ are decreasing and converge to 0.
b) Prove that the sequence $\{2^n a_n\}$ is increasing, the sequence $\{2^n b_n\}$ is decreasing and those two sequences converge to the same limit.
c) Prove that there is a positive constant c such that for all n the following inequality holds $0 < b_n - a_n < \frac{c}{8^n}$.

7.6.4 Suppose $\sum_{n=1}^{\infty} a_n < \infty$. Do the following sums have to converge as well?
a) $a_1 + a_2 + a_4 + a_3 + a_8 + a_7 + a_6 + a_5 + a_{16} + a_{15} + ... + a_9 + a_{32} + ...$,
b) $a_1 + a_2 + a_3 + a_4 + a_5 + a_7 + a_6 + a_8 + a_9 + a_{11} + a_{13} + a_{15} + a_{10} + a_{12} + a_{14} + a_{16} + ...$

7.6.5 Suppose that $a_n > 0$, $\sum_{n=1}^{\infty} a_n = \infty$ and $s_n = a_1 + ... + a_n$. Prove that
a) $\sum_{n=1}^{\infty} \frac{a_n}{a_n+1} = \infty$,
b) $\sum_{n=1}^{\infty} \frac{a_n}{s_n} = \infty$,
c) $\sum_{n=1}^{\infty} \frac{a_n}{s_n^2} < \infty$.

7.6.6 Let $\{\varepsilon_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers such that $\lim_{n\to\infty} \varepsilon_n = 0$. Find $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \ln\left(\frac{k}{n} + \varepsilon_n\right)$.

7.7. **Calculus.**

7.7.1 Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x,y) = (X^2 - y^2)e^{-x^2-y^2}$.
a) Prove that f attains its minimum and its maximum.
b) Determine all points (x,y) such that $\frac{\partial f}{\partial x}(x,y) = \frac{\partial f}{\partial y}(x,y) = 0$ and determine for which of them f has global or local minimum or maximum.

7.7.2 Find $\lim\limits_{t\to 1-} (1-t) \sum\limits_{n=1}^{\infty} \dfrac{t^n}{1+t^n}$, where $t \to 1-$ means that t approaches 1 from below.

7.7.3 Let f be a continuous function on $[0,1]$ such that for every $x \in [0,1]$ we have

$$\int_x^1 f(t)\, dt \geq \frac{1-x^2}{2}.$$

Show that

$$\int_0^1 f(t)^2\,dt \geq \frac{1}{3}.$$

7.7.4 Suppose n is a natural number. Evaluate

$$\int_{-\pi}^{\pi} \frac{\sin nx}{(1+2^x)\sin x}\,dx$$

.

7.7.5 Let $F\colon (1,\infty) \to \mathbb{R}$ be the function defined by

$$F(x) := \int_x^{x^2} \frac{dt}{\ln t}.$$

Show that F is one-to-one (i.e. injective) and find the range (i.e. set of values) of F.

7.7.6 Prove that

$$\int_0^1 \int_0^1 \frac{dx\,dy}{x^{-1} + |\ln y| - 1} \leq 1.$$

8. Hints

8.1. **Basics.**

8.1.1 Write down occurences for first few k.
8.1.2 How many elements can be larger than $1/k$?
8.1.3 Both parts can be solved by looking at only two types subdivision operations, such as one cube is divided into 2^d cubes.
8.1.4 "Plot" the function. How many fixed-points can f have?
8.1.5 There exist integers u, v such that $ru + sv = 1$.
8.1.6 It is easier to construct such a sequence if you already have a sequence of lenght $k-1\ldots$

8.2. **Abstract algebra.**

8.2.1 You're on your own here, sorry...
8.2.2 Let π be the permutation in question. $\pi^i(a_1) = a_1$ if and only if i is a multiple of $k-1$. Similarly for b_1 and c_1. If π^j fixes these three elements, then it fixes everything.
8.2.3 In the graph you move from x to ax to a^2x to a^3x and so on.
8.2.4 H is isomorphic to (has the same structure as) a subgroup of $\mathbb{R}$. For part b recall that the generators in particular are elements of H.
8.2.5 Look at squares like $(a+b)^2$, $(ab+c^2)$, etc.
8.2.6 What is $\sum b_i - \sum a_i$? Recall that the characteristic is always a prime.

8.3. **Polynomials over $\mathbb{C}$.**

8.3.1 Use complex numbers.
8.3.2 Find the product of the roots of $Q(z) = 2zP'(z) - nP(z)$. Then try to prove that Q has no roots outside the unit circle; the formula for P'/P from the lecture notes could be useful.
8.3.3 It is enough to prove the inequality for $x = 0$.
8.3.4 Try interpolation, for example.
8.3.5 It is not very difficult now, just write down the system and be patient in doing calculations.
8.3.6 If a is a root of p of multiplicity k, then a is a root of p' of multiplicity $k-1$.

8.4. **Linear algebra.**

8.4.1 This is really a second degree problem.
8.4.2 Use four elements, two in each of T and U.
8.4.3 Assume that $f_1, \ldots, f_k$ are linearly independent, and determine what the coefficient of f would have to be.
8.4.4 $\sqrt{2} \notin \mathbb{Q}$.
8.4.5 Check what A does to basis vectors.
8.4.6 A can be diagonalized (see Section 4.2). What are the Eigenvalues of A?
8.4.7 A matrix A always commutes with its...
8.4.8 Use a telescoping sum. For (b), consider eigenvalues of a suitable operator (not of F or G).
8.4.9 Block matrices are useful here.
8.4.10 Construct the eigenvectors of $L_M(X)$ out of the eigenvectors of M and their transposes.
8.4.11 Construct S as a linear combination of A and I_n.