

Further linear algebra. Chapter I. Integers.

Andrei Yafaev

Number theory is the theory of $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$.

1 Euclid's algorithm, Bézout's identity and the greatest common divisor.

- We say that $a \in \mathbb{Z}$ divides $b \in \mathbb{Z}$ iff there exists $c \in \mathbb{Z}$ such that $b = ac$. We write $a|b$.
- A *common divisor* of a and b is an integer d that divides both a and b .
- The *greatest common divisor* (sometimes called the *highest common factor*) of a and b is a common divisor $d > 0$ of a and b such that any other positive common divisor is smaller than d . This is written $d = \gcd(a, b)$.

Every $a \in \mathbb{Z}$ is a divisor of 0, since $0 = 0 \times a$, therefore it makes sense to define $\gcd(a, 0) = a$ if $a > 0$. However $\gcd(0, 0)$ does not exist as any integer is a common factor of 0 and 0.

Two integers a and b are called *coprime* if $\gcd(a, b) = 1$.

Recall that an integer $p > 1$ is called a *prime* if p is only divisible by 1 and by itself. An integer which is not prime is called *composite*.

Lemma 1.1 *Any integer $n > 1$ is divisible by a prime number.*

Proof. By induction. If $n = 2$ then n is obviously divisible by 2 which is prime.

Fix $n > 1$ and suppose the property holds by all integers $< n$. If n is prime, then we can take $p = n$. Suppose n is composite, $n = n_1 n_2$ with $n_1 < n$. By induction assumption, n_1 , and hence n is divisible by a prime number. \square

The following obvious remark : any divisor of $a \geq 0$ is *smaller or equal* to a , is often used in the proofs.

Let $a \geq b > 0$ be two integers. When a is not divisible by b , one can still divide with the *remainder*. For example:

$$a = 5, b = 2, a = 2b + 1$$

$$a = 20, b = 3, a = 6b + 2$$

etc..

This leads to the following (fundamental) theorem.

Theorem 1.2 (Euclidean division) *Let $a \geq b > 0$ be two integers. There exists a UNIQUE pair of integers (q, r) satisfying*

$$a = qb + r$$

and $0 \leq r < b$.

Proof. Two things need to be proved : the existence of (q, r) and its uniqueness.

Let us prove the *existence*.

Consider the set

$$S = \{x, x \text{ integer } \geq 0 : a - xb \geq 0\}$$

The set S is not empty : 1 belongs to S . The set S is bounded : any element x of S satisfies $x \leq \frac{a}{b}$. Therefore, S being a bounded set of positive integers, S is finite and hence contains a maximal element. Let q be this maximal element and let $r := a - qb$.

We need to prove that $0 \leq r < b$. By definition $r \geq 0$ (it belongs to S). To prove that $r < b$, let us argue by contradiction. Suppose that $r \geq b$. As $r = a - qb$, we get

$$a - (q + 1)b \geq 0$$

This means that $q + 1 \in S$ but $q + 1 > q$. This contradicts the *maximality* of q . Therefore $r < b$ and the existence is proved.

Let us now prove the *uniqueness*.

Again we argue by contradiction. Suppose that there exists a pair (q', r') satisfying $a = q'b + r'$ with $0 \leq r' < b$ and such that $q' \neq q$. By subtracting the inequality $0 \leq r < b$ to this inequality, we get $-b < r' - r < b$ i.e.

$$|r - r'| < b$$

Now by subtracting $a = q'b + r'$ to $a = qb + r$ and taking the modulus, we get

$$|r - r'| = |q - q'|b$$

By assumption $q \neq q'$, hence $|q' - q| \geq 1$ and we get the inequality

$$|r - r'| \geq b$$

The two inequalities satisfied by $r - r'$ contradict each other, hence $q = q'$. Now $|r - r'| = |q - q'|b = 0$, hence $r = r'$. The uniqueness is proved. \square

Theorem 1.3 Let $a \geq b > 0$ be two integers and (q, r) such that

$$a = bq + r, \quad 0 \leq r < b$$

Then

$$\gcd(a, b) = \gcd(b, r).$$

Proof. Let $A := \gcd(a, b)$ and $B := \gcd(b, r)$. As $r = a - bq$ and A divides a and b , A divides r . Therefore A is a common factor of b and r . As B is the highest common divisor of b and r , $A \leq B$.

In exactly the same way, one proves (left to the reader), that $B \leq A$ and therefore $A = B$. \square

This leads to the following algorithm (the so-called **Euclid's algorithm**).

Let $a \geq b > 0$ be two integers. We wish to calculate $\gcd(a, b)$.

The method is this: Set $r_1 = a$ and $r_2 = b$. We have

$$r_1 = r_2q_1 + r_3$$

with, by the above proposition, $\gcd(a, b) = \gcd(r_1, r_2) = \gcd(r_2, r_3)$ with $0 \leq r_3 < r_2$.

- If $r_3 = 0$, then $\gcd(r_2, r_3) = r_2$ and we are done.
- If $r_3 \neq 0$, then divide r_2 by r_3 :

$$r_2 = r_3q_2 + r_4$$

with $0 \leq r_4 < r_3$. Again, if $r_4 = 0$, then $\gcd(a, b) = r_3$, otherwise carry on...

This way one constructs a sequence:

$$r_i = r_{i+1}q_i + r_{i+2}$$

where $0 \leq r_{i+2} < r_{i+1}$.

Notice that r_{i+2} goes **strictly down** hence one **must** at some point find $r_{i+2} = 0$ and then $\gcd(a, b) = r_{i+1}$.

Remark 1.4 *When performing Euclid's algorithm, be very careful not to divide q_i by r_i . This is a mistake very easy to make.*

Example 1.5 *Take $a = 27$ and $b = 7$. We have*

$$\begin{aligned} 27 &= 3 \times 7 + 6r_1 = 27, & r_2 &= 7, r_3 = 6 \\ 7 &= 1 \times 6 + 1, & r_3 &= 6, r_4 = 1 \\ 6 &= 6 \times 1 + 0 & r_5 &= 0 \end{aligned}$$

Therefore

$$\gcd(27, 7) = \gcd(7, 6) = \gcd(6, 1) = \gcd(1, 0) = 1.$$

Another example:

$$\begin{aligned} 555 &= 155 \cdot 3 + 90 \\ 155 &= 90 \cdot 1 + 65 \\ 90 &= 65 \cdot 1 + 25 \\ 65 &= 25 \cdot 2 + 15 \\ 15 &= 10 \cdot 1 + 5 \\ 10 &= 5 \cdot 2 + 0 \\ \gcd(555, 155) &= 5 \end{aligned}$$

Euclid's algorithm is an **algorithm** meaning that no matter what the initial data is, it will yield a gcd in a finite number of steps.

It is easy to implement on a computer. Suppose that you have some standard computer language (Basic, Pascal, Fortran,...) and that it has an instruction $r := a \bmod b$ which returns the remainder of the Euclidean division of a by b .

The implementation of the algorithm would be something like this:

```

Procedure gcd(a, b)
  If a < b then
    Swap(a, b)
  While b ≠ 0
  Begin
    r := a mod b
    a := b
    b := r
  End
  Return a
End

```

The following lemma is very important for what will follow. It is essentially ‘the Euclid’s algorithm’ run backwards.

Theorem 1.6 (Bézout’s Identity) *As usual, let $a \geq b > 0$ be integers. Let $d = \gcd(a, b)$. Then there are integers $h, k \in \mathbb{Z}$ such that*

$$d = ha + kb.$$

Note that in this lemma, the integers h and k are not positive, in fact exactly one of them is negative or zero. Prove it !

Proof. Consider the sequence given by Euclid’s algorithm:

$$r_i = r_{i+1}q_i + r_{i+2}$$

where $0 \leq r_{i+2} < r_{i+1}$ with $r_1 = a, r_2 = b$.

We will show that each r_i can be expressed as $h_i a + k_i b$ with $h_i, k_i \in \mathbb{Z}$. In particular, as by Euclid’s algorithm, $\gcd(a, b)$ is some r_i , the result will follow.

This is certainly true for $i = 1, 2$ since $r_1 = 1 \times a + 0 \times b$ and $r_2 = 0 \times a + 1 \times b$.

For the inductive step, assume it is the case for r_{i-1} and r_{i-2} , i.e.

$$r_{i-1} = ha + bk, \quad r_{i-2} = h'a + k'b.$$

We have

$$r_{i-2} = q_{i-2}r_{i-1} + r_i.$$

Therefore

$$r_i = h'a + k'b - q_{i-2}(ha + kb) = (h' - q_{i-2}h)a + (k' - q_{i-2}k)b.$$

□

Example 1.7 Again we take $a = 27$ and $b = 7$.

$$27 = 3 \times 7 + 6$$

$$7 = 1 \times 6 + 1$$

$$6 = 6 \times 1 + 0.$$

Therefore

$$\begin{aligned} 1 &= 7 - 1 \times 6 \\ &= 7 - 1 \times (27 - 3 \times 7) \\ &= 4 \times 7 - 1 \times 27. \end{aligned}$$

So we take $h = -1$ and $k = 4$.

Another example

Example 1.8 Take $a = 819$ and $b = 165$.

$$819 = 165 \times 4 + 159$$

$$165 = 159 \times 1 + 6$$

$$159 = 26 \times 6 + 3$$

$$6 = 3 \times 2$$

Therefore

$$\begin{aligned} 3 &= (1 \times 159) + (-26 \times 6) \\ &= (-26 \times 165) + (27 \times 159) \\ &= (27 \times 819) + (-134 \times 165) \\ &= (-134 \times 165) + (27 \times 819) \end{aligned}$$

So we take $h = 27$ and $k = -134$.

Proposition 1.9 a and b are coprime if and only if there exist integers k and h such that $ha + kb = 1$.

Proof. If a and b are coprime, then it's just the Bézout's identity. Suppose that there exist integers k and h such that $ha + kb = 1$. Let $d = \gcd(a, b)$. Then d divides $ha + kb$, hence d divides 1, hence $d = 1$. \square

For example, for any positive integer k , $6 \cdot (7k + 6) + (-7) \cdot (6k + 5) = 1$, hence $\gcd(7k + 6, 6k + 5) = 1$.

Another example. What values can $\gcd(3a + 5, 5a + 7)$ take?

We have $5(3a + 5) - 3(5a + 7) = 4$ hence \gcd can be 1, 2 or 4.

One has the following properties of the \gcd .

- if $ha + kb = m$ for some integers h, k , then $\gcd(a, b)$ divides m .
- $\gcd(ca, cb) = c \gcd(a, b)$
- If d divides a and b , then $\gcd(a/d, b/d) = \gcd(a, b)/d$. In particular, if $d = \gcd(a, b)$, then $\gcd(a/d, a/b) = 1$.

The first property is obvious. For the second, write the Bézout's identity $ha + kb = d$ with $d = \gcd(a, b)$. Hence $cd = hac + kbc$ and hence any common divisor of ac and bc divides cd and hence smaller or equal to cd . In addition, cd divides ac and bc hence $\gcd(ca, cb) = cd$. The next property follows from this.

We also have the following property.

Proposition 1.10 *Let a, b be two positive integers. Let d be any integer dividing both a and b . Then d divides $\gcd(a, b)$.*

Proof. Write Bézout's identity:

$$\gcd(a, b) = ah + bk$$

As d divides both a and b , d divides the right hand side, hence it divides $\gcd(a, b)$. \square

Theorem 1.11 *Suppose that $a|bc$ and a and b are coprime. Then $a|c$.*

In particular, if p is a prime and $p|ab$ then $p|a$ or $p|b$.

Proof. a and b are coprime, hence there exist h and k such that $ha + kb = 1$. Multiply by c and get $c = hac + kbc$. a divides ac and bc , hence a divides the right hand side. It follows that a divides c .

The second statement follows trivially as if p does not divide a , then a and p are coprime. \square

We have the following useful consequence:

Corollary 1.12 *Let p be a prime number and $n \geq 1$ an integer. Suppose that $p|a^n$, then $p|a$.*

Proof. The proof is by induction.

If $n = 1$, then $p|a$.

Fix $n > 1$ and suppose the statement holds for all $i = 1, \dots, n - 1$. We have $p|a^n$ hence, by the above theorem, either $p|a$ or $p|a^{n-1}$. In the second case, the induction assumption implies that $p|a$. \square

Another property is the following:

Proposition 1.13 *Suppose a and b are coprime and suppose $a|c$ and $b|c$, then $ab|c$.*

Proof. As a and b are coprime, there exists integers h and k such that

$$1 = ah + bk$$

Hence

$$c = ach + cbk$$

As a divides c , we have $c = aa'$. Similarly $c = bb'$. Hence we get

$$c = ab(b'h) + ab(a'k) = ab(b'h + a'k)$$

It follows that $ab|c$. \square

Obviously, this assertion is wrong when a and b are not coprime: $4|4$ and $2|4$ but 8 does not divide 4 .

Finally, we have seen that if p is a prime, then $p|bc$ implies p divides b or p divides c . The converse to this is also true.

Proposition 1.14 *Let a be an integer such that for any integers b and c such that $a|bc$, a divides either b or c . Then a is prime.*

Proof. By contradiction. Suppose that a is not prime, write $a = a_1a_2$ with $a_1 < a$ and $a_2 < a$. Then a divides neither a_1 nor a_2 which is a contradiction. Hence a is prime. \square

We now apply the Euclid's algorithm and Bézout's identity to the solution of *linear diophantine equations*.

Let a, b, c be three positive integers. A linear diophantine equation (in two variables) is the equation

$$ax + by = c$$

A solution is a pair (x, y) of integers (not necessarily positive) that satisfy this relation.

Such an equation may or may not have solutions. For example, consider $2x + 4y = 5$. Quite clearly, if there was a solution, then 2 will divide the right hand side, which is 5. This is not the case, therefore, this equation does not have a solution.

On the other hand, the equation $2x + 4y = 6$ has many solutions : $(1, 1), (5, -1), \dots$. This suggests that the existence of solutions depends on whether or not c is divisible by the $\gcd(a, b)$ and that if such is the case, there are many solutions to the equation. This indeed is the case, as shown in the following theorem.

Theorem 1.15 (Solution to linear diophantine equations) *Let a, b, c be three positive integers, let $d := \gcd(a, b)$ and consider the equation*

$$ax + by = c$$

1. *This equation has a solution if and only if d divides c*
2. *Suppose that $d|c$ and let (x_0, y_0) be a solution. The set of all solutions is $(x_0 + n\frac{b}{d}, y_0 - n\frac{a}{d})$ where n runs through the set of all integers (positive and negative).*

Proof. For the 'if' part : Suppose there is a solution (x, y) . Then d divides $ax + by$. But, as $ax + by = c$, d divides c .

For the 'only if' part : Suppose that d divides c and write $c = dm$ for some integer m . By Bézout's lemma there exist integers h, k such that

$$d = ha + kb$$

Multiply this relation by m and get

$$c = dm = (mh)a + (mk)b$$

This shows that $(x_0 = mh, y_0 = mk)$ is a solution to the equation. That finishes the ‘only if’ part.

Let us now suppose that the equation has a solution (in particular d divides c) (x_0, y_0) . Let (x, y) be any other solution. Subtract $ax + by = c$ from $ax_0 + by_0 = c$ to get

$$a(x_0 - x) + b(y_0 - y) = 0$$

Divide by d to get

$$\frac{a}{d}(x_0 - x) = -\frac{b}{d}(y_0 - y)$$

This relation shows that $\frac{a}{d}$ divides $\frac{b}{d}(y_0 - y)$ but the integers $\frac{a}{d}$ and $\frac{b}{d}$ are coprime hence $\frac{a}{d}$ divides $y_0 - y$ (by 1.11)

Therefore, there exists an integer n such that

$$y = y_0 - n\frac{a}{d}$$

Now plug this into the equality $\frac{a}{d}(x_0 - x) = -\frac{b}{d}(y_0 - y)$ to get that

$$x = x_0 + n\frac{b}{d}$$

□

The proof of this theorem gives a *procedure* for finding solutions, it is as follows:

1. Calculate $d = \gcd(a, b)$. If d does not divide c , then there are no solutions and you’re done. If d divides c , $c = md$ then there are solutions.
2. Run Euclid’s algorithm backwards to find h, k such that $d = ha + kb$. Then $(x_0 = mh, y_0 = mk)$ is a solution.
3. All solutions are

$$\left(x_0 + n\frac{b}{d}, y_0 - n\frac{a}{d}\right)$$

where n runs through all integers.

Example 1.16 Take $a = 27, b = 7, c = 5$. We have found that $\gcd(a, b) = 1$ (in particular there will be solutions with any c) and that $1 = 4 \times 7 - 1 \times 27$ hence $h = -1$ and $k = 4$.

Our procedure gives a particular solution : $(-5, 20)$ and the general one $(-5 + 7n, 20 - 27n)$.

Take $a = 666, b = 153, c = 43$. We have found that $\gcd(a, b) = 9$, it does not divide 43, hence no solutions.

Take $c = 45 = 5 \times 9$. There will be solutions. We had $9 = 3 \times 666 - 13 \times 153$. A particular solution is $(15, -65)$ and the general one is $(15 + 17n, -65 - 74n)$.

(in particular there will be solutions with any c) and that $1 = 4 \times 7 - 1 \times 27$ hence $h = -1$ and $k = 4$.

Our procedure gives a particular solution : $(-5, 20)$ and the general one is $(-5 + 7n, 20 - 27n)$.

2 Factorisation into primes.

Lemma 2.1 If $p|a_1 \cdots a_n$ then there exists $1 \leq i \leq n$ such that $p|a_i$.

Proof. One proceeds by induction. True for $i = 1, 2$ so suppose true for $n - 1$ and suppose that $p|a_1 \cdots a_n$. Let $A = a_1 \cdots a_{n-1}$ and $B = a_n$ then $p|AB$ implies $p|A$ or $p|B$. In the latter case we are done and in the former case the inductive hypothesis implies that $p|a_i$ for some $1 \leq i \leq n - 1$. \square

Theorem 2.2 (Unique Factorisation Theorem) If $a \geq 2$ is an integer then there are primes $p_i > 0$ such that

$$a = p_1 p_2 \cdots p_s.$$

Moreover this factorisation is unique in the sense that if

$$a = q_1 q_2 \cdots q_t$$

for primes $q_j > 0$ then

$$s = t$$

and

$$\{p_1, \dots, p_s\} = \{q_1, \dots, q_s\}$$

(equality of sets) In other words, the p_i s and the q_i s are the same prime numbers up to reordering.

Proof. For existence suppose the result does not hold. Then there an integer which can not be written as a product of primes. Among all those integers, there is a smallest one (the integers under consideration are greater than two !). Let a be this smallest integer which is not a product of primes. Certainly a is not prime so $a = bc$ with $1 < b, c < a$. As b and c are strictly smaller than a , they are products of primes. Write

$$b = p_1 \cdots p_k$$

and

$$c = p_{k+1} \cdots p_l$$

hence

$$a = p_1 \cdots p_l,$$

This contradicts the definition of a hence the factorisation exists.

For uniqueness suppose that we have an example where there are two distinct factorisations. Again we can choose a *smallest* integer with two different factorisations

$$a = p_1 \cdots p_s = q_1 \cdots q_t.$$

Then $p_1 | q_1 \cdots q_t$ so by lemma 2.1 we have $p_1 | q_j$ for some $1 \leq j \leq t$ then since p_1 and q_j are primes we have $p_1 = q_j$. But then dividing a by p_1 we have a smaller integer with two distinct factorisations, a contradiction. \square

Remark 2.3 *Of course, the primes in the factorisation $a = p_1 \cdots p_s$ need not be distinct. For example : $4 = 2^2$, here $p_1 = p_2 = 2$. Similarly $8 = 2^3$, $p_1 = p_2 = p_3 = 2$. Also $12 = 3 \times 2^2$, $p_1 = 3, p_2 = p_3 = 2$*

In fact we have that for any integer $a \geq 2$, there exist s distinct primes p_1, \dots, p_s and t integers $e_i \geq 1$ such that

$$a = p_1^{e_1} \cdots p_s^{e_t}$$

Examples of factorisations:

$$1000 = 2^3 \times 5^3$$

$$144 = 2^4 \times 3^2$$

$$975 = 2^3 \times 5^3$$

Factoring a given integer is hard as there is no procedure like Euclidean algorithm. One usually does it by trial and error. The following trivial lemma helps.

Lemma 2.4 (Square root test) *Let n be a composite (not prime) integer. Then n has a prime divisor $\leq \sqrt{n}$.*

Proof. Write $n = ab$ with $1 < a, b < n$. Suppose that $a \geq \sqrt{n}$, then $n = ab \geq \sqrt{n}b$ hence $b \leq \sqrt{n}$ and therefore any prime divisor of b is $\leq \sqrt{n}$.
□

For example, suppose you were to factor 3372. Clearly it's divisible by 2 : $3372 = 2 \times 1686$. Now, 1686 is again divisible by two : $1686 = 2 \times 843$ and $3372 = 2^2 \times 843$. Now we notice that 3 divides $843 = 3 \times 281$. Now the primes $< \sqrt{281}$ are 2, 3, 5, 7, 11, 13 and 281 is not divisible by any of these. Hence 281 is prime and we get a factorisation:

$$3372 = 2^2 \cdot 3 \cdot 281$$

How many primes there are ? Here is the answer.

Theorem 2.5 (Euclid's Theorem) *There exist infinitely many primes.*

Proof. Suppose not, let p_1, \dots, p_n be all the primes there are. Consider $Q = p_1 p_2 \cdots p_n + 1$. Since Q has a prime factorisation, there is a prime p that divides Q . This prime p has to belong to our list, after reordering we can assume that $p = p_1$. Then p_1 divides $Q - p_1 \cdots p_n = 1$ which is not possible because p_1 is prime. □

The idea we used here is this : suppose the set of all primes is finite, we *construct* an integer that is not divisible by any of the primes from this set. This is a contradiction.

Can we use the same idea to prove that there are infinitely many primes of a certain form? In some cases yes.

Quite clearly Euclid's theorem shows that there are infinitely many *odd* primes since the only even prime is 2. Put in another way, it shows that there are infinitely many primes of the form $2k + 1$.

Let's look at primes of the form $4k + 3$. Are there infinitely many of them ?

Suppose there are finitely many and list them p_1, \dots, p_r . Note that $p_1 = 3$. Consider $Q = 4p_2 \cdots p_r + 3$ (note that we started at p_2 !!!).

The integer Q is clearly not divisible by 3 (otherwise 3 would divide $p_2 \cdots p_r$ and $p_i \neq 3$ for all $i > 1$).

None of the p_i , $i > 2$ divides Q . Indeed suppose some p_i , $i > 2$ divides Q . Then

$$4p_2 \cdots p_r + 3 = p_i k$$

which shows that p_i divides 3 which is not the case.

To get a contradiction, we need to prove that Q is divisible by a prime of the form $4k + 3$, for it will have necessarily be one of the p_i s and they do not divide Q .

This is precisely what we are proving.

Lemma 2.6 *Every integer of the form $4k + 3$ has a prime factor of the form $4k + 3$.*

Proof. Let $N = 4k + 3$.

The smallest positive integer of this form is 3 which is prime, hence the property holds for 3.

Suppose that the property holds for all integers $< N$ of the form $4k + 3$. If N is prime, then take for the factor N itself.

We can and do assume that N is composite. Write $N = N_1 N_2$ with $1 < N_i < N$. As N is odd, N_1 and N_2 are odd. Any odd number is of the form $4k + 1$ or $4k + 3$.

Suppose $N_1 = 4a + 1$ and $N_2 = 4b + 1$. Then $N = N_1 N_2 = (4a + 1)(4b + 1) = 4(4ab + a + b) + 1$ is of the form $4k + 1$ which contradicts the fact that N is of the form $4k + 3$. Hence one of the N_i s, N_1 say has a prime factor of the form $4k + 3$. As $N_1 < N$, by induction assumption, N_1 and hence N has a prime factor of the form $4k + 3$. This finishes the proof. \square

Note that the proof does not work if you try to prove that there are infinitely many primes of the form $4k + 1$. This is where it fails. The first prime of this form is $5 = 4 \times 1 + 1$ but when you try to construct your Q , you get $Q = 4 \times 5 + 1 = 21 = 3 \times 7$. The divisors of Q are of the form $4k + 3$, not $4k + 1$...

In other words, the method fails because the divisors of Q can have no divisor of the form $4k + 1$.

It is however true that there are infinitely many primes of the form $4k + 1$, in fact, there is the following spectacular theorem :

Theorem 2.7 (Dirichlet's theorem on primes in arithmetic progressions)

Let a and d be two coprime integers. There exist infinitely many primes of the form $a + kd$.

The proof of this theorem is well beyond the scope of this course.

2.1 Congruences.

We define $a \equiv b \pmod m$ iff $m|(a-b)$ in other words iff there exists an integer $k \in \mathbb{Z}$ such that $a = b + km$.

We say a is *congruent* to b modulo m .

The *congruency class* of a is the set of numbers congruent to a modulo m . This is written $[a]$. In other words

$$[a] = \{a + km : k \in \mathbb{Z}\}$$

Every integer is congruent to one of the numbers $0, 1, \dots, m-1$ (can be seen using Euclidean division), so the set of all congruency classes is

$$\mathbb{Z}/m\mathbb{Z} = \{[0], \dots, [m-1]\}$$

Ex. Take $m = 3$, then $[8] = [5] = [2] = [-1] = [-4] = \dots$

For an integer k , $4k + 1 \equiv 1 \pmod 4$, $4k + 3 \equiv 3 \pmod 4$ and $4k \equiv 0 \pmod 4$.

An integer is *even* if and only if it is zero mod 2. An integer is *odd* if and only if it is one mod 2.

Let $a \geq b$ be two positive integers and let (q, r) be such that $a = bq + r$. Then $a \equiv r \pmod b$. It may help to think of congruences as the remainders of the Euclidean division.

Another trivial but useful observation is that if $a \equiv b \pmod m$ and d divides a, b and m , then $\frac{a}{d} \equiv \frac{b}{d} \pmod{\frac{m}{d}}$.

Proposition 2.8 *If $a \equiv b \pmod m$ then $b \equiv a \pmod m$.*

If $a \equiv a' \pmod m$ and $b \equiv b' \pmod m$ then $a + b \equiv a' + b' \pmod m$ and $ab \equiv a'b' \pmod m$.

Proof. trivial □

We can rewrite this proposition by simply saying:

$$[a] + [b] = [a + b] \text{ and } [a][b] = [ab]$$

The proposition says that these operations $+$ and \times are well defined operations on $\mathbb{Z}/m\mathbb{Z}$.

Ex. Write down addition and multiplication tables in $\mathbb{Z}/3\mathbb{Z}$, $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}$.

By an *inverse* of a modulo m we mean a number c such that $ac \equiv 1 \pmod{m}$. This is written $c \equiv a^{-1} \pmod{m}$.

An element may or may not have an inverse mod m .

Take $m = 6$. $[5]$ has an inverse in $\mathbb{Z}/6\mathbb{Z}$:

$$[5] \times [5] = [25] = [1]$$

While $[3]$ does not have an inverse : in $\mathbb{Z}/6\mathbb{Z}$ we have $[3][2] = [6] = [0]$. So if $[3]$ had an inverse, say $[a]$, we would have $[3][a] = [1]$, and by multiplying by $[2]$ we would get $[0] = [2]$ which is not the case.

This suggests that the existence of the inverse of $a \pmod{m}$ has something to do with common factors of a and m . This is indeed the case as shown in the following lemma.

Lemma 2.9 *An integer a has an inverse modulo m if and only if a and m are coprime ($\gcd(a, m) = 1$).*

Proof. The integer a has an inverse mod m if and only if the equation

$$ax + my = 1$$

has a solution. This equation has a solution if and only if $\gcd(a, m)$ divides 1 which is only possible if $\gcd(a, m) = 1$. \square

As usual, the proof of the lemma gives a procedure for finding inverses. Use Euclidean algorithm to calculate $\gcd(a, m)$. If it's not one, there is no inverse. If it is one run the algorithm backwards to find h and k such that $ah + mk = 1$ and

$$[a]^{-1} = [h]$$

Notice in particular that if p is a prime number, then any class $[x] \neq [0]$ is invertible. The set $\mathbb{Z}/p\mathbb{Z}$ is a **field** and it is denoted \mathbb{F}_p .

Example 2.10 Find $43^{-1} \pmod{7}$.

Euclid's algorithm :

$$43 = 6 \times 7 + 1$$

They are coprime and $1 = -6 \times 7 + 1 \times 43$. Hence $43^{-1} = 1 \pmod{7}$
Same with $32^{-1} \pmod{7}$.

$$32 = 4 * 7 + 4$$

$$7 = 1 * 4 + 3$$

$$4 = 1 * 3 + 1$$

And

$$1 = (1*4)+(-1*3) = (-1*7)+(2*4) = (2*32)+(-9*7) = (-9*7)+(2*32)$$

Hence $32^{-1} = 2 \pmod{7}$.

Same with $49^{-1} \pmod{15}$.

$$49 = 3 * 15 + 4$$

$$15 = 3 * 4 + 3$$

$$4 = 1 * 3 + 1$$

And get

$$1 = (1*4)+(-1*3) = (-1*15)+(4*4) = (4*49)+(-13*15) = (-13*15)+(4*49)$$

Hence $49^{-1} \pmod{15} = 4$.

The equation $ax \equiv b \pmod{m}$.

More generally, suppose we want to solve an equation

$$ax = b \pmod{m}$$

By this we mean, find all integers $x \pmod{m}$ that satisfy the equation.

The equation is equivalent to the existence of an integer y such that

$$ax + my = b$$

And we know how to solve this !

This equation has a solution if and only if $d = \gcd(a, b)$ divides b and we know how to find all the solutions.

In particular, the equation has solutions if and only if d divides b .

If this is the case, then to solve the equation, divide it by d , let $a' = a/d, c' = c/d$. Write $a'x + m'y = b'$. Bézout's identity gives (h, k) such that $a'h + m'k = 1$.

By the theorem on solutions of linear diophantine equations, all values of x are $\{b'h + nm'\}$ and the solutions of the equation are the $\{[b'h + nm']\}$. Notice that there are exactly d of them.

Let's see a few examples.

$$2x \equiv 4 \pmod{10}.$$

We have $\gcd(2, 10) = 2$, it divides 4, there are solutions. Dividing by 2 we get $x \equiv 2 \pmod{5}$ i.e $x = 2 + 5n$.

Now the solutions are $\{[2], [7]\}$ (classes $\pmod{10}$).

The equation $2x = 4 \pmod{5}$ has no solutions.

Another example:

$$3x \equiv 6 \pmod{18}$$

$\gcd(3, 18) = 3$ divides 6. We find $x \equiv 2 \pmod{6}$. Solutions are $\{[2], [8], [14]\}$.

Another: $10x \equiv 14 \pmod{18}$.

We have $\gcd(10, 18) = 2$, divides 14, we'll find 2 solutions.

Euclid's algorithm gives:

$$18 = 10 + 8$$

$$10 = 8 + 2$$

$$8 = 4 \times 2 + 0$$

and Bézout's identity:

$$2 = 10 + (-1) \times 8 = (-1) \times 18 + 2 \times 10$$

hence

$$14 = -7 \times 18 + 14 \times 10$$

The general solution is $x = 14 + 9n$.

The solutions to the congruence are $\{[14], [5]\}$.

Notice that when a and m are coprime, then there is a unique solution and it is given by $[a]^{-1}[b]$.

For example, solve $99x \equiv 100 \pmod{101}$.

99 and 101 are coprime, hence there is a unique solution.

Euclid's algorithm gives:

$$101 = 99 + 2$$

$$99 = 2 \times 49 + 1$$

$$2 = 1 \times 2 + 0$$

Bézout's identity:

$$1 = (1 * 99) + (-49 * 2) = (-49 * 101) + (50 * 99)$$

One finds $[99]^{-1} = [50]$. The unique solution is $[50] \times [100] = [5000] = [51]$.

Corollary 2.11 $\mathbb{F}_p^\times = \{[1], [2], \dots, [p-1]\}$ is a group with the operation of multiplication.

Proof. A group is a set with a binary operation (in this case multiplication), such that (i) the operation is associative; (ii) there is an identity element; (iii) every element has an inverse. Clearly $[1]$ is the identity element, and the every element has an inverse because $1, 2, 3, \dots, p-1$ are coprime with p . \square

Recall that Lagrange's theorem states that if G is a finite group and H is a subgroup, then $|H|$ divides $|G|$. The corollary of this theorem is that if $a \in G$ and $k \geq 0$ is the smallest integer such that $a^k = 1$, then k divides $|G|$.

Theorem 2.12 (Fermat's Little Theorem) If p is prime and $a \in \mathbb{Z}$ then

$$a^p \equiv a \pmod{p}.$$

Hence if $p \nmid a$ then $a^{p-1} \equiv 1 \pmod{p}$.

Proof. If $p|a$ then $a \equiv 0 \pmod{p}$ and $a^p \equiv 0 \pmod{p}$ so suppose p does not divide a , and so $a \in \mathbb{F}_p^\times$. Recall that by a corollary to Lagrange's Theorem, the order of an element of a group divides the order of the group. Let n be the order of a , so $a^n \equiv 1$. But by the corollary to Lagrange's theorem, $n|p-1$. \square

Let's look at an example. What is $33^{22} \pmod{23}$? 23 is prime so $33^{22} \equiv 1 \pmod{23}$.

How about $3^{101} \pmod{103}$? Well 103 is prime so $3^{102} \equiv 1 \pmod{103}$ So $3^{101} \equiv 3^{-1} \pmod{103}$. To find $3^{-1} \pmod{103}$ use Euclid's algorithm.

$$103 = 3 \times 34 + 1.$$

So $3^{-1} \equiv 34 \pmod{103}$. Hence $3^{101} \equiv 34 \pmod{103}$.

Another example : $32^6 \pmod{7}$. We know that $32^7 \pmod{7} = 32 \pmod{7}$. It follows that $32^7 = 32^{-1} \pmod{7}$. It suffices to calculate $32^{-1} \pmod{7}$. We get

$$\begin{aligned} 32 &= 4 * 7 + 4 \\ 7 &= 1 * 4 + 3 \\ 4 &= 1 * 3 + 1 \end{aligned}$$

and

$$1 = (1*4) + (-1*3) = (-1*7) + (2*4) = (2*32) + (-9*7) = (-9*7) + (2*32)$$

Hence $32^{-1} \equiv 2 \pmod{7}$ and $32^6 \equiv 2 \pmod{7}$.

Yet another example : $45^{35} \pmod{13}$.

We have $13 \times 2 = 26$ and $45^{13} \equiv 45 \pmod{13}$. Hence, as $35 = 13 \times 2 + 9$, we have $45^{35} = 45^2 \times 45^9 = 45^{11} \pmod{13}$. As $45^{12} \cong 1 \pmod{13}$, we have $45^{11} = 45^{-1} \pmod{13}$

We need to calculate $45^{-1} \pmod{13}$.

Euclidian algorithm : We get

$$\begin{aligned} 45 &= 3 * 13 + 6 \\ 13 &= 2 * 6 + 1 \end{aligned}$$

and

$$1 = (1 * 13) + (-2 * 6) = (-2 * 45) + (7 * 13) = (7 * 13) + (-2 * 45)$$

Hence $45^{35} \equiv -2 \pmod{13} \equiv 11 \pmod{13}$.

Let's do $43^{42} \pmod{13}$. We have $43^{39} \equiv 43^3 \pmod{13}$. Hence $43^{42} \equiv 43^6 \pmod{13}$. Now $43 \pmod{13} \equiv 4 \pmod{13}$. Hence $43^{42} \equiv 4^6 \pmod{13}$. Now $4^2 = 16 = 3 \pmod{13}$. Hence $4^6 = 4^{2^3} = 3^3 = 27 \pmod{13} = 1 \pmod{13}$. Hence $43^{42} \equiv 1 \pmod{13}$.

And now we get to yet another application of the Bézout's lemma.

We would like to find integers z that satisfy **two** congruences: $z \equiv x \pmod{m}$ and $z \equiv y \pmod{n}$.

This is not always possible as the example $z \equiv 3 \pmod{4}$ and $z \equiv 5 \pmod{8}$ shows. If such a z existed, one would get $0 = 1 \pmod{8}$ which is not the case. The reason is that 4 and 8 are not coprime. However, when n and m are coprime, we have the following theorem.

Theorem 2.13 (Chinese Remainder Theorem) Suppose m and n are coprime; let x and y be two integers. Then there is a unique $[z] \in \mathbb{Z}/nm$ such that $z \equiv x \pmod{m}$ and $z \equiv y \pmod{n}$.

Proof. (existence) By Bezout's Lemma, we can find $h, k \in \mathbb{Z}$ such that

$$hn + km = 1.$$

Notice that $hn \equiv 1 \pmod{m}$ and $km \equiv 1 \pmod{n}$.

Given x, y we choose z by

$$z = hnx + kmy.$$

Clearly $z \equiv hnx \equiv x \pmod{m}$ and $z \equiv y \pmod{n}$.

(uniqueness) For uniqueness, suppose z' is another solution. Then $z \equiv z' \pmod{n}$ and $z \equiv z' \pmod{m}$. Hence there exist integers r, s such that

$$z - z' = nr = ms.$$

Since $hn + km = 1$ we have

$$z - z' = (z - z')hn + (z - z')km = mshn + nrkm = nm(sh + rk).$$

Hence $z \equiv z'(nm)$. □

As usual the proof gives you a procedure to find z . To find z , find h and k as in the Bézouts lemma (run Euclidean algorithm backwards). Then z is $hnx + kmy$.

Find the unique solution of $x \equiv 3 \pmod{7}$ and $x \equiv 9 \pmod{11}$ satisfying $0 \leq x \leq 76$.

Solution find h, k such that $7h + 11k = 1$ using Euclid:

$$11 = 7 + 4$$

$$7 = 4 + 3$$

$$4 = 3 + 1$$

$$\text{So } 1 = 4 - 3 = 4 - (7 - 4) = 2 \cdot 4 - 7 = 2 \cdot (11 - 7) - 7 = 2 \cdot 11 - 3 \cdot 7.$$

$$\text{Hence let } h = -3 \text{ and } k = 2 \text{ so take } x = -3 \cdot 7 \cdot 9 + 2 \cdot 11 \cdot 3 = -189 + 66 = -123 \equiv 31 \pmod{77}.$$

Further linear algebra. Chapter II. Polynomials.

Andrei Yafaev

1 Definitions.

In this chapter we consider a field k . Recall that examples of fields include \mathbb{Q} , \mathbb{R} , \mathbb{C} , \mathbb{F}_p where p is prime.

A polynomial is an expression of the form

$$f(x) = a_0 + a_1x + \cdots + a_dx^d = \sum a_nx^n, \quad a_0, \dots, a_d \in k$$

The elements a_i s are called **coefficients** of f . If all a_i s are zero, then f is called a **zero** polynomial (notation: $f = 0$).

If $f \neq 0$, then the **degree** of f (notation $\deg(f)$) is by definition the largest integer $n \geq 0$ such that $a_n \neq 0$.

If $f = 0$, then, by convention, $\deg(f) = -\infty$.

Addition and multiplication are defined as one expects: if $f(x) = \sum a_nx^n$ and $g(x) = \sum b_nx^n$ then we define

$$(f + g)(x) = \sum (a_n + b_n)x^n,$$

$$(fg)(x) = \sum c_nx^n,$$

where

$$c_n = \sum_{i=0}^n a_ib_{n-i}.$$

Notice that we always have:

$$\deg(f \times g) = \deg(f) + \deg(g).$$

(we are using the convention that $-\infty + n = -\infty$). Notice also that

$$\deg(f + g) \leq \max\{\deg(f), \deg(g)\}$$

If $f = \sum a_n X^n \neq 0$ has degree d , the coefficient a_d is called the **leading coefficient** of f . If f has leading coefficient 1 then f is called **monic**.

Two polynomials are **equal** if all their coefficients are equal.

Example 1.1 $f(x) = x^3 + x + 2$ has degree 3, and is monic.

The set of all polynomials with coefficients in k is denoted by $k[x]$.

The polynomials of the form $f(x) = a_0$ are called **constant** and a constant polynomial of the form $f(x) = a_0 \neq 0$ is called a **unit** in $k[x]$. Notice that units are exactly polynomials of degree zero. In other words, units are precisely non-zero constant polynomials. Another way to put it: units are precisely polynomials of degree zero. Units are analogous to $\pm 1 \in \mathbb{Z}$. Notice that a unit is monic if it is just 1.

Given $f, g \in k[x]$, we say that g **divides** f if there exists a polynomial $h \in k[x]$ such that

$$f = gh$$

Clearly, a unit divides any polynomial. Also for any polynomial f , f divides f .

A non-zero polynomial is called **irreducible** if it is not a unit and whenever $f = gh$ with $g, h \in k[x]$, either g or h must be a unit. In other words, the only polynomials that divide f are units and f itself. Irreducible polynomials are analogues of prime numbers from Chapter I.

If f divides g i.e. $f = gh$, then

$$\deg(f) = \deg(g) + \deg(h) \leq \deg(g)$$

We prove the following:

Proposition 1.2 *Let $f \in k[x]$. If $\deg(f) = 1$ then f is irreducible.*

Proof. Suppose $f = gh$. Then $\deg(g) + \deg(h) = 1$. Therefore the degrees of g and h are 0 and 1, so one of them is a unit. \square

The property of being irreducible depends on the field k !

For example, the polynomial $f(x) = x^2 + 1$ is irreducible no matter what k is. If $k = \mathbb{R}$, then $f(x) = x^2 + 1$ is irreducible. However, if $k = \mathbb{C}$, then $x^2 + 1 = (x + i)(x - i)$ is reducible.

Similarly $x^2 - 2$ factorises in $\mathbb{R}[X]$ as $(x + \sqrt{2})(x - \sqrt{2})$, but is irreducible in $\mathbb{Q}[X]$ (since $\sqrt{2}$ is irrational).

We have the following theorem:

Theorem 1.3 (Fundamental Theorem of Algebra) *Let $f \in \mathbb{C}[x]$ be a non-zero polynomial. Then f factorises as a product of linear factors (i.e. polynomials of degree one):*

$$f(X) = c(x - \lambda_1) \cdots (x - \lambda_d)$$

where c is the leading coefficient of f .

The proof of this uses complex analysis and is omitted here.

The theorem means that in $\mathbb{C}[x]$ the irreducible polynomials are exactly the polynomials of degree 1, with no exceptions. In $\mathbb{R}[x]$ the description of the irreducible polynomials is a little more complicated (we'll do it later). In $\mathbb{Q}[x]$ things are much more complicated and it can take some time to determine whether a polynomial is irreducible or not.

2 Euclid's algorithm in $k[x]$.

The rings \mathbb{Z} and $k[x]$ are very similar. This is because in both rings we are able to divide with remainder in such a way that the remainder is smaller than the element we divided by. In \mathbb{Z} if we divide a by b we find:

$$a = qb + r, \quad 0 \leq r < b.$$

In $k[x]$, we have something identical:

Theorem 2.1 Euclidean division *Given $f, g \in k[X]$ with $g \neq 0$ and $\deg(f) \geq \deg(g)$ there exist unique $q, r \in k[x]$ such that*

$$f = qg + r \quad \text{and} \quad \deg(r) < \deg(b).$$

Proof. The proof is **IDENTICAL** to the one for integers.

Existence:

Choose q so that $\deg(f - qg)$ is minimal. Write

$$(f - qg)(x) = c_k x^k + \cdots + c_0,$$

$c_k \neq 0$.

If g has degree $m \leq k$ say

$$g(x) = b_m x^m + \cdots + b_0,$$

where $b_m \neq 0$. Let us subtract $c_k b_m^{-1} x^{k-m} g$ from $(f - qg)$ to get

$$q' = q + c_k b_m^{-1} x^{k-m}.$$

Then

$$f - q'g = f - qg - c_k b_m^{-1} x^{k-m} g = c_k x^k - c_k x^k + \text{terms of order at most } k - 1.$$

This contradicts the minimality of $\deg(f - qg)$. Hence we can choose q such that $\deg(f - qg) < \deg(g)$ and then set $r = f - qg$.

Uniqueness:

Suppose we have $f = q_1 g + r_1 = q_2 g + r_2$. Then

$$g(q_1 - q_2) = r_2 - r_1.$$

So if $q_1 \neq q_2$ then $\deg(q_1 - q_2) \geq 0$ so $\deg(g(q_1 - q_2)) \geq \deg(g)$. But then $\deg(r_2 - r_1) \leq \max\{\deg(r_2), \deg(r_1)\} < \deg(g) \leq \deg(g(q_1 - q_2)) = \deg(r_2 - r_1)$, a contradiction. So $q_1 = q_2$ and $r_1 = r_2$. □

The procedure for finding q and r is the following. Write:

$$f = a_0 + a_1 x + \cdots + a_m x^m$$

where $a_m \neq 0$ and

$$g = b_0 + b_1 x + \cdots + b_n x^n$$

with $b_n \neq 0$ and $m \geq n$.

We calculate

$$r_1 = f - \frac{a_m}{b_n} x^{m-n} g$$

if $\deg(r_1) < \deg(g)$ then we are done; if not, we continue until we found $\deg(r_i) < \deg(g)$.

For example: in $\mathbb{Q}[x]$:

$$f(x) = x^3 + x^2 - 3x - 3, \quad g(x) = x^2 + 3x + 2$$

Then

$$f - xg = -2x^2 - 5x - 3$$

$$(f - xg) + 2g = x + 1$$

Hence

$$f = (x - 2)g + x + 1$$

hence $q = x - 2, r = x + 1$.

Another example: still in $\mathbb{Q}[x]$

$$f(x) = 3x^4 + 2x^3 + x^2 - 4x + 1, \quad g(x) = x^2 + x + 1$$

Then

$$f - 3x^2g = -x^3 - 2x^2 - 4x + 1$$

$$(f - 3x^2g) + xg = -x^2 - 3x + 1$$

$$(f - 3x^2g) + xg + g = -2x + 2$$

Hence

$$f = (3x^2 - x - 1)g + (-2x + 2)$$

hence $q = 3x^2 - x - 1, r = -2x + 2$.

We now define the **greatest common divisor** of two polynomials:

Definition 2.1 Let f and g be two polynomials in $k[x]$ with one of them non-zero. The **greatest common divisor** of f and g is the unique **monic** polynomial $d = \gcd(f, g)$ with the following properties:

1. d divides f and g
2. c divides f and g implies c divides d

Why is it unique? Suppose we had two gcd's d_1 and d_2 , then d_1 divides d_2 i.e. $d_1 = hd_2$. Similarly d_2 divides d_1 : $d_2 = kd_1$. It follows that

$$\deg(h) + \deg(k) = 0$$

therefore $h, k \in k \setminus \{0\}$. As polynomials d_1 and d_2 are monic, we have $h = k = 1$ hence $d_1 = d_2$.

The greatest common divisor of f and g is also the unique monic polynomial d such that:

1. d divides f and g
2. if c divides f and g , then $\deg(c) \leq \deg(d)$

Let us see that this definition is equivalent to the previous one. Let $d_1 = \gcd(f, g)$ and d_2 the monic polynomial satisfying

1. d_2 divides f and g
2. if c divides f and g , then $\deg(c) \leq \deg(d_2)$

We need to show that $d_1 = d_2$.

As $d_1|f$ and $d_1|g$, we have

$$\deg(d_1) \leq \deg(d_2)$$

by definition of d_2 .

Now, $d_2|f$ and $d_2|g$ hence $d_2|d_1$ by definition of d_1 . In particular $\deg(d_2) \leq \deg(d_1)$.

It follows that $\deg(d_2) = \deg(d_1)$ and $d_2|d_1$.

Hence $d_1 = \alpha d_2$ with $\deg(\alpha) = 0$ i.e. α is a unit. As both d_1 and d_2 are monic, it follows that

$$d_1 = d_2$$

From Euclidean division, just like in the case of integers, we derive a Euclidean algorithm for calculating the gcd.

The Euclidean division gives $f = qg + r$, $\deg(r) < \deg(g)$; then

$$\gcd(f, g) = \gcd(g, r)$$

To see this, just like in the case of integers, let $A := \gcd(f, g)$ and $B := \gcd(g, r)$. We have $f = qg + r$. As A divides f and g , A divides r . Therefore A divides g and r . As B is the greatest common divisor of g and r , $A|B$.

Similarly, B divides g and r , hence $B|f$. It follows that $B|A$.

The same argument we used to show that the gcd is unique now shows that $A = B$.

Running the algorithm backwards, we get the **Bézout's identity**: there exist two polynomials h and k such that

$$\gcd(f, g) = hf + kg$$

Just like in the case of integers, it follows that

1. f and g are coprime iff there exist polynomials h and k such that

$$hf + gk = 1$$

2. If $f|gh$ and f and g are coprime, then $f|h$

We say that f and g are coprime if $\gcd(f, g) = 1$ and, using Bézout's identity, one sees that f and g are coprime if and only if there exist (h, k) , polynomials, such that

$$1 = hf + kg$$

Let's do an example : Calculate $\gcd(f, g)$ and find h, k such that $\gcd(f, g) = hf + kg$ with $f = x^4 + 1$ and $g = x^2 + x$.

We write: $f - x^2g = -x^3 + 1$, then $f - x^2g + xg = x^2 + 1$ and $f - x^2g + xg - g = 1 - x$ and we are finished.

We find:

$$f = (x^2 - x + 1)g + 1 - x$$

And then

$$x^2 + x = (-x + 1)(-x - 2) + 2$$

As 2 is invertible, we find that the gcd is one !

Now, we do it backwards:

$$\begin{aligned} 2 &= g - (1 - x)(-x - 2) = \\ &g + (1 - x)(x + 2) = \\ &g + (x + 2)(f - (x^2 - x + 1)g) = \\ &g[1 - (x + 2)(x^2 - x + 1)] + (x + 2)f = \\ &g[-1 - x^3 - x^2 + x] + (x + 2)f \end{aligned}$$

hence $h = (1/2)(x + 2)$ and $k = (1/2)(-x^3 - x^2 + x - 1)$.
Now, suppose we considered the same example in $\mathbb{F}_2[x]$. In $\mathbb{F}_2[x]$,

$$f = x^4 + 1 = x^4 - 1 = (x - 1)^4$$

and

$$g = x(x + 1) = x(x - 1)$$

Clearly in $\mathbb{F}_2[x]$, $\gcd(f, g) = x - 1$ and the Bézout's identity is

$$x - 1 = (x^2 - x + 1)g - f$$

An element $a \in k$ is called a **root** of a polynomial $f \in k[x]$ if $f(a) = 0$.
We have the following consequence of the Euclidean division:

Theorem 2.2 (The Remainder Theorem) *If $f \in k[x]$ and $a \in k$ then*

$$f(a) = 0 \iff (x - a)|f.$$

Proof. If $(x - a)|f$ then there exists $g \in k[x]$ such that $f(x) = (x - a)g(x)$.
Then $f(a) = (a - a)g(a) = 0g(a) = 0$.

Conversely by Euclidean division we have $q, r \in k[x]$ with $\deg(r) < \deg(x - a) = 1$ such that $f(x) = q(x)(x - a) + r(x)$. So $r(x) \in k$. Then

$$r(a) = f(a) - q(x)(a - a) = 0 + 0 = 0.$$

Hence $(x - a)|f$. □

A consequence of this theorem is the following:

Lemma 2.3 *A polynomial $f \in k[x]$ of degree 2 is reducible if and only if f has a root in k .*

Proof. If f has a root a in k , then the above theorem shows that $(x - a)$ divides f and as $\deg(f) > 1$, f is reducible. Conversely, suppose that f is reducible i.e.

$$f = gh$$

where neither g nor h is a unit.

Therefore, we have $\deg(g) = \deg(h) = 1$. Dividing by the leading coefficient of g , we may assume that $g = x - a$ for some a in k , hence $f(a) = 0$, a is a root of f . □

For example, $x^2 + 1$ in $\mathbb{R}[x]$ is of degree two and has no roots in \mathbb{R} , hence it is irreducible in $\mathbb{R}[x]$.

The polynomial $x^2 + 1$ is also irreducible in $\mathbb{F}_3[x]$: it suffices to check that 0, 1 and 2 are not roots in \mathbb{F}_3 .

We have the following corollary of the fundamental theorem of algebra and euclidean division.

Proposition 2.4 *No polynomial $f(x)$ in $\mathbb{R}[x]$ of degree > 2 is irreducible in $\mathbb{R}[x]$.*

Proof. Let $f \in \mathbb{R}[x]$ be a polynomial of degree > 2 . By fundamental theorem f has a root in \mathbb{C} , call it α . Then $\bar{\alpha}$ (complex conjugate) is another root (because $f \in \mathbb{R}[x]$). Let

$$p(x) = (x - \alpha)(x - \bar{\alpha}) = x^2 - (\alpha + \bar{\alpha})x + \alpha\bar{\alpha}$$

The polynomial p is in $\mathbb{R}[x]$ and is irreducible (if it was reducible it would have a real root).

Divide f by p .

$$f(x) = p(x)q(x) + r(x)$$

with $\deg(r) \leq 1$. We can write $r = sx + r$ with $s, r \in \mathbb{R}$. But $f(\alpha) = p(\alpha)q(\alpha) + r(\alpha) = 0 = r(\alpha)$. As α not real we must have $r = s = 0$. This implies that p divides f but $\deg(p) = 2 < \deg(f)$. It follows that f is not irreducible. \square

Notice that the proof above shows that any polynomial of degree three in $\mathbb{R}[x]$ has a root in \mathbb{R} . This is not true for polynomials of degree > 3 . For example $x^4 + 1$ is not irreducible in $\mathbb{R}[x]$:

$$x^4 + 1 = (x^2 - \sqrt{2}x + 1)(x^2 + \sqrt{2}x + 1)$$

However, the polynomial $x^4 + 1$ has no roots in \mathbb{R} . The proposition above does not hold for $\mathbb{Q}[x]$. For example, it can be shown that $x^4 + 1$ is irreducible in $\mathbb{Q}[x]$. The reason why the proof does not work is that although $\alpha + \bar{\alpha}$ and $\alpha\bar{\alpha}$ are in \mathbb{R} , they have no reason to be in \mathbb{Q} .

Lemma 2.5 *Suppose f in $k[x]$ is irreducible. Then $f|g_1 \cdots g_r$ implies $f = g_i$ for some i .*

Proof. Copy the proof for integers. □

Theorem 2.6 (Unique Factorisation Theorem) *Let $f \in k[x]$ be monic. Then there exist $p_1, p_2, \dots, p_n \in k[x]$ monic irreducibles such that*

$$f = p_1 p_2 \cdots p_n.$$

If q_1, \dots, q_s are monic and irreducible and $f = q_1 \cdots q_s$ then $r = s$ and (after reordering) $p_1 = q_1, \dots, p_r = q_r$.

Proof. (Existence): We prove the existence by induction on $\deg(f)$. If f is linear then it is irreducible and the result holds. So suppose the result holds for polynomials of smaller degree. Either f is irreducible and so the result holds or $f = gh$ for g, h non-constant polynomials of smaller degree. By our inductive hypothesis g and h can be factorized into irreducibles and hence so can f .

(Uniqueness): Factorization is obviously unique for linear polynomials (or even irreducible polynomials). For the inductive step, assume all polynomials of smaller degree than f have unique factorization. Let

$$f = g_1 \cdots g_s = h_1 \cdots h_t,$$

with g_i, h_j monic irreducible.

Now g_1 is irreducible and $g_1 | h_1 \cdots h_t$. By the Lemma, there is $1 \leq j \leq t$ such $g_1 | h_j$. This implies $g_1 = h_j$ since they are both monic irreducibles. After reordering, we can assume $j = 1$, so

$$g_2 \cdots g_s = h_2 \cdots h_t,$$

is a polynomial of smaller degree than f . By the inductive hypothesis, this has unique factorization. I.e. we can reorder things so that $s = t$ and

$$g_2 = h_2, \dots, g_s = h_t.$$

□

The fundamental theorem of algebra tells you exactly that any monic polynomial in $\mathbb{C}[x]$ is a product of irreducibles (recall that polynomials of degree one are irreducible).

A consequence of factorisation theorem and fundamental theorem of algebra is the following: any polynomial of **odd degree** has a root in \mathbb{R} .

Indeed, in the decomposition we can have polynomials of degree one and two. Because the degree is odd, we have a factor of degree one, hence a root.

Another example : $x^2 + 2x + 1 = (x + 1)^2$ in $k[x]$.

Look at $x^2 + 1$. This is irreducible in $\mathbb{R}[x]$ but in $\mathbb{C}[x]$ it is reducible and decomposes as $(x + i)(x - i)$ and in $\mathbb{F}_2[x]$ it is also reducible : $x^2 + 1 = (x + 1)(x - 1) = (x + 1)^2$ in $\mathbb{F}_2[x]$. In $\mathbb{F}_5[x]$ we have $2^2 = 4 = -1$ hence $x^2 + 1 = (x + 2)(x - 2)$ (check : $(x - 2)(x + 2) = x^2 - 4 = x^2 + 5$).

In fact one can show that $x^2 + 1$ is reducible in $\mathbb{F}_p[x]$ is and only if $p \equiv 1 \pmod{4}$.

In $\mathbb{F}_p[x]$, the polynomial $x^p - x$ decomposes as product of polynomials of degree one.

Suppose you want to decompose $x^4 + 1$ in $\mathbb{R}[x]$. It is not irreducible puisque degree est > 2 . Also, $x^4 + 1$ does not have a root in $\mathbb{R}[x]$ but it does in $\mathbb{C}[x]$. The idea is to decompose into factors of the form $(x - a)$ in $\mathbb{C}[x]$ and then group the conjugate factors.

This is in general how you decompose a polynomial into irreducibles in $\mathbb{R}[x]$!

So here, the roots are

$$a_1 = e^{i\pi/4}, a_2 = e^{3i\pi/4}, a_3 = e^{5i\pi/4}, a_4 = e^{7i\pi/4}.$$

Now note that $a_4 = \overline{a_1}$ and the polynomial $(x - a_1)(x - a_4)$ is irreducible over \mathbb{R} . The middle coefficient is $-(a_1 + a_4) = -2 \cos(\pi/4) = -\sqrt{2}$. Hence we find : $(x - a_1)(x - a_4) = x^2 - \sqrt{2}x + 1$.

Similarly $a_2 = \overline{a_3}$ and $(x - a_2)(x - a_3) = x^2 + \sqrt{2}x + 1$.

We get the decomposition into irreducibles over \mathbb{R} :

$$x^4 + 1 = (x^2 - \sqrt{2}x + 1)(x^2 + \sqrt{2}x + 1)$$

In $\mathbb{Q}[x]$ one can show that $x^4 + 1$ is irreducible.

In $\mathbb{F}_2[x]$ we can also decompose $x^4 + 1$ into irreducibles. Indeed :

$$x^4 + 1 = x^4 - 1 = (x^2 - 1)^2 = (x - 1)^4$$

Further linear algebra. Chapter III. Revision of linear algebra.

Andrei Yafaev

As in the previous chapter, we consider a field k . Typically, $k = \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{F}_p$.

A vector space over a field k (one also says a k -vector space) is a set V with two operations: addition and scalar multiplication by elements of k . Elements of V are called vectors, and elements of k are called scalars. The axioms are:

- $(V, +)$ is an abelian (commutative) group (in particular there is an identity for addition called zero 0).
- $(xy)v = x(yv)$ for $x, y \in k, v \in V$.
- $(x + y)(v + w) = xv + xw + yv + yw$ for $x, y \in k, v, w \in V$.
- $1v = v$.

A typical example of a vector space is the space k^n of n -tuples of elements in k . In particular k itself is a vector space over itself.

Another example is $k[X]$. The set of polynomials with coefficients in k is a vector space. Fix $n \geq 0$ and let $k[X]_n$ be the set of polynomials of degree less or equal to n . This is a vector space. If $n = 0$, then this vector space is just k .

The set of polynomials of degree exactly n is not a vector space. For example because 0 is not there.

Take $k = \mathbb{R}$ and let C be the set of all continuous functions from $[0, 1]$ to \mathbb{R} . Then C is an \mathbb{R} -vector space.

Similarly, take $k = \mathbb{C}$ and let H be the set of all holomorphic functions from the unit ball to \mathbb{C} . This is a \mathbb{C} vector space. Of course it also an \mathbb{R} -vector space.

Another example. Let $a, b \in \mathbb{R}$ and consider the set of all twice differentiable functions f such that

$$\frac{d^2 f}{dx^2} + a \frac{df}{dx} + bf = 0$$

This is an \mathbb{R} vector space (exercise).

Let k be a field, the set of matrices $M_n(k)$ with coefficients in k is a k -vector space.

- A linear combination of $\{v_1, \dots, v_n\}$ is a vector of the form $x_1 v_1 + \dots + x_n v_n$.

For example, consider the vector space $k[x]_n$ as before. This vector space is in fact the set of all linear combinations of the elements $1, x, \dots, x^n$.

- The span of a set of vectors is the set of linear combinations of those vectors.

As above, $k[x]_n$ is the span of the set $\{1, x, \dots, x^n\}$. We say that the vectors $\{1, x, \dots, x^n\}$ span or generate this vector space.

Consider k^2 and the vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Then the set of vectors $\{e_1, e_2\}$ spans k^2 for any field k .

- Let V be a k -vector space. A subset A of V is said to generate V if V is the span of A .

In the examples above $\{1, x, \dots, x^n\}$ generates $k[x]_n$ but $1, \{1, x\}, \{1, x, x^2\}, \{1, x, x^2, \dots, x^{n-1}\}$ do not generate V .

The set $\{e_1, e_2\}$ certainly generates \mathbb{R}^2 while $\{e_1\}$ or $\{e_2\}$ do not.

- Let V be a k -vector space. A subset $W \subset V$ is called a subspace if any linear combination of elements in W is in W . In other words, a subspace is a subset which is a vector space with the same addition and scalar multiplication as V .

Let V be a k -vector space and take $v \in V$. The set kv of all multiples of v by elements of k is a subspace. More generally, take any set $A \subset V$, then the set of linear combinations of elements of A is a vector subspace.

As an (easy) exercise, prove that given any collection $W_i, i \in I$ (I is some set, finite or infinite) of subspaces of V , the intersection $\bigcap_{i \in I} W_i$ is a subspace. The union is not! For example, let $V = k^2$ and $W_1 = ke_1$ and $e_2 = ke_2$. It is quite clear that $W_1 \cup W_2$ not a subspace, for example $e_1 + e_2$ is not in it.

Let C be the set of continuous functions $[0, 1] \rightarrow \mathbb{R}$. We have seen that this is an \mathbb{R} -vector space. Let $W = \{f \in C : f(0) = 0\}$. This is a subspace (easy exercise).

We have seen that $k[x]$ is a vector space. The space $k[x]_n$ is a vector subspace.

- Vectors $v_1, \dots, v_n \in V$ are called linearly independent if whenever $\sum_{i=1}^n \lambda_i v_i = 0$ (for some $\lambda_i \in \mathbb{R}$), then $\lambda_i = 0$ for all i .

For example, in k^2 , vectors e_1, e_2 are linearly independent. Clearly e_1 and $2e_1$ are not linearly independent. If v is any vector, v and 0 are obviously never linearly independent.

If $k = \mathbb{R}$ then the vectors e_1 and $2e_2$ are linearly independent. Not if $k = \mathbb{F}_2$, indeed in this case $2e_2 = 0$.

In $k[x]$, the vectors $\{1, x, x^2, \dots\}$, are linearly independent.

- A set $\{v_1, \dots, v_n\}$ of vectors is a basis for V if it is linearly independent and its span is V (it generates V). Such a set is called a *basis*. If this is the case then every vector has a unique expression as a linear combination of $\{v_1, \dots, v_n\}$.

For example $\{e_1, e_2\}$ is a basis of k^2 . The set $\{1, x, x^2, \dots, x^n\}$ is basis of $k[x]_n$.

The set $\{1, x, x^2, \dots\}$, is a basis of $k[x]$.

A basis of the space $M_n(k)$ consists of the matrices $E_{i,j}$ which have 1 at the position (i, j) and zero elsewhere.

- Any vector space has a basis. If it has a basis consisting of finitely many elements, it is called finite dimensional.

The dimension of a vector space is the number of vectors in a basis. This does not depend on the basis: any two bases have the same number of elements.

For example k^n has dimension n , $k[x]_n$ has dimension $n + 1$ while $k[x]$ is infinite dimensional and so is C .

\mathbb{R} viewed as vector space over itself has dimension 1 but viewed as vector space over \mathbb{Q} is infinite dimensional.

\mathbb{C} viewed as a vector space over itself has dimension 1 but as a vector space over \mathbb{R} it has dimension 2 : a basis is $\{1, i\}$.

The space $M_n(k)$ has dimension n^2 .

- The fundamental properties are the following.

Suppose that V is a finite dimensional vector space.

1. Let $\{v_1, \dots, v_n\}$ be a *linearly independent* set of vectors. Then one can find v_{n+1}, \dots, v_r such that $\{v_1, \dots, v_r\}$ is a basis of V .

In particular, if $W \subset V$ is a subspace, then any basis of W can be completed to a basis of V .

2. Let $\{w_1, \dots, w_r\}$ be a family of vectors spanning V , then one can extract a basis from $\{w_1, \dots, w_r\}$

- Direct sums.

Let V be a vector space and U, W two subspaces. The sum $U + W = \{u + w : u \in U, w \in W\}$ and the intersection $U \cap W$ are subspaces of V . The sum $U + W$ is called *direct* if $U \cap W = \{0\}$, the notation is $U \oplus W$. One has $\dim(U \oplus W) = \dim(U) + \dim(W)$.

For example, take k^n with its standard basis, then k^n is the direct sum

$$k^n = (ke_1) \oplus \dots \oplus (ke_n)$$

Consider in $V = k^2$ the vectors $v_1 = e_1 + e_2$ and $v_2 = e_1 - e_2$. If $k = \mathbb{R}$, then $k^2 = \text{Span}(v_1) \oplus \text{Span}(v_2)$.

Indeed, let $v = \alpha e_1 + \beta e_2 \in V$, then

$$v = \frac{\alpha}{2}(v_1 + v_2) + \frac{\beta}{2}(v_1 - v_2)$$

hence $\text{Span}(v_1) + \text{Span}(v_2) = V$.

Now suppose $v \in \text{Span}(v_1) \cap \text{Span}(v_2)$. Then $v = \lambda v_1 = \mu v_2$. As e_1 and e_2 are linearly independent, we get $\lambda = -\mu = \mu$ hence, because $k = \mathbb{R}$, $\mu = 0$, therefore $v = 0v_2 = 0$ hence $\text{Span}(v_1) \cap \text{Span}(v_2) = \{0\}$ and the sum is direct.

If $k = \mathbb{F}_2$, then it fails as in this case $v_1 = v_2$.

Let V, W be vector spaces. A function $T : V \rightarrow W$ is a linear map if

- $T(v + w) = T(v) + T(w)$,
- $T(xv) = xT(v)$.

or equivalently, $T(v + xv) = T(v) + xT(v)$. A bijective linear map is called an isomorphism of vector spaces.

For example the map $T : \mathbb{C} \rightarrow \mathbb{C}$ that sends z to \bar{z} is not a linear map of \mathbb{C} -vector spaces : $T(\lambda z) = \overline{\lambda T(z)}$! But it is a map of real vector spaces : if $\lambda \in \mathbb{R}$, then $\bar{\lambda} = \lambda$.

Similarly, the map $k \rightarrow k$ sending $x \mapsto x^2$ is not linear when $k = \mathbb{R}$ or $k = \mathbb{C}$, but it is linear if $k = \mathbb{F}_2$.

The map $\text{tr} : M_n(k) \rightarrow k$ sending a matrix M to $\sum_{i,j} M_{i,j}$ is linear. If T is linear, we define its kernel and image:

$$\ker(T) = \{v \in V : T(v) = 0\},$$

$$\text{im}(T) = \{T(v) : v \in V\}.$$

The rank of T is the dimension of the image of T , and the nullity of T is the dimension of the kernel of T .

This implies the following:

Theorem 0.1 (Rank-Nullity Theorem) *Let $T : V \rightarrow W$ be a linear map. Then*

$$\text{rank}(T) + \text{nullity}(T) = \dim V.$$

Proof. Let $\{v_1, \dots, v_r\}$ be a basis of $\ker(T)$ and $\{w_1, \dots, w_s\}$ be a basis of $\text{im}(T)$. By definition of the image, there exist $\{u_1, \dots, u_s\}$ vectors of V such that

$$T(u_i) = w_i$$

We claim that $\{u_1, \dots, u_s, v_1, \dots, v_r\}$ form a basis of V which will conclude the proof.

First we show linear independence. Suppose that

$$a_1v_1 + \dots + a_sv_r + b_1u_1 + \dots + b_ru_s = 0$$

Apply T , we get

$$0 = T(0) = b_1T(u_1) + \dots + b_rT(u_s) = b_1w_1 + \dots + b_sw_s$$

(note that $a_1v_1 + \dots + a_sv_r = 0$ because the v_i s are in the kernel of T). Now, as $\{w_1, \dots, w_s\}$ is a basis of $\text{im}(T)$ (in particular it is linearly independent), we get that $b_i = 0$ for all i .

So we have $a_1v_1 + \dots + a_sv_r = 0$ and, because v_i s for a basis of $\ker(T)$ (and in particular are linearly independent), we get that $a_i = 0$ for all i .

We have shown that a_i s and b_i s are all zero hence $\{u_1, \dots, u_s, v_1, \dots, v_r\}$ is linearly independent.

It remains to show that $\{u_1, \dots, u_s, v_1, \dots, v_r\}$ spans V .

Let $x \in V$. By the choice of $\{w_1, \dots, w_s\}$ as a basis of the image of T , we have

$$T(x) = \sum_{i=1}^s a_iw_i = \sum_{i=1}^s a_iT(u_i) = T\left(\sum_{i=1}^s a_iu_i\right)$$

Therefore

$$T\left(x - \sum_{i=1}^s a_iu_i\right) = 0$$

and hence

$$x - \sum_{i=1}^s a_iu_i \in \ker(T)$$

and now, by the choice of $\{v_i\}$ as basis of $\ker(T)$, there exist b_i s such that

$$x = \sum_{i=1}^s a_iu_i + \sum_{j=1}^r b_jv_j$$

which shows that $\{u_1, \dots, u_s, v_1, \dots, v_r\}$ generates V .

This finishes the proof. \square

Here are some consequences of this theorem.

Definition 0.1 A linear map $T: V \rightarrow W$ is called **isomorphism** if there exists

1. $T_1: W \rightarrow V$ such that $TT_1 = I_V$ (identity of V)
2. $T_2: W \rightarrow V$ such that $T_2T = I_W$ (identity of W)

In particular, a linear map $T^{-1}: W \rightarrow V$ is an isomorphism if there exists T^{-1} such that $T^{-1}T$ is the identity.

It is easy (and left as exercise) to see that $T: V \rightarrow W$ is an isomorphism if and only if T is both surjective and injective. (for the converse you will need to construct T_1 and T_2 as maps and then show that they are linear.)

If $T: V \rightarrow V$ is an isomorphism, one also says that T is invertible.

Corollary 0.2 Let $T: V \rightarrow W$ be a linear map **with** $\dim V = \dim W$. If T is injective, then T is an isomorphism. If T is surjective, then T is an isomorphism.

Proof. If T is injective, then $\ker(T) = \{0\}$. By the above theorem, $\dim(\text{im}(T)) = \dim(V) = \dim(W)$ and hence $\text{im}(T) = W$ and T is surjective. Injective + Surjective = Isomorphism.

Similarly, if T is surjective, then $\dim(\text{im}(T)) = \dim(W) = \dim(V)$ and hence $\dim(\ker(T)) = 0$. It follows that T is injective. Injective + Surjective = Isomorphism. \square

Corollary 0.3 Let V and W be two vector spaces of same dimension. Then V is isomorphic to W (i.e there is an isomorphism between V and W).

Proof. Let $r = \dim(V) = \dim(W)$ and let $\{v_1, \dots, v_r\}$ be a basis of V and $\{w_1, \dots, w_r\}$ be a basis of W . Define T by $T(v_i) = w_i$. By construction, T is surjective and by the theorem it's also injective hence an isomorphism. \square

If $T: V \rightarrow W$ and $T': W \rightarrow U$, then we denote by TT' the composition $T \circ T': V \rightarrow U$.

If $T: V \rightarrow V$ and n is an integer, we write T^n for T composed with itself n times.

1 Matrix representation of linear maps.

Let V and W be two finite dimensional vector space over a field k . Suppose that V is of dimension r and W is of dimension t .

Let $B = \{b_1, \dots, b_r\}$ be a basis for V and $B' = \{b'_1, \dots, b'_s\}$ be a basis for W .

For any vector $v \in V$ we shall write $[v]_B$ (in the future, we will by abuse of notation simply call this column vector v when it is obvious which basis we are referring to) for the column vector of coefficients of v with respect to the basis B , i.e.

$$[v]_B = \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}, \quad v = x_1 b_1 + \dots + x_r b_r.$$

Given a linear map $T : V \rightarrow W$ we have

$$T(v) = T\left(\sum_{i=1}^r x_i b_i\right) = \sum_{i=1}^r x_i T(b_i)$$

Now we have

$$T(b_i) = \sum_{j=1}^s a_{ji} b'_j$$

We get :

$$T(v) = \sum_{1 \leq i \leq r, 1 \leq j \leq s} x_i a_{ji} b'_j$$

In other words it is the $s \times r$ matrix , usually denoted by $[T]_{B, B'}$, with entries

$$([T]_{B, B'})_{i, j} = a_{ji}$$

In practice, to write a matrix of T with respect to given bases, decompose $T(b_i)$ in the basis of W and write column vectors, this gives the matrix $[T]$

The matrix $[T]_{B, B'}$ is called the matrix of T with respect to bases B and B' .

A LINEAR TRANSFORMATION IS THE MATRIX WITH RESPECT TO SPECIFIED BASES OF THE SOURCE AND THE TARGET SPACES.

Example. $V = W = \mathbb{R}^3$ with canonical bases $\{e_1, e_2, e_3\}$.

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

(notice that this is the projection onto the plane $z = 0$).

One finds

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

One can find the kernel and the image. In this case, clearly the image is the span of e_1 and e_2 hence $\dim(\text{im}T) = 2$. By rank-nullity theorem, $\dim \ker(T) = 1$ and quite clearly it is generated by e_3 .

Let us look at $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ given by

$$T: \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x - y + 3z \\ 4x - 2y + 6z \\ -6x + 3y - 9z \end{pmatrix}$$

One finds :

$$\begin{pmatrix} 2 & -1 & 3 \\ 4 & -2 & 6 \\ -6 & -3 & -9 \end{pmatrix}$$

Quite clearly, the first column vector in this matrix is -2 times the second and the third is the first minus the second, therefore $\text{im}(T)$ is one dimensional

and spanned by $\begin{pmatrix} 2 \\ 4 \\ -6 \end{pmatrix}$. The rank-nullity theorem implies that the dimension

of $\ker(T)$ is 2. To find $\ker(T)$ one needs to solve $[T]v = 0$. By elimination,

one easily shows that the kernel has equation $2x - y + 3z = 0$, hence can be spanned by the vectors $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}$

Another example : $k[x]_n \rightarrow k[x]_{n-1}$ sending f to its derivative f' . Quite clearly it's a linear map. Find its matrix, image and kernel.

Same question with $k[x]_n \rightarrow k[x]_n$ sending f to $f + f'$.

Let us consider the transformation $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + y \\ x - y \end{pmatrix}$$

and $B_1 = B_2 = \{v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, v_2 = \begin{pmatrix} -3 \\ 2 \end{pmatrix}\}$.

One calculates $T(v_1) = v_1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix} = -6v_1 - 2v_2$ and $T(v_2) = v_1 = \begin{pmatrix} -3 \\ 2 \end{pmatrix} = 17v_1 + 6v_2$.

The matrix of T with respect to these bases is

$$\begin{pmatrix} -6 & 17 \\ -2 & 6 \end{pmatrix}$$

In the canonical bases, of course the matrix is:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Ex. Let $T: M_2(k) \rightarrow M_2(k)$ sending M to M^t . Write down the matrix of this linear map in the standard bases.

If $T: V \rightarrow V$ is a linear map and B is a basis of V , then we will simply write $[T]_B$ for the matrix $[T]_B^B$.

We also have the following :

Proposition 1.1 *Let $T_1: V \rightarrow W$ and $T_2: W \rightarrow U$ be two linear maps and suppose we are given bases B, B_1, B_2 of the vector spaces V, W, U . Then for the composed map $T_2 \circ T_1$ (usually simply denoted by $T_2 T_1$) the matrix is*

$$[T_2 T_1]_{B_2}^B = [T_2]_{B_2}^{B_1} [T_1]_{B_1}^B$$

*In particular if $T: V \rightarrow V$ (such a map is called **endomorphism**) and B is a basis for V , then*

$$[T^n]_B = [T]_B^n$$

and, if we suppose that T is an isomorphism

$$[T^{-1}]_B = [T]_B^{-1}$$

Let V be a vector space and B and B' be two bases for V . Let I_V be the identity map from V to V . We let

$$P = [I_V]_{B_2}^{B_1}$$

We have $P^{-1} = [I_V]_{B_1}^{B_2}$. Writing $T = I_V T I_V$ and applying the proposition, we obtain

$$[T]_{B_2}^{B_2} = P [T]_{B_1}^{B_1} P^{-1}$$

Further linear algebra. Chapter IV. Jordan normal form.

Andrei Yafaev

In what follows V is a vector space of dimension n and B is a basis of V . In this chapter we are concerned with linear maps $T: V \rightarrow V$.

Let A be the matrix representing T in the basis B . Because A is an $n \times n$ matrix, we can form powers A^k for any k with the convention that $A^0 = I_n$. Note that A^k represents the transformation T composed k times.

Notice for example that when the matrix is diagonal with coefficients λ_i on the diagonal, then A^n is diagonal with coefficients λ_i^n . Notice also that such a matrix is invertible if and only if all λ_i s are non-zero, then A^{-1} is the diagonal matrix with coefficients λ_i^{-1} on the diagonal.

Definition 0.1 Let $f(X) = \sum a_i X^i \in k[X]$. We define

$$f(T) = \sum a_i T^i.$$

where we define $T^0 = \text{Id}$. This is a linear transformation.

If $A \in M_n(k)$ is a matrix then we define

$$f(A) = \sum a_i A^i,$$

This matrix $f(A)$ represents $f(T)$ in the basis B .

What this means is that we can ‘evaluate’ a polynomial at a $n \times n$ matrix and get another $n \times n$ matrix. We write $[f(T)]_B$ for this matrix in the basis B , obviously it is the same as $f([T]_B)$.

Let’s look at an example.

Take $A = \begin{pmatrix} -1 & 3 \\ 4 & 7 \end{pmatrix}$ and $f(x) = x^2 - 5x + 3$. Then

$$f(A) = A^2 - 5A + 3 = \begin{pmatrix} 21 & 3 \\ 4 & 29 \end{pmatrix}$$

Another example: $V = M_n(k)$ and T sends M to M^t . Consider $f(x) = x^2 - 1$. As $T^2 = I$, we see that $f(T) = 0$.

Notice that

$$[f(T)]_B = f([T]_B)$$

It follows that if T is a linear map, $f(T) = 0$ if and only if $f(A) = 0$ where A is the matrix of T in some (equivalently any) basis.

Another property worth noting is that if $f, g \in k[x]$, then

$$f(T)g(T) = (fg)(T) = (gf)(T) = g(T)f(T)$$

Definition 0.2 Recall that the characteristic polynomial of an $n \times n$ matrix A is defined by

$$\text{ch}_A(x) = \det(x \cdot I_n - A) = (-1)^n \det(A - x \cdot I_n).$$

This is a monic polynomial of degree n over k . Now suppose $T : V \rightarrow V$ is a linear map. We can define ch_T to be $\text{ch}_{[T]_B}$ but we need to check that this does not depend on the basis B . If C is another basis with transition matrix M then we have:

$$\begin{aligned} \text{ch}_{[T]_C}(X) &= \det(X \cdot I_n - M^{-1}[T]_B M) \\ &= \det(M^{-1}(X \cdot I_n - [T]_B)M) \\ &= \det(M)^{-1} \det(X \cdot I_n - [T]_B) \det(M) \\ &= \det(X \cdot I_n - [T]_B) \\ &= \text{ch}_{[T]_B}(X) \end{aligned}$$

In other words, the characteristic polynomial does not depend on the choice of the basis in which we write our matrix.

The following (important !) theorem was proved in the first year courses.

Theorem 0.1 (Cayley–Hamilton Theorem) For any A be an $n \times n$ matrix. We have $\text{ch}_A(A) = 0$.

We therefore have:

Theorem 0.2 (Cayley–Hamilton Theorem) For any $T : V \rightarrow V$ linear map, we have $\text{ch}_T(T) = 0$.

Example 0.3 Take $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ Then $ch_A(x) = (x - \lambda_1)(x - \lambda_2)$ and clearly $ch_A(A) = 0$.

Take $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Calculate $ch_A(x)$ and check that $ch_A(A) = 0$.

There are plenty of polynomials f such that $f(A) = 0$, all the multiples of ch_A for example.

What can also happen is that some divisor g of ch_A is already such that $g(A) = 0$. Take the identity I_n for example. Its characteristic polynomial is $(x - 1)^n$ but in fact $g = x - 1$ is already such that $g(I_n) = 0$. This leads us to the notion of **minimal polynomial**.

0.1 Minimal polynomials.

Definition 0.3 Let V be a finite dimensional vector space over a field k and $T : V \rightarrow V$ a linear map. A minimal polynomial of T is the **monic** polynomial $m \in k[X]$ such that

- $m(T) = 0$;
- if $f(T) = 0$ and $f \neq 0$ then $\deg f \geq \deg m$.

Notice that if m is a polynomial as in the definition, and f a polynomial such that $f(T) = 0$ and $\deg(f) < \deg(m)$, then $f = 0$.

Indeed, if f was not zero, then dividing f by the coefficient of the leading term, one would obtain a monic polynomial with degree strictly less than m which contradicts the definition of m .

Theorem 0.4 Every linear map $T : V \rightarrow V$ has a unique minimal polynomial m_T . Furthermore if $f \in k[x]$ is such that $f(T) = 0$ iff $m_T | f$.

Proof. Firstly the Cayley-Hamilton theorem implies that there exists a polynomial f satisfying $f(T) = 0$, namely $f = ch_T$. Among monic polynomials satisfying $f(T) = 0$ we choose one of smallest degree, call it m_T . This shows that the minimal polynomial exists.

Suppose that m_T is not unique then there exists another monic polynomial $n(x) \in k[X]$ satisfying the conditions of the definition. Because $n(T) = 0$, $\deg(n) \geq \deg(m_T)$ and because $m_T(T) = 0$, $\deg(m_T) \geq \deg(n)$ hence $\deg(m) = \deg(n)$.

If $f(x) = m_T(x) - n(x)$ then

$$f(T) = m_T(T) - n(T) = 0,$$

also $\deg(f) < \deg(m_T) = \deg(n)$. By the remark following the definition of the minimal polynomial, we see that $f = 0$ i.e. $m_T = n$. This proves the uniqueness.

Suppose $f \in k[X]$ and $f(T) = 0$. By the Division Algorithm for polynomials there exist unique $q, r \in k[X]$ with $\deg(r) < \deg(m)$ and

$$f = qm + r.$$

Then

$$r(T) = f(T) - q(T)m(T) = 0 - q(T) \cdot 0 = 0.$$

So r is the zero polynomial (by the remark following the definition.) Hence $f = qm$ and so $m|f$.

Conversely if $f \in k[X]$ and $m|f$ then $f = qm$ for some $q \in k[X]$ and so $f(T) = q(T)m(T) = q(T) \cdot 0 = 0$. \square

Corollary 0.5 *If $T : V \rightarrow V$ is a linear map then $m_T | \text{ch}_T$.*

Proof. By the Cayley-Hamilton Theorem $\text{ch}_T(T) = 0$. \square

Using the corollary we can calculate the minimal polynomial as follows:

- Calculate ch_T and factorise it into irreducibles.
- Make a list of all the factors.
- Find the monic factor m of smallest degree such that $m(T) = 0$.

Example 0.6 *Suppose T is represented by the matrix $\begin{pmatrix} 2 & 1 & \\ & 2 & \\ & & 2 \end{pmatrix}$. The characteristic polynomial is*

$$\text{ch}_T(X) = (X - 2)^3.$$

The factors of this are:

$$1, \quad (X - 2), \quad (X - 2)^2, \quad (X - 2)^3.$$

The minimal polynomial is $(X - 2)^2$.

In fact this method can be speeded up: there are certain factors of the characteristic polynomial which cannot arise. To explain this we recall the definition of an *eigenvalue*

Definition 0.4 Recall that a number $\lambda \in k$ is called an eigenvalue of T if there is a **non-zero** vector v satisfying

$$T(v) = \lambda \cdot v.$$

The non-zero vector v is called an eigenvector

Remark 0.7 It is important that an eigenvector be non-zero. If you allow zero to be an eigenvector, then **any** λ would be an eigenvalue.

Proposition 0.8 Let v be an eigenvector of T with eigenvalue $\lambda \in k$. Then for any polynomial $f \in k[X]$,

$$(f(T))(v) = f(\lambda) \cdot v.$$

Proof. Just use that $T(v) = \lambda v$. □

Theorem 0.9 If $T : V \rightarrow V$ is linear and $\lambda \in k$ then the following are equivalent:

- (i) λ is an eigenvalue of T .
- (ii) $m_T(\lambda) = 0$.
- (iii) $\text{ch}_T(\lambda) = 0$.

Proof. (i) \Rightarrow (ii): Assume $T(v) = \lambda v$ with $v \neq 0$. Then by the proposition,

$$(m_T(T))(v) = m_T(\lambda) \cdot v.$$

But $m_T(T) = 0$ so we have $m_T(\lambda) \cdot v = 0$. Since $v \neq 0$ this implies $m_T(\lambda) = 0$.

(ii) \Rightarrow (iii): This is trivial since we have already shown that m_T is a factor of ch_T .

(iii) \Rightarrow (i): Suppose $\text{ch}_T(\lambda) = 0$. Therefore $\det(\lambda \cdot \text{Id} - T) = 0$. It follows that $(\lambda \cdot \text{Id} - T)$ is not invertible hence $\lambda \cdot \text{Id} - T$ has a non-zero kernel. Therefore there exists $v \in V$ such that $(\lambda \cdot \text{Id} - T)(v) = 0$. But then $T(v) = \lambda \cdot v$. □

Now suppose the characteristic polynomial of T factorises into irreducibles as

$$\text{ch}_T(X) = \prod_{i=1}^r (X - \lambda_i)^{a_i}.$$

By fundamental theorem of algebra, if $k = \mathbb{C}$, we can always factorise it like this.

Then the minimal polynomial has the form

$$m_T(X) = \prod_{i=1}^r (X - \lambda_i)^{b_i}, \quad 1 \leq b_i \leq a_i.$$

This makes it much quicker to calculate the minimal polynomial. Indeed, in practice, the number of factors and the a_i s are quite small.

Example 0.10 Suppose T is represented by the matrix $\text{diag}(2, 2, 3)$. The characteristic polynomial is

$$\text{ch}_T(X) = (X - 2)^2(X - 3).$$

The possibilities for the minimal polynomial are:

$$(X - 2)(X - 3), \quad (X - 3)^2(X - 3).$$

The minimal polynomial is $(X - 2)(X - 3)$.

0.2 Generalised Eigenspaces

Definition 0.5 Let V be a finite dimensional vector space over a field k , and let $\lambda \in k$ be an eigenvalue of a linear map $T : V \rightarrow V$. We define for $t \in \mathbb{N}$ the t -th generalised eigenspace by:

$$V_t(\lambda) = \ker((\lambda \cdot \text{Id} - T)^t).$$

Note that $V_1(\lambda)$ is the usual eigenspace (i.e. the set of eigenvectors together with zero).

Remark 0.11 We obviously have

$$V_1(\lambda) \subseteq V_2(\lambda) \subseteq \dots$$

and by definition,

$$\dim V_i(\lambda) = \text{Nullity}((\lambda \cdot \text{Id} - T)^i).$$

Another property is that generalised eigenspaces are T invariant:

$$T(V_i(\lambda)) \subset V_i(\lambda)$$

To see this, let $v \in V_i(\lambda)$. Then $T(T - \lambda \text{Id})^i v = 0 = (T - \lambda \text{Id})^i T v$, therefore $T(v) \in V_i(\lambda)$.

Example 0.12 Let

$$A = \begin{pmatrix} 2 & 2 & 2 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}.$$

We have $\text{ch}_A(X) = (X - 2)^3$ so 2 is the only eigenvalue. We'll now calculate the generalised eigenspaces $V_t(2)$:

$$V_1(2) = \ker \begin{pmatrix} 0 & 2 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

We calculate the kernel by row-reducing the matrix:

$$V_1(2) = \ker \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \text{Span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

Similarly

$$V_2(2) = \ker \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \text{Span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

$$V_3(2) = \ker \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \text{Span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Example 0.13 Let

$$A = \begin{pmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \\ 1 & 1 & -2 \end{pmatrix}.$$

Let V be a vector space and U and W be two subspaces. Then (exercise on Sheet 4), $U + W$ is a subspace of V . If furthermore $U \cap W = \{0\}$, then we call this subspace the **direct sum** of U and W and denote it by $U \oplus W$. In this case

$$\dim(U \oplus W) = \dim U + \dim W$$

When we have r subspaces U_1, \dots, U_r of V , then we say that V is the direct sum of the U_i s if every element of V can be written *uniquely* as a sum of elements of the U_i s:

For any $v \in V$, there exists a unique r -tuple (u_1, \dots, u_r) , $u_i \in U_i$ such that

$$v = u_1 + \dots + u_r$$

Notice that when this is the case the union of bases of the U_i s is a basis of V .

Theorem 0.14 (Primary Decomposition Theorem) *If V is a finite dimensional vector space over k and $T : V \rightarrow V$ is linear, with distinct eigenvalues $\lambda_1, \dots, \lambda_r \in k$.*

Let f be a monic polynomial such that $f(T) = 0$ and suppose that f factorises in $k[x]$ into a product of degree one polynomials

$$f(X) = \prod_{i=1}^r (X - \lambda_i)^{b_i},$$

then

$$V = V_{b_1}(\lambda_1) \oplus \dots \oplus V_{b_r}(\lambda_r).$$

Lemma 0.15 *Let k be a field. If $f, g \in k[x]$ satisfy $\gcd(f, g) = 1$ and T is as above then*

$$\ker(fg(T)) = \ker(f(T)) \oplus \ker(g(T)).$$

Proof. (of the theorem) By induction on r .

If $r = 1$, then $f(x) = (x - \lambda_1)^{b_1}$. We have $f(T) = 0$ hence

$$V = \ker(f(T)) = V_{b_1}(\lambda_1)$$

Now suppose that the statement holds for all finite dimensional vector spaces V , linear maps $T: V \rightarrow V$ and f such that $f(T) = 0$ and $f = \prod_{i=1}^r (X - \lambda_i)^{b_i}$ where λ_i are distinct.

Now let $T: V \rightarrow V$ and $f = \prod_{i=1}^{r+1} (X - \lambda_i)^{b_i}$ such that $f(T) = 0$. The subspace $W = \ker \prod_{i=1}^r (X - \lambda_i)^{b_i}$ is stable by T , by induction assumption, we have

$$W = V_{b_1} \oplus \cdots \oplus V_{b_r}$$

The polynomials $\prod_{i=1}^r (X - \lambda_i)^{b_i}$ and $(x - \lambda_{r+1})^{b_{r+1}}$ are coprime (because λ_i s are distinct), hence by induction assumption, we have

$$V = V_{b_1}(\lambda_1) \oplus \cdots \oplus V_{b_r}(\lambda_{r+1}).$$

This finishes the proof of the theorem. □

Proof. (of the lemma) Let $f, g \in k[x]$ satisfy $\gcd(f, g) = 1$.

Firstly if $v \in \ker f(T) + \ker g(T)$, say $v = w_1 + w_2$, with $w_1 \in \ker f(T)$ and $w_2 \in \ker g(T)$ then

$$fg(T)v = fg(T)(w_1 + w_2) = f(g(T)w_1) + f(g(T)w_2) = f(g(T)w_1)$$

Now, f and g are polynomials in $k[x]$, hence $fg = gf$, therefore

$$f(g(T)w_1) = g(f(T)w_1) = 0$$

because $w_2 \in \ker(f(T))$.

Therefore $\ker(f(T)) + \ker(g(T)) \subseteq \ker(fg(T))$.

We need to prove the equality, here we will use that $\gcd(f, g) = 1$.

Now since $\gcd(f, g) = 1$ there exist $a, b \in k[x]$ such that

$$af + bg = 1.$$

So

$$a(T)f(T) + b(T)g(T) = 1 \quad (\text{the identity map}).$$

Let $v \in \ker(fg(T))$. If

$$v_1 = a(T)f(T)v, \quad v_2 = b(T)g(T)v$$

then $v = v_1 + v_2$ and

$$g(T)v_1 = (gaf)(T)v = a(fg(T)v) = a(T)0 = 0.$$

So $v_1 \in \ker(g(T))$. Similarly $v_2 \in \ker(f(T))$ since

$$f(T)v_2 = (fbg)(T)v = b(fg(T)v) = b(T)0 = 0.$$

Hence $\ker fg(T) = \ker f(T) + \ker g(T)$. Moreover, if $v \in \ker f(T) \cap \ker g(T)$ then $v_1 = 0 = v_2$ so $v = 0$. Hence

$$\ker fg(T) = \ker f(T) \oplus \ker g(T).$$

□

Notice that, by fundamental theorem of algebra, in $\mathbb{C}[x]$ every polynomial factorises into a product of degree one ones. The theorem applies to $\text{ch}_T(x)$ and $m_T(x)$.

Definition 0.6 Recall that a linear map $T : V \rightarrow V$ is diagonalisable if there is a basis \mathcal{B} of V such that $[T]_{\mathcal{B}}$ is a diagonal matrix. This is equivalent to saying that the basis vectors in \mathcal{B} are all eigenvectors.

Theorem 0.16 Let V is a finite dimensional vector space over a field k and let $T : V \rightarrow V$ be a linear map with **distinct** eigenvalues $\lambda_1, \dots, \lambda_r \in k$. Then T is diagonalizable iff we have (in $k[X]$):

$$m_T(X) = (X - \lambda_1) \dots (X - \lambda_r).$$

Proof. First suppose that T is diagonalisable and let \mathcal{B} be a basis of eigenvectors. Let $f(X) = (X - \lambda_1) \dots (X - \lambda_r)$. We already know that $f|m_T$, so to prove that $f = m_T$ we just have to check that $f(T) = 0$. To show this, it is sufficient to check that $f(T)(v) = 0$ for each basis vector $v \in \mathcal{B}$. Suppose $v \in \mathcal{B}$, so v is an eigenvector with some eigenvalue λ_i . Then we have

$$f(T)(v) = f(\lambda) \cdot v = 0 \cdot v = 0.$$

Therefore $m_T = f$.

Conversely if $m_T = f$ then by the primary decomposition theorem we have

$$V = V_1(\lambda_1) \oplus \dots \oplus V_1(\lambda_r).$$

Let \mathcal{B}_i be a basis for $V_1(\lambda_i)$. Then obviously the elements of \mathcal{B}_i are eigenvectors and $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_r$ is a basis of V . Therefore T is diagonalisable.

□

Remark 0.17 Observe that if the characteristic polynomial splits as $\text{ch}_T(x) = (x - \lambda_1) \cdots (x - \lambda_r)$ where λ_i s are distinct eigenvalues, then $m_T = \text{ch}_T$ and the matrix is diagonalisable.

The converse, of course, does not hold.

Example 0.18 Let $k = \mathbb{C}$ and let

$$A = \begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix}.$$

The characteristic polynomial is $(x - 1)(x - 6)$. The minimal polynomial is the same. The matrix is diagonalisable.

One finds that the basis of eigenvectors is

$$\begin{pmatrix} 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

In fact this matrix is diagonalisable over \mathbb{R} or even \mathbb{Q} .

Example 0.19 Let $k = \mathbb{R}$ and let

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The characteristic polynomial is $x^2 + 1$. It is irreducible over \mathbb{R} . The minimal polynomial is the same. The matrix is not diagonalisable over \mathbb{R} , however over \mathbb{C} $x^2 + 1 = (x - i)(x + i)$ and the matrix is diagonalisable.

Example 0.20 Let $k = \mathbb{C}$ and let

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}.$$

The characteristic polynomial is X^2 . Since $A \neq 0$ the minimal polynomial is also X^2 . Since this is not a product of distinct linear factors, A is not diagonalizable over \mathbb{C} .

Example 0.21 Let $k = \mathbb{C}$ and let

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

The minimal polynomial is $(x - 1)^2$. Not diagonalisable.

0.3 Jordan Bases in the one eigenvalue case

Let $T : V \rightarrow V$ be a linear map. Fix a basis for V and let A be the matrix of T in this basis. As we have seen above, it is not always the case that T can be diagonalised; i.e. there is not always a basis of V consisting of eigenvalues of T . This the case that there is no basis of eigenvalues, the best kind of basis is a Jordan basis. We shall define a Jordan basis in several steps.

Suppose $\lambda \in k$ is the only eigenvalue of a linear map $T : V \rightarrow V$. We have defined generalized eigenspaces:

$$V_1(\lambda) \subseteq V_2(\lambda) \subseteq \dots \subseteq V_b(\lambda),$$

where b is the power of $X - \lambda$ in the minimal polynomial m_T .

We can choose a basis \mathcal{B}_1 for $V_1(\lambda)$. Then we can choose \mathcal{B}_2 so that $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for $V_2(\lambda)$ etc. Eventually we end up with a basis $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_b$ for $V_b(\lambda)$. We'll call such a basis a *pre-Jordan* basis.

Consider

$$A = \begin{pmatrix} 3 & -2 \\ 8 & -5 \end{pmatrix}.$$

One calculates the characteristic polynomial and finds $(x+1)^2$ hence $\lambda = -1$ is the only eigenvalue. The unique eigenvector is $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Hence $V_1(-1) = \text{Span}(v)$. Of course we have $(A - \lambda I_2)^2 = 0$ hence $V_2(-1) = \mathbb{C}^2$ and we complete v to a basis of $\mathbb{C}^2 = V_2(-1)$, by $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ for example. We have $Av_2 = -2v_1 - v_2$ and hence in the basis $\{v_1, v_2\}$ the matrix of A

$$\begin{pmatrix} -1 & -2 \\ 0 & -1 \end{pmatrix}$$

The basis $\{v_1, v_2\}$ is a pre-Jordan basis for A .

Example 0.22

$$A = \begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & -2 \\ 1 & 1 & -1 \end{pmatrix}$$

We have $\text{ch}_A(X) = (X - 1)^3$ and $m_A(X) = (X - 1)^2$. There is only one eigenvalue $\lambda = 1$, and we have generalized eigenspaces

$$V_1(1) = \ker \begin{pmatrix} 1 & 1 & -2 \end{pmatrix}, \quad V_2(1) = \ker(0) = \mathbb{C}^3.$$

So we can choose a pre-Jordan basis as follows:

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \right\}, \quad \mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

This in fact also works over \mathbb{R} .

Now note the following:

Lemma 0.23 *If $v \in V_t(\lambda)$ with $t > 1$ then*

$$(T - \lambda \cdot \text{Id})(v) \in V_{t-1}(\lambda).$$

Proof. Clear from the definition of the generalised eigenspaces. □

Now suppose we have a pre-Jordan basis $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_b$. We call this a *Jordan basis* if in addition we have the condition:

$$(T - \lambda \cdot \text{Id})\mathcal{B}_t \subset \mathcal{B}_{t-1}, \quad t = 2, 3, \dots, b.$$

If we have a pre-Jordan basis $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_b$, then to find a Jordan basis, we do the following:

- For each basis vector $v \in \mathcal{B}_b$, replace one of the vectors in \mathcal{B}_{b-1} by $(T - \lambda \cdot \text{Id})(v)$. When choosing which vector to replace, we just need to take care that we still have a basis at the end.
- For each basis vector $v \in \mathcal{B}_{b-1}$, replace one of the vectors in \mathcal{B}_{b-2} by $(T - \lambda \cdot \text{Id})(v)$. When choosing which vector to replace, we just need to take care that we still have a basis at the end.
- etc.
- For each basis vector $v \in \mathcal{B}_2$, replace one of the vectors in \mathcal{B}_1 by $(T - \lambda \cdot \text{Id})(v)$. When choosing which vector to replace, we just need to take care that we still have a basis at the end.

Once finished, order the vectors of the basis appropriately.

We'll prove later that this method always works.

Example 0.24 Let's look again at

$$A = \begin{pmatrix} 3 & -2 \\ 8 & -5 \end{pmatrix}.$$

We have seen that $\{v_1, v_2\}$ is a pre-Jordan basis, here v_2 is the second vector in the standard basis.

Replace v_1 by the vector $(A + I_2)v_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$. Then $\{v_1, v_2\}$ still forms a basis of \mathbb{C}^2 . This is the Jordan basis for A .

We have $Av_1 = -v_1$ and $Av_2 = v_1 - v_2$ (you do not need to calculate, just use $(A + I_2)v_2 = v_1$). Hence in the new basis the matrix is

$$\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$$

Example 0.25 In the example above, we replace one of the vectors in \mathcal{B}_1 by

$$(A - I_3) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

So we can choose a Jordan basis as follows:

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}, \quad \mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

Example 0.26 Take $k = \mathbb{R}$.

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}.$$

Here, we have seen that the characteristic and minimal polynomials are x^2 . Therefore, 0 is the only eigenvalue.

Clearly $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ generates the eigenspace and $V_2(0) = \mathbb{R}^2$. We complete the basis by taking $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. We get a pre-Jordan basis.

Let's construct a Jordan basis. Replace v_1 by $Av_2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$. This is a Jordan basis. The matrix of A in the new basis is

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Example 0.27

$$A = \begin{pmatrix} 2 & 2 & 2 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}.$$

Clearly, the characteristic polynomial is $(x-2)^3$ and it is equal to the minimal polynomial, 2 is the only eigenvalue.

$V_1(2)$ has equations $y = z = 0$, hence it's spanned by $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

$V_2(2)$ is $z = 0$ and $V_3(2)$ is \mathbb{R}^3 . Therefore, the standard basis is a pre-Jordan basis.

We have

$$(A - 2I_3)^2 = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

hence we get

Now,

$$A - 2I_3 = \begin{pmatrix} 0 & 2 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

We have

$$(A - 2I_3)v_3 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}$$

and we replace v_2 by this vector.

$$\text{Now } (A - 2I_3)v_2 = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}$$

We get:

$$v_1 = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

We have

$$Av_1 = 2v_1, \quad Av_2 = v_1 + 2v_2, \quad Av_3 = v_2 + 2v_3$$

In this basis the matrix of A is:

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

This is a Jordan basis.

0.4 Jordan Canonical (or Normal) Form in the one eigenvalue case

The Jordan canonical form of a linear map $T : V \rightarrow V$ is essentially the matrix of T with respect to a Jordan basis. We just need to order the vectors appropriately. Everything is over a field k , often k will be \mathbb{C} .

Suppose for the moment that $m_T = (x - \lambda)^b$, in particular T has only one eigenvalue λ . Choose a Jordan basis:

$$\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_b,$$

Of course, as $(A - \lambda id)^b = 0$, we have $V_b(\lambda) = V$.

We have a chain of subspaces

$$V_1(\lambda) \subset V_2(\lambda) \subset \dots \subset V_b(\lambda) = V$$

and the **pre-Jordan** basis was constructed by starting with a basis of $V_1(\lambda)$ and completing it successfully to get a basis of $V_b(\lambda) = V$. We then altered this basis so that

$$(T - \lambda id)\mathcal{B}_i \subset \mathcal{B}_{i-1}$$

Notice that we can arrange the basis elements in **chains** : starting with a vector $v \in \mathcal{B}_b$ we get a chain

$$v, (T - \lambda id)v, (T - \lambda id)^2v, \dots, (T - \lambda id)^{b-1}v$$

This last vector $w = (T - \lambda id)^{b-1}v$ is in $V_1(\lambda)$. Indeed

$$(T - \lambda id)^b v = 0$$

hence

$$(T - \lambda id)w = 0$$

therefore

$$Tw = \lambda w$$

therefore $w \in V_1(\lambda)$.

We have the following

Lemma 0.28 For any $v \in \mathcal{B}_b$ (in particular $v \notin \mathcal{B}_i$ for $i < b$!), the vectors

$$v, (T - \lambda id)v, (T - \lambda id)^2v, \dots, (T - \lambda id)^{b-1}v$$

are linearly independent.

Proof. Suppose that

$$\sum_i \mu_i (T - \lambda id)^i v = 0$$

Then

$$\mu_0 v + (T - \lambda id)w = 0$$

where w is a linear combination of $(T - \lambda id)^k v$. Multiplying by $(T - \lambda id)^{b-1}$, we get

$$\mu_0 (T - \lambda id)^{b-1} v = 0$$

but, as $v \notin V_{b-1}(\lambda)$, we see that

$$(T - \lambda id)^{b-1} v \neq 0$$

hence $\mu_0 = 0$.

Repeating the process inductively, we get that $\mu_i = 0$ for all i and the vectors we consider are linearly independent. \square

Let us number the vectors in this chain as $v_b = (T - \lambda id)^{b-1}v, \dots, v_0 = v$. In other words

$$v_i = (T - \lambda id)^{b-i} v$$

Then

$$(T - \lambda id)v_i = v_{i-1}$$

i.e.

$$Tv_i = \lambda v_i + v_{i-1}$$

In other words,

$$T(v_i) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

in the basis formed by this chain.

This gives a **Jordan block** i.e. $b \times b$ matrix:

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}.$$

In this way, we arrange our Jordan basis in chains starting with \mathcal{B}_i (for $i = b, b-1, \dots, 1$) and terminating at $V_1(\lambda)$.

By putting the chains together, we get that in the Jordan basis, the matrix is of the following form:

$$\begin{pmatrix} \lambda & 1 & & & & & & & & & \\ & \lambda & 1 & & & & & & & & \\ & & \lambda & 1 & & & & & & & \\ & & & \lambda & 1 & & & & & & \\ & & & & \lambda & 1 & & & & & \\ & & & & & \lambda & 1 & & & & \\ & & & & & & \lambda & 1 & & & \\ & & & & & & & \lambda & 1 & & \\ & & & & & & & & \lambda & 1 & \\ & & & & & & & & & \lambda & 1 \end{pmatrix}.$$

We had write it as

$$[T]_{\mathcal{B}} = \text{diag}(J_{h_1}(\lambda), \dots, J_{h_w}(\lambda)).$$

where the J_{h_i} s are blocks corresponding to a chain of length h_i .

We can actually say more; in fact the following results determines the number of blocks:

Lemma 0.29 *The number of blocks is the dimension of the eigenspace $V_1(\lambda)$.*

Proof. Let (v_1, \dots, v_k) be the Jordan basis of the subspace U corresponding to one block. It is a chain, we have

$$Tv_1 = \lambda v_1$$

and

$$Tv_i = \lambda v_i + v_{i-1}$$

for $2 \leq i \leq k$.

Let $v \in U$ be an eigenvector : $Tv = \lambda v$. Write $v = \sum_{i=1}^k c_i v_i$. Then

$$Tv = c_1 \lambda v_1 + \sum_{i \geq 2} c_i (\lambda v_i + v_{i-1}) = \lambda v + \sum_{i \geq 2} c_i v_i$$

It follows that $Tv = \lambda v$ if and only if $\sum_{i \geq 2} c_i v_i = 0$ which implies that $c_2 = \dots = c_n = 0$ and hence v is in the subspace generated by v_1 . Therefore, each block determines exactly one eigenvector for eigenvalue λ . As eigenvectors from different blocks are linearly independent : they are members of a basis, the number of blocks is exactly the dimension of the eigenspace $V_1(\lambda)$. \square

SUMMARY :

To summarise what we have seen so far. Suppose T has one eigenvalue λ , let $m_T(x) = (x - \lambda id)^b$ be its minimal polynomial. We construct a **pre-Jordan** basis by choosing a basis \mathcal{B}_1 for the eigenspace $V_1(\lambda)$ and then complete by \mathcal{B}_2 (a certain number of vectors in $V_2(\lambda)$) and then to $\mathcal{B}_3, \dots, \mathcal{B}_b$. Note that $V_b(\lambda) = V$. We get a pre-Jordan basis $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_b$.

Now we **alter** the pre-Jordan basis by doing the following. Start with a vector $v_b \in \mathcal{B}$, replace one of the vectors in \mathcal{B}_{b-1} by $v_{b-1} = (T - \lambda id)v_b$ making sure that this v_{b-1} is linearly independent of the other vectors in \mathcal{B}_{b-1} . Then replace a vector in \mathcal{B}_{b-2} by $v_{b-2} = (T - \lambda id)v_{b-1}$ (again choose a

vector to replace by choosing one such that v_{b-2} is linearly independent of the others)... continue until you get to $V_1(\lambda)$. The last vector will be $v_1 \in V_1(\lambda)$ i.e. v_1 is an eigenvector.

We obtain a chain of vectors

$$v_1 = (T - \lambda id)v_2, v_2 = (T - \lambda id)v_3, \dots, v_{b-1} = (T - \lambda id)v_b, v_b$$

Hence in particular

$$Tv_k = v_{k-1} + \lambda v_k$$

The subspace U spanned by this chain is T -stable (because $Tv_k = v_{k-1} + \lambda v_k$) and this chain is **linearly independent** hence the chain forms a basis of U . In restriction to U and with respect to this basis the matrix of T is

$$J(b)(\lambda) = \begin{pmatrix} \lambda & 1 & & & & & & & \\ & \lambda & 1 & & & & & & \\ & & \lambda & 1 & & & & & \\ & & & \lambda & 1 & & & & \\ & & & & \lambda & 1 & & & \\ & & & & & \lambda & 1 & & \\ & & & & & & \lambda & 1 & \\ & & & & & & & \lambda & 1 \\ & & & & & & & & \lambda \end{pmatrix}.$$

One constructs such chains with all elements of \mathcal{B}_b . Once done, one looks for elements in \mathcal{B}_{b-1} **which are not in the previously constructed chains starting at \mathcal{B}_b** and constructs chains with them. Then with \mathcal{B}_{b-2} , etc...

In the end, the union of chains will be a **Jordan basis** and in it the matrix of T is of the form :

$$\text{diag}(J_{h_1}(\lambda), \dots, J_{h_w}(\lambda)).$$

Notice the following two observations :

1. **There is always a block of size $b \times b$. Hence by knowing the degree of the minimal polynomial, in some cases it is possible to determine the shape of Jordan normal form.**
2. **The number of blocks is the dimension of the eigenspace $V_1(\lambda)$**

Here are some examples:

Suppose you have a matrix such that

$$ch_A = (x - \lambda)^5$$

and

$$m_A(x) = (x - \lambda)^4$$

There is always a block of size 4×4 , hence the Jordan normal form has one 4×4 block and one 1×1 block.

Suppose ch_A is the same but $m_A(x) = (x - \lambda)^3$. Here you need to know more. There is one 3×3 block and then either two 1×1 blocks or one 2×2 block. This is determined by the dimension of $V_1(\lambda)$. If it's three then the first possibility, if it's two then the second.

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 0 & 2 \end{pmatrix}$$

One calculates that $ch_A(x) = (x - 1)^4$. We have

$$A - I = \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix}$$

Clearly the rank of $A - I$ is 1, hence $\dim V_1(\lambda) = 3$.

This means that the Jordan normal form will have three blocks. Therefore there will be two blocks of size 1×1 and one of size 2×2 . The Jordan normal form is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Another example:

$$A = \begin{pmatrix} -2 & 0 & -1 & 1 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

One calculates that $ch_A(x) = (x + 2)^4$. We have

$$A + 2I = \begin{pmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We see that $(A + 2I)^2 = 0$ and therefore $m_A(x) = (x + 2)^2$. As there is always a block of size two, there are two possibilities : either two 2×2 blocks or one 2×2 and two 1×1 .

To decide which one it is, we see that the rank of $A + 2I$ is 2 hence the dimension of the kernel is 2. There are therefore 2 blocks and the Jordan normal form is

$$\begin{pmatrix} -2 & 1 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

0.5 Jordan canonical form in general.

Once we know how to determine Jordan canonical form in one eigenvalue case, the general case is easy. Let T be a linear transformation and $\lambda_1, \dots, \lambda_r$ its eigenvalues. Suppose that the minimal polynomial decomposes as

$$m_T(x) = \prod_{i=1}^r (x - \lambda_i)^{b_i}$$

(recall again that this is always true over \mathbb{C} .)

The we have seen that

$$V = V_{b_1}(\lambda_1) \oplus \dots \oplus V_{b_r}(\lambda_r)$$

and each $V_{b_i}(\lambda_i)$ is stable by T . Therefore in restriction to each $V_{b_i}(\lambda_i)$, T is a linear transformation with one eigenvalue, namely λ_i and the minimal polynomial of T restricted to $V_{b_i}(\lambda_i)$ is $(x - \lambda_i)^{b_i}$.

One gets the Jordan basis by taking the union of Jordan bases for each $V_{b_i}(\lambda_i)$ which are constructed as previously.

Let's look at an example.

$$A = \begin{pmatrix} -1 & 1 & 1 \\ -2 & 2 & 1 \\ -1 & 1 & 1 \end{pmatrix}$$

One calculates that $ch_A(x) = x(x - 1)^2$. Then 0 and 1 are the only eigenvalues and

$$V = V_1(0) \oplus V_2(1)$$

One finds that $V_1(0)$ is generated by

$$v_0 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

That will be the first vector of the basis.

For $\lambda = 1$.

We have

$$A - I = \begin{pmatrix} -2 & 1 & 1 \\ -2 & 1 & 1 \\ -1 & 1 & 0 \end{pmatrix}$$

We find that $V_1(\lambda)$ is spanned by

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Then

$$(A - I)^2 = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

and $V_2(\lambda)$ is spanned by

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

and

$$v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Notice that $(A - I)v_2 = v_1$ and therefore this is already a Jordan basis. The matrix in this basis is

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Another example :

$$A = \begin{pmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{pmatrix}$$

One calculates that $ch_A(x) = (x - 1)^2(x - 10)$. Then 1 and 10 are the only eigenvalues.

One finds

$$V_1(1) = \text{Span}(u_1 = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}, u_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix})$$

The dimension is two, therefore there will be two blocks of size 1×1 corresponding to the eigenvalue 1.

For $V_1(10)$, one finds

$$V_1(10) = \text{Span}(u_3 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix})$$

In the basis (u_1, u_2, u_3) , the matrix is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

It is diagonal, the matrix is diagonalisable, in fact $m_A = (x - 1)(x - 10)$.

And a last example : find Jordan basis and normal form of :

$$A = \begin{pmatrix} 4 & 0 & 1 & 0 \\ 2 & 2 & 3 & 0 \\ -1 & 0 & 2 & 0 \\ 4 & 0 & 1 & 2 \end{pmatrix}$$

One finds that the characteristic polynomial is $ch_A(x) = (x - 2)^2(x - 3)^2$.

Hence 2 and 3 are eigenvalues and we have

$$A - 2I = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 2 & 0 & 3 & 0 \\ -1 & 0 & 0 & 0 \\ 4 & 0 & 1 & 0 \end{pmatrix}$$

Clearly the dimension of the kernel is 2 and spanned by e_2 and e_4 .

The eigenspace has dimension two.

So we will have two blocks of size 1×1 corresponding to eigenvalue 2.

For the eigenvalue 3:

$$A - 3I = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 2 & -1 & 3 & 0 \\ -1 & 0 & -1 & 0 \\ 4 & 0 & 1 & -1 \end{pmatrix}$$

We see that rows one and three are identical, others are linearly independent. It follows that the eigenspace is one-dimensional and spanned by

$$u = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 3 \end{pmatrix}$$

We will have one block.

Let us calculate:

$$(A - 3I)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -3 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 1 \end{pmatrix}$$

We take the vector $v = \begin{pmatrix} 0 \\ 4 \\ 1 \\ -2 \end{pmatrix}$ to complete the basis of $\ker(A - 3I)^2$.

Now, we have $(A - 3I)v = u$ hence we already have a Jordan basis.

The basis (e_2, e_4, u, v) is a Jordan basis and in this basis the matrix

is

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Further linear algebra. Chapter V. Bilinear and quadratic forms.

Andrei Yafaev

1 Matrix Representation.

Definition 1.1 Let V be a vector space over k . A bilinear form on V is a function $f : V \times V \rightarrow k$ such that

- $f(u + \lambda v, w) = f(u, w) + \lambda f(v, w)$;
- $f(u, v + \lambda w) = f(u, v) + \lambda f(u, w)$.

I.e. $f(v, w)$ is linear in both v and w .

An obvious example is the following : take $V = \mathbb{R}$ and $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x, y) = xy$.

Notice here the difference between linear and bilinear : $f(x, y) = x + y$ is linear, $f(x, y) = xy$ is bilinear.

More generally $f(x, y) = \lambda xy$ is bilinear for any $\lambda \in \mathbb{R}$.

More generally still, given a matrix $A \in M_n(k)$, the following is a bilinear form on k^n :

$$f(v, w) = v^t A w = \sum_{i,j} v_i a_{i,j} w_j, \quad v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}.$$

We'll see that in fact all bilinear form are of this form.

Example 1.1 If $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ then the corresponding bilinear form is

$$f\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) = \begin{pmatrix} x_1 & y_1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = x_1 x_2 + 2x_1 y_2 + 3y_1 x_2 + 4y_1 y_2.$$

Recall that if $\mathcal{B} = \{b_1, \dots, b_n\}$ is a basis for V and $v = \sum x_i b_i$ then we write $[v]_{\mathcal{B}}$ for the column vector

$$[v]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Definition 1.2 If f is a bilinear form on V and $\mathcal{B} = \{b_1, \dots, b_n\}$ is a basis for V then we define the matrix of f with respect to \mathcal{B} by

$$[f]_{\mathcal{B}} = \begin{pmatrix} f(b_1, b_1) & \dots & f(b_1, b_n) \\ \vdots & & \vdots \\ f(b_n, b_1) & \dots & f(b_n, b_n) \end{pmatrix}$$

Proposition 1.2 Let \mathcal{B} be a basis for a finite dimensional vector space V over k , $\dim(V) = n$. Any bilinear form f on V is determined by the matrix $[f]_{\mathcal{B}}$. Moreover for $v, w \in V$,

$$f(v, w) = [v]_{\mathcal{B}}^t [f]_{\mathcal{B}} [w]_{\mathcal{B}}.$$

Proof. Let

$$v = x_1 b_1 + x_2 b_2 + \dots + x_n b_n,$$

so

$$[v]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Similarly suppose

$$[w]_{\mathcal{B}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Then

$$\begin{aligned}
 f(v, w) &= f\left(\sum_{i=1}^n x_i b_i, w\right) \\
 &= \sum_{i=1}^n x_i f(b_i, w) \\
 &= \sum_{i=1}^n x_i f\left(b_i, \sum_{j=1}^n y_j b_j\right) \\
 &= \sum_{i=1}^n x_i \sum_{j=1}^n y_j f(b_i, b_j) \\
 &= \sum_{i=1}^n \sum_{j=1}^n x_i y_j f(b_i, b_j) \\
 &= \sum_{i=1}^n \sum_{j=1}^n a_{i,j} x_i y_j \\
 &= [v]_{\mathcal{B}}^t [f]_{\mathcal{B}} [w]_{\mathcal{B}}.
 \end{aligned}$$

□

Let us give some examples.

Suppose that $f: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by

$$f\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) = 2x_1x_2 + 3x_1y_2 + x_2y_1$$

Let us write the matrix of f in the standard basis.

$$f(e_1, e_1) = 2, f(e_1, e_2) = 3, f(e_2, e_1) = 1, f(e_2, e_2) = 0$$

hence the matrix in the standard basis is

$$\begin{pmatrix} 2 & 3 \\ 1 & 0 \end{pmatrix}$$

Now suppose $\mathcal{B} = \{b_1, \dots, b_n\}$ and $\mathcal{C} = \{c_1, \dots, c_n\}$ are two bases for V . We may write one basis in terms of the other:

$$c_i = \sum_{j=1}^n \lambda_{j,i} b_j.$$

The matrix

$$M = \begin{pmatrix} \lambda_{1,1} & \dots & \lambda_{1,n} \\ \vdots & & \vdots \\ \lambda_{n,1} & \dots & \lambda_{n,n} \end{pmatrix}$$

is called the transition matrix from \mathcal{B} to \mathcal{C} . It is always an invertible matrix: its inverse is the transition matrix from \mathcal{C} to \mathcal{B} . Recall that for any vector $v \in V$ we have

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

and for any linear map $T : V \rightarrow V$ we have

$$[T]_{\mathcal{C}} = M^{-1}[T]_{\mathcal{B}}M.$$

We'll now describe how bilinear forms behave under change of basis.

Theorem 1.3 (Change of Basis Formula) *Let f be a bilinear form on a finite dimensional vector space V over k . Let \mathcal{B} and \mathcal{C} be two bases for V and let M be the transition matrix from \mathcal{B} to \mathcal{C} .*

$$[f]_{\mathcal{C}} = M^t[f]_{\mathcal{B}}M.$$

Proof. Let $u, v \in V$ with $[u]_{\mathcal{B}} = x$, $[v]_{\mathcal{B}} = y$, $[u]_{\mathcal{C}} = s$ and $[v]_{\mathcal{C}} = t$.

Let $A = (a_{i,j})$ be the matrix representing f with respect to \mathcal{B} .

Now $x = Ms$ and $y = Mt$ so

$$\begin{aligned} f(u, v) &= (Ms)^t A (Mt) \\ &= (s^t M^t) A (Mt) \\ &= s^t (M^t A M) t. \end{aligned}$$

We have $f(b_i, b_j) = (M^t A M)_{i,j}$. Hence

$$[f]_{\mathcal{C}} = M^t A M = M^t [f]_{\mathcal{B}} M.$$

□

For example, let f be the linear form from the previous example. It is given by

$$\begin{pmatrix} 2 & 3 \\ 1 & 0 \end{pmatrix}$$

in the standard basis. We want to write this matrix in the basis

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

The transition matrix is :

$$M = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

it's transpose is the same. The matrix of f in the new basis is

$$\begin{pmatrix} 6 & 3 \\ 5 & 2 \end{pmatrix}$$

2 Symmetric bilinear forms and quadratic forms.

As before let V be a finite dimensional vector space over a field k .

Definition 2.1 A bilinear form f on V is called symmetric if it satisfies $f(v, w) = f(w, v)$ for all $v, w \in V$.

Definition 2.2 Given a symmetric bilinear form f on V , the associated quadratic form is the function $q(v) = f(v, v)$.

Notice that q has the property that $q(\lambda v) = \lambda^2 q(v)$.

For example, take the bilinear form f defined by

$$\begin{pmatrix} 6 & 0 \\ 0 & 5 \end{pmatrix}$$

The corresponding quadratic form is

$$q\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = 6x^2 + 5y^2$$

Proposition 2.1 Let f be a bilinear form on V and let \mathcal{B} be a basis for V . Then f is a symmetric bilinear form if and only if $[f]_{\mathcal{B}}$ is a symmetric matrix (that means $a_{i,j} = a_{j,i}$).

Proof. This is because $f(e_i, e_j) = f(e_j, e_i)$. □

Theorem 2.2 (Polarization Theorem) *If $1 + 1 \neq 0$ in k then for any quadratic form q the underlying symmetric bilinear form is unique.*

Proof. If $u, v \in V$ then

$$\begin{aligned} q(u + v) &= f(u + v, u + v) \\ &= f(u, u) + 2f(u, v) + f(v, v) \\ &= q(u) + q(v) + 2f(u, v). \end{aligned}$$

So $f(u, v) = \frac{1}{2}(q(u + v) - q(u) - q(v))$. □

Let's look at an example :

Consider

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$$

it is a symmetric matrix.

Let f be the corresponding bilinear form. We have

$$f((x_1, y_1), (x_2, y_2)) = 2x_1x_2 + x_1y_2 + x_2y_1$$

and

$$q(x, y) = 2x^2 + 2xy = f((x, y), (x, y))$$

Let $u = (x_1, y_1), v = (x_2, y_2)$ and let us calculate

$$\begin{aligned} \frac{1}{2}(q(u+v) - q(u) - q(v)) &= \frac{1}{2}(2(x_1+x_2)^2 + 2(x_1+x_2)(y_1+y_2) - x_1^2 - 2x_1y_1 - 2x_2^2 - 2x_1y_2) \\ &= \frac{1}{2}(4x_1x_2 + 2(x_1y_2 + x_2y_1)) = f((x_1, y_1), (x_2, y_2)) \end{aligned}$$

If $A = (a_{i,j})$ is a symmetric matrix, then the corresponding form is

$$f(x, y) = \sum_i a_{i,i}x_iy_i + \sum_{i < j} a_{i,j}(x_iy_j + x_jy_i)$$

and the corresponding quadratic form is

$$q(x) = \sum_{i=1}^n a_{i,i}x_i^2 + 2 \sum_{i < j} a_{i,j}x_ix_j$$

then the symmetric matrix $A = (a_{i,j})$ is the matrix representing the underlying bilinear form f .

3 Orthogonality and diagonalisation

Definition 3.1 Let V be a vector space over k with a symmetric bilinear form f . We call two vectors $v, w \in V$ orthogonal if $f(v, w) = 0$. It is a good idea to imagine this means that v and w are at right angles to each other. This is written $v \perp w$. If $S \subset V$ be a non-empty subset, then the orthogonal complement of S is defined to be

$$S^\perp = \{v \in V : \forall w \in S, w \perp v\}.$$

Proposition 3.1 S^\perp is a subspace of V .

Proof. Let $v, w \in S^\perp$ and $\lambda \in k$. Then for any $u \in S$ we have

$$f(v + \lambda w, u) = f(v, u) + \lambda f(w, u) = 0.$$

Therefore $v + \lambda w \in S^\perp$. □

Definition 3.2 A basis \mathcal{B} is called an orthogonal basis if any two distinct basis vectors are orthogonal. Thus \mathcal{B} is an orthogonal basis if and only if $[f]_{\mathcal{B}}$ is diagonal.

Theorem 3.2 (Diagonalisation Theorem) Let f be a symmetric bilinear form on a finite dimensional vector space V over a field k in which $1 + 1 \neq 0$. Then there is an orthogonal basis \mathcal{B} for V ; i.e. a basis such that $[f]_{\mathcal{B}}$ is a diagonal matrix.

Notice that the existence of an orthogonal basis is indeed equivalent to the matrix being diagonal.

Let $B = \{v_1, \dots, v_n\}$ be an orthogonal basis. By definition $f(v_i, v_j) = 0$ if $i \neq j$ hence the only possible non-zero values are $f(v_i, v_i)$ i.e. on the diagonal.

And of course the converse holds : if the matrix is diagonal, then $f(v_i, v_j) = 0$ if $i \neq j$.

The quadratic form associated to such a bilinear form is

$$q(x_1, \dots, x_n) = \sum_i \lambda_i x_i^2$$

where λ_i s are elements on the diagonal.

Let U, W be two subspaces of V . The sum of U and W is the subspace

$$U + W = \{u + w : u \in U, w \in W\}.$$

We call this a direct sum $U \oplus W$ if $U \cap W = \{0\}$. This is the same as saying that every element of $U + W$ can be written uniquely as $u + w$ with $u \in U$ and $w \in W$.

Theorem 3.3 (Key Lemma) *Let $v \in V$ and assume that $q(v) \neq 0$. Then*

$$V = \text{Span}\{v\} \oplus \{v\}^\perp.$$

Proof. For $w \in V$, let

$$w_1 = \frac{f(v, w)}{f(v, v)}v, \quad w_2 = w - \frac{f(v, w)}{f(v, v)}v.$$

Clearly $w = w_1 + w_2$ and $w_1 \in \text{Span}\{v\}$. Note also that

$$f(w_2, v) = f\left(w - \frac{f(v, w)}{f(v, v)}v, v\right) = f(w, v) - \frac{f(v, w)}{f(v, v)}f(v, v) = 0.$$

Therefore $w_2 \in \{v\}^\perp$. It follows that $\text{Span}\{v\} + \{v\}^\perp = V$. To prove that the sum is direct, suppose that $w \in \text{Span}\{v\} \cap \{v\}^\perp$. Then $w = \lambda v$ for some $\lambda \in k$ and we have $f(w, v) = 0$. Hence $\lambda f(v, v) = 0$. Since $q(v) = f(v, v) \neq 0$ it follows that $\lambda = 0$ so $w = 0$. \square

Proof. [of the theorem] We use induction on $\dim(V) = n$. If $n = 1$ then the theorem is true, since any 1×1 matrix is diagonal. So suppose the result holds for vector spaces of dimension less than $n = \dim(V)$.

If $f(v, v) = 0$ for every $v \in V$ then using Theorem 5.3 for any basis \mathcal{B} we have $[f]_{\mathcal{B}} = [0]$, which is diagonal. [This is true since

$$f(e_i, e_j) = \frac{1}{2}(f(e_i + e_j, e_i + e_j) - f(e_i, e_i) - f(e_j, e_j)) = 0.]$$

So we can suppose there exists $v \in V$ such that $f(v, v) \neq 0$. By the Key Lemma we have

$$V = \text{Span}\{v\} \oplus \{v\}^\perp.$$

Since $\text{Span}\{v\}$ is 1-dimensional, it follows that $\{v\}^\perp$ is $n - 1$ -dimensional. Hence by the inductive hypothesis there is an orthonormal basis $\{b_1, \dots, b_{n-1}\}$ of $\{v\}^\perp$.

Now let $\mathcal{B} = \{b_1, \dots, b_{n-1}, v\}$. This is a basis for V . Any two of the vectors b_i are orthogonal by definition. Furthermore $b_i \in \{v\}^\perp$, so $b_i \perp v$. Hence \mathcal{B} is an orthogonal basis. \square

4 Examples of Diagonalisation.

Definition 4.1 Two matrices $A, B \in M_n(k)$ are congruent if there is an invertible matrix P such that

$$B = P^t A P.$$

We have shown that if \mathcal{B} and \mathcal{C} are two bases then for a bilinear form f , the matrices $[f]_{\mathcal{B}}$ and $[f]_{\mathcal{C}}$ are congruent.

Theorem 4.1 Let $A \in M_n(k)$ be symmetric, where k is a field in which $1 + 1 \neq 0$, then A is congruent to a diagonal matrix.

Proof. This is just the matrix version of the previous theorem. □

We shall next find out how to calculate the diagonal matrix congruent to a given symmetric matrix.

There are three kinds of row operation:

- swap rows i and j ;
- multiply $row(i)$ by $\lambda \neq 0$;
- add $\lambda \times row(i)$ to $row(j)$.

To each row operation there is a corresponding elementary matrix E ; the matrix E is the result of doing the row operation to I_n . The row operation transforms a matrix A into EA .

We may also define three corresponding column operations:

- swap columns i and j ;
- multiply $column(i)$ by $\lambda \neq 0$;
- add $\lambda \times column(i)$ to $column(j)$.

Doing a column operation to A is the same as doing the corresponding row operation to A^t . We therefore obtain $(EA^t)^t = AE^t$.

Definition 4.2 By a double operation we shall mean a row operation followed by the corresponding column operation.

If E is the corresponding elementary matrix then the double operation transforms a matrix A into EAE^t .

Lemma 4.2 *If we do a double operation to A then we obtain a matrix congruent to A .*

Proof. EAE^t is congruent to A . □

Recall that a symmetric bilinear forms are represented by symmetric matrices. If we change the basis then we will obtain a congruent matrix. We've seen that if we do a double operation to matrix A then we obtain a congruent matrix. This corresponds to the same quadratic form with respect to a different basis. We can always do a sequence of double operations to transform any symmetric matrix into a diagonal matrix.

Example 4.3 *Consider the quadratic form $q(x, y)^t = x^2 + 4xy + 3y^2$*

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

This shows that there is a basis $\mathcal{B} = \{b_1, b_2\}$ such that

$$q(xb_1 + yb_2) = x^2 - y^2.$$

Notice that when we have done the first operation, we have multiplied A on the left by $E_{2,1}(-2) = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$ and when we have done the second, we have

multiplied on the right by $E_{2,1}(-2)^t = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$

We find that

$$E_{2,1}(-2)AE_{2,1}(-2)^t = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Hence in the basis

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

the matrix of the corresponding quadratic form is $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

Example 4.4 Consider the quadratic form $q(x, y)^t = 4xy + y^2$

$$\begin{aligned} \begin{pmatrix} 0 & 2 \\ 2 & 1 \end{pmatrix} &\rightarrow \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & 2 \\ 0 & -4 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned}$$

This shows that there is a basis $\{b_1, b_2\}$ such that

$$q(xb_1 + yb_2) = x^2 - y^2.$$

The last step in the previous example transformed the -4 into a -1. In general, once we have a diagonal matrix we are free to multiply or divide the diagonal entries by squares:

Lemma 4.5 For $\mu_1, \dots, \mu_n \in k^\times = k \setminus \{0\}$ and $\lambda_1, \dots, \lambda_n \in k$

$$D(\lambda_1, \dots, \lambda_n) \text{ is congruent to } D(\mu_1^2 \lambda_1, \dots, \mu_n^2 \lambda_n).$$

Proof. Since $\mu_1, \dots, \mu_n \in k \setminus \{0\}$ then $\mu_1 \cdots \mu_n \neq 0$. So

$$P = D(\mu_1, \dots, \mu_n)$$

is invertible. Then

$$\begin{aligned} P^t D(\lambda_1, \dots, \lambda_n) P &= D(\mu_1, \dots, \mu_n) D(\lambda_1, \dots, \lambda_n) D(\mu_1, \dots, \mu_n) \\ &= D(\mu_1^2 \lambda_1, \dots, \mu_n^2 \lambda_n). \end{aligned}$$

□

Definition 4.3 Two bilinear forms f, f' are equivalent if they are the same up to a change of basis.

Definition 4.4 The rank of a bilinear form f is the rank $[f]_{\mathcal{B}}$ for any basis \mathcal{B} .

Clearly if f and f' have different rank then they are not equivalent.

5 Canonical forms over \mathbb{C}

Definition 5.1 Let q be a quadratic form on vector space V over \mathbb{C} , and suppose there is a basis \mathcal{B} of V such that

$$[q]_{\mathcal{B}} = \begin{pmatrix} I_r & \\ & 0 \end{pmatrix}.$$

We call the matrix $\begin{pmatrix} I_r & \\ & 0 \end{pmatrix}$ a canonical form of q (over \mathbb{C}).

Theorem 5.1 (Canonical forms over \mathbb{C}) Let V be a finite dimensional vector space over \mathbb{C} and let q be a quadratic form on V . Then q has exactly one canonical form.

Proof. (Existence) We first choose an orthogonal basis $\mathcal{B} = \{b_1, \dots, b_n\}$. After reordering the basis we may assume that $q(b_1), \dots, q(b_r) \neq 0$ and $q(b_{r+1}), \dots, q(b_n) = 0$. Since every complex number has a square root in \mathbb{C} , we may divide b_i by $\sqrt{q(b_i)}$ if $i \leq r$.

(Uniqueness) Change of basis does not change the rank. □

Corollary 5.2 Two quadratic forms over \mathbb{C} are equivalent iff they have the same canonical form.

6 Canonical forms over \mathbb{R}

Definition 6.1 Let q be a quadratic form on vector space V over \mathbb{R} , and suppose there is a basis \mathcal{B} of V such that

$$[q]_{\mathcal{B}} = \begin{pmatrix} I_r & & \\ & -I_s & \\ & & 0 \end{pmatrix}.$$

We call the matrix $\begin{pmatrix} I_r & & \\ & -I_s & \\ & & 0 \end{pmatrix}$ a canonical form of q (over \mathbb{R}).

Theorem 6.1 (Sylvester's Law of Inertia) Let V be a finite dimensional vector space over \mathbb{R} and let q be a quadratic form on V . Then q has exactly one (real) canonical form.

Proof. (existence) Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be an orthogonal basis. We can reorder the basis so that

$$q(b_1), \dots, q(b_r) > 0, \quad q(b_{r+1}), \dots, q(b_{r+s}) < 0, \quad q(b_{r+s+1}), \dots, q(b_n) = 0.$$

Then define a new basis by

$$c_i = \begin{cases} \frac{1}{\sqrt{|q(b_i)|}} b_i & i \leq r + s, \\ b_i & i > r + s. \end{cases}$$

The matrix of q with respect to \mathcal{C} is a canonical form.

(uniqueness) Suppose we have two bases \mathcal{B} and \mathcal{C} with

$$[q]_{\mathcal{B}} = \begin{pmatrix} I_r & & \\ & -I_s & \\ & & 0 \end{pmatrix}, \quad [q]_{\mathcal{C}} = \begin{pmatrix} I_{r'} & & \\ & -I_{s'} & \\ & & 0 \end{pmatrix}.$$

By comparing the ranks we know that $r + s = r' + s'$. It's therefore sufficient to prove that $r = r'$. Define two subspaces of V by

$$U = \text{Span}\{b_1, \dots, b_r\}, \quad W = \text{Span}\{c_{r'+1}, \dots, c_n\}.$$

If u is a non-zero vector of U then we have $u = x_1 b_1 + \dots + x_r b_r$. Hence

$$q(u) = x_1^2 + \dots + x_r^2 > 0.$$

Similarly if $w \in W$ then $w = y_{r'+1} c_{r'+1} + \dots + y_n c_n$, and

$$q(w) = -y_{r'+1}^2 - \dots - y_{r'+s'}^2 \leq 0.$$

It follows that $U \cap W = \{0\}$. Therefore

$$U + W = U \oplus W \subset V.$$

From this we have

$$\dim U + \dim W \leq \dim V.$$

Hence

$$r + (n - r') \leq n.$$

This implies $r \leq r'$. A similar argument (consider $U = \text{Span}\{c_1, \dots, c_{r'}\}$ and $W = \text{Span}\{b_{r+1}, \dots, b_n\}$) shows that $r' \leq r$, so we have $r = r'$. \square

The **rank** of a quadratic form is the rank of the corresponding matrix. Clearly, in the complex case it is the integer r that appears in the canonical form.

In the real case, it is $r + s$.

For a **real** quadratic form, the signature is the pair (r, s) . In this case $q(v) > 0$ for all non-zero vectors v .

A real form q is **positive definite** if its signature is $(r, 0)$, **negative definite** if its signature is $(0, s)$. In this case $q(v) < 0$ for all non-zero vectors v .

There exists a non-zero vector v such that $q(v) = 0$ if and only if the signature is (r, s) with $r > 0$ and $s > 0$.

Further linear algebra. Chapter VI. Inner product spaces.

Andrei Yafaev

1 Geometry of Inner Product Spaces

Definition 1.1 Let V be a vector space over \mathbb{R} and let $\langle -, - \rangle$ be a symmetric bilinear form on V . We shall call the form positive definite if for all non-zero vectors $v \in V$ we have

$$\langle v, v \rangle > 0.$$

Notice that a symmetric bilinear form is positive definite if and only if its canonical form (over \mathbb{R}) is I_n .

Clearly $x_1^2 + \dots + x_n^2$ is positive definite on \mathbb{R}^n . Conversely, suppose \mathcal{B} is a basis such that the matrix with respect to \mathcal{B} is the canonical form. For any basis vector b_i , the diagonal entry satisfies $\langle b_i, b_i \rangle > 0$ and hence $\langle b_i, b_i \rangle = 1$.

Definition 1.2 Let V be a vector space over \mathbb{C} . A Hermitian form on V is a function $\langle -, - \rangle : V \times V \rightarrow \mathbb{C}$ such that:

- For all $u, v, w \in V$ and all $\lambda \in \mathbb{C}$,

$$\langle u + \lambda v, w \rangle = \langle u, w \rangle + \lambda \langle v, w \rangle;$$

- For all $u, v \in V$,

$$\langle u, v \rangle = \overline{\langle v, u \rangle}.$$

Example 1.1 The simplest example is the following : take $V = \mathbb{C}$, then $\langle z, w \rangle = z\bar{w}$ is a hermitian form on \mathbb{C} .

A matrix $A \in M_n(\mathbb{C})$ is called a Hermitian matrix if $A^t = \bar{A}$. Here \bar{A} is the matrix obtained from A by applying complex conjugation to the entries.

If A is a Hermitian matrix then the following is a Hermitian form on \mathbb{C}^n :

$$\langle v, w \rangle = v^t A \bar{w}.$$

In fact every Hermitian form on \mathbb{C}^n is one of these.

To see why, suppose we are given a Hermitian form \langle, \rangle . Choose a basis $B = (b_1, \dots, b_n)$. Let $v = \sum_i \lambda_i b_i$ and $w = \sum_j \mu_j b_j$. We calculate

$$\langle v, w \rangle = \left\langle \sum_i \lambda_i b_i, \sum_j \mu_j b_j \right\rangle = \sum_{i,j} \lambda_i \bar{\mu}_j \langle b_i, b_j \rangle = v^t A \bar{w}$$

where $A = (\langle b_i, b_j \rangle)$. Of course $A^t = \bar{A}$ because $\langle b_i, b_j \rangle = \overline{\langle b_j, b_i \rangle}$.

A matrix A satisfying $A^t = \bar{A}$ is called hermitian.

Example 1.2 If $V = \mathbb{R}^n$, then \langle, \rangle defined by $\langle x_1, \dots, x_n, y_1, \dots, y_n \rangle = \sum_{i,j} x_i y_j$ is called the standard inner product.

If $V = \mathbb{C}^n$, then \langle, \rangle defined by $\langle z_1, \dots, z_n, w_1, \dots, w_n \rangle = \sum_{i,j} z_i \bar{w}_j$ is called the standard (hermitian) inner product.

Note that a Hermitian form is conjugate-linear in the second variable, i.e.

$$\langle u, v + \lambda w \rangle = \langle u, v \rangle + \bar{\lambda} \langle u, w \rangle.$$

Note also that by the second axiom

$$\langle u, u \rangle \in \mathbb{R}.$$

Definition 1.3 A Hermitian form is positive definite if for all non-zero vectors v we have

$$\langle v, v \rangle > 0.$$

In other words, $\langle v, v \rangle \geq 0$ for all v and $\langle v, v \rangle = 0$ if and only if $v = 0$.

Clearly, the form $z \bar{w}$ is positive definite.

Definition 1.4 By an inner product space we shall mean one of the following:

either A finite dimensional vector space V over \mathbb{R} with a positive definite symmetric bilinear form;

or A finite dimensional vector space V over \mathbb{C} with a positive definite Hermitian form.

We shall often write K to mean the field \mathbb{R} or \mathbb{C} , depending on which is relevant.

Example 1.3 Consider the vector space V of all continuous functions $[0, 1] \rightarrow \mathbb{C}$.

Then we can define

$$\langle f, g \rangle = \int_0^1 f(x)\overline{g(x)}dx.$$

This defines an inner product on V (easy exercise).

Another example. Let $V = M_n(\mathbb{R})$ the vector space of $n \times n$ -matrices with real entries. Then

$$\langle A, B \rangle = \text{tr}(AB^t)$$

is an inner product on V .

Similarly, if $V = M_n(\mathbb{C})$ and $\langle A, B \rangle = \text{tr}(A\overline{B}^t)$ is an inner product.

Definition 1.5 Let V be an inner product space. We define the norm of a vector $v \in V$ by

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

Lemma 1.4 For $\lambda \in K$ we have $\lambda\overline{\lambda} = |\lambda|^2$ for for $v \in V$ we have $\|\lambda v\| = |\lambda| \|v\|$.

The proof is obvious.

Theorem 1.5 (Cauchy-Schwarz inequality) If V is an inner product space then

$$\forall u, v \in V \quad |\langle u, v \rangle| \leq \|u\| \cdot \|v\|.$$

Proof. If $v = 0$ then the result holds so suppose $v \neq 0$. We have for all $\lambda \in K$,

$$\langle u - \lambda v, u - \lambda v \rangle \geq 0.$$

Expanding this out we have:

$$\|u\|^2 - \lambda\langle v, u \rangle - \overline{\lambda}\langle u, v \rangle + |\lambda|^2\|v\|^2 \geq 0.$$

Setting $\lambda = \frac{\langle u, v \rangle}{\|v\|^2}$ we have:

$$\|u\|^2 - \frac{\langle u, v \rangle}{\|v\|^2} \langle v, u \rangle - \frac{\langle v, u \rangle}{\|v\|^2} \langle u, v \rangle + \left| \frac{\langle u, v \rangle}{\|v\|^2} \right|^2 \|v\|^2 \geq 0.$$

Multiplying by $\|v\|^2$ we get

$$\|u\|^2 \|v\|^2 - 2|\langle u, v \rangle|^2 + |\langle u, v \rangle|^2 \geq 0.$$

Hence

$$\|u\|^2 \|v\|^2 \geq |\langle u, v \rangle|^2.$$

Taking the square root of both sides we get the result. \square

Theorem 1.6 (Triangle inequality) *If V is an inner product space with norm $\|\cdot\|$ then*

$$\forall u, v \in V \quad \|u + v\| \leq \|u\| + \|v\|.$$

Proof. We have

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v \rangle \\ &= \|u\|^2 + 2\Re \langle u, v \rangle + \|v\|^2. \end{aligned}$$

Notice that $|\Re \langle u, v \rangle| \leq |\langle u, v \rangle|$ hence

$$\|u + v\|^2 \leq \|u\|^2 + 2|\langle u, v \rangle| + \|v\|^2$$

So the Cauchy–Schwarz inequality implies that

$$\|u + v\|^2 \leq \|u\|^2 + 2\|u\| \|v\| + \|v\|^2 = (\|u\| + \|v\|)^2.$$

Hence

$$\|u + v\| \leq \|u\| + \|v\|.$$

\square

Definition 1.6 *Two vectors v, w in an inner product space are called orthogonal if $\langle v, w \rangle = 0$.*

Theorem 1.7 (Pythagoras' Theorem) Let (V, \langle, \rangle) be an inner product space. If $v, w \in V$ are orthogonal, then

$$\|v\|^2 + \|w\|^2 = \|v + w\|^2$$

Proof. Since

$$\|v + w\|^2 = \langle v + w, v + w \rangle = \|v\|^2 + 2\Re\langle v, w \rangle + \|w\|^2,$$

so we have

$$\|v\|^2 + \|w\|^2 = \|v + w\|^2$$

if $\langle v, w \rangle = 0$. □

2 Gram–Schmidt Orthogonalisation

Definition 2.1 Let V be an inner product space. We shall call a basis \mathcal{B} of V an orthonormal basis if $\langle b_i, b_j \rangle = \delta_{i,j}$.

Proposition 2.1 If \mathcal{B} is an orthonormal basis then for $v, w \in V$ we have:

$$\langle v, w \rangle = [v]_{\mathcal{B}}^t \overline{[w]_{\mathcal{B}}}.$$

Proof. If the basis $\mathcal{B} = (b_1, \dots, b_n)$ is orthonormal, then the matrix of \langle, \rangle in this basis is the identity I_n . The proposition follows. □

Theorem 2.2 (Gram–Schmidt Orthogonalisation) Let \mathcal{B} be any basis. Then the basis \mathcal{C} defined by

$$\begin{aligned} c_1 &= b_1 \\ c_2 &= b_2 - \frac{\langle b_2, c_1 \rangle}{\langle c_1, c_1 \rangle} c_1 \\ c_3 &= b_3 - \frac{\langle b_3, c_1 \rangle}{\langle c_1, c_1 \rangle} c_1 - \frac{\langle b_3, c_2 \rangle}{\langle c_2, c_2 \rangle} c_2 \\ &\vdots \\ c_n &= b_n - \sum_{r=1}^{n-1} \frac{\langle b_n, c_r \rangle}{\langle c_r, c_r \rangle} c_r, \end{aligned}$$

is orthogonal. Furthermore the basis \mathcal{D} defined by

$$d_r = \frac{1}{\|c_r\|} c_r,$$

is orthonormal.

Proof. Clearly each b_i is a linear combination of \mathcal{C} , so \mathcal{C} spans V . As the cardinality of \mathcal{C} is $\dim V$, \mathcal{C} is a basis. It follows also that \mathcal{D} is a basis. We'll prove by induction that $\{c_1, \dots, c_r\}$ is orthogonal. Clearly any one vector is orthogonal. Suppose $\{c_1, \dots, c_{r-1}\}$ are orthogonal. Then for $s < r$ we have

$$\langle c_r, c_s \rangle = \langle b_r, c_s \rangle - \sum_{t=1}^{r-1} \frac{\langle b_r, c_t \rangle}{\langle c_t, c_t \rangle} \langle c_t, c_s \rangle.$$

By the inductive hypothesis we have

$$\langle c_r, c_s \rangle = \langle b_r, c_s \rangle - \frac{\langle b_r, c_s \rangle}{\langle c_s, c_s \rangle} \langle c_s, c_s \rangle = \langle b_r, c_s \rangle - \langle b_r, c_s \rangle = 0.$$

(notice that $\langle c_t, c_s \rangle = 0$ unless $t = s$). This shows that $\{c_1, \dots, c_r\}$ are orthogonal. Hence \mathcal{C} is an orthogonal basis. It follows easily that \mathcal{D} is orthonormal. \square

This theorem shows in particular that an orthonormal basis **always exists**. Indeed, take any basis and turn it into an orthonormal one by applying Gram-Schmidt process to it.

Proposition 2.3 *If V is an inner product space with an orthonormal basis $\mathcal{B} = \{b_1, \dots, b_n\}$, then any $v \in V$ can be written as $v = \sum_{i=1}^n \langle v, e_i \rangle e_i$.*

Proof. We have $v = \sum_{i=1}^n \lambda_i e_i$ and $\langle v, e_j \rangle = \sum_{i=1}^n \lambda_i \langle e_i, e_j \rangle = \lambda_j$. \square

Definition 2.2 *Let S be a subspace of an inner product space V . The orthogonal complement of S is defined to be*

$$S^\perp = \{v \in V : \forall w \in S \langle v, w \rangle = 0\}.$$

Theorem 2.4 *If (V, \langle, \rangle) is an inner product space and W is a subspace of V then*

$$V = W \oplus W^\perp,$$

and hence any $v \in V$ can be written as

$$v = w + w^\perp,$$

for unique $w \in W$ and $w^\perp \in W^\perp$.

Proof. We show first that $V = W + W^\perp$.

Let $\mathcal{E} = \{e_1, \dots, e_n\}$ be an orthonormal basis for V , such that $\{e_1, \dots, e_r\}$ is a basis for W . This can be constructed by Gram-Schmidt orthogonalisation. (choose a basis $\{b_1, \dots, b_r\}$ for W and complete to a basis $\{b_1, \dots, b_n\}$ of V .)

Then apply Gram-Schmidt process. Notice that in Gram-Schmidt process, when constructing orthonormal basis, the vectors c_1, \dots, c_k lie in the space generated by c_1, \dots, c_{k-1}, b_k . It follows that the process will give an orthonormal basis e_1, \dots, e_n such that e_1, \dots, e_r is an orthonormal basis of W .)

If $v \in V$ then

$$v = \sum_{i=1}^r \lambda_i e_i + \sum_{i=r+1}^n \lambda_i e_i.$$

Now

$$\sum_{i=1}^r \lambda_i e_i \in W.$$

If $w \in W$ then there exist $\mu_i \in \mathbb{R}$ such that

$$w = \sum_{i=1}^r \mu_i e_i.$$

So

$$\left\langle w, \sum_{j=r+1}^n \lambda_j e_j \right\rangle = \sum_{i=1}^r \sum_{j=r+1}^n \mu_i \lambda_j \langle e_i, e_j \rangle = 0.$$

Hence

$$\sum_{i=r+1}^n \lambda_i e_i \in W^\perp.$$

Therefore

$$V = W + W^\perp.$$

Next suppose $v \in W \cap W^\perp$. So $\langle v, v \rangle = 0$ and so $v = 0$.

Hence $V = W \oplus W^\perp$ and so any vector $v \in V$ can be expressed uniquely as

$$v = w + w^\perp,$$

where $w \in W$ and $w^\perp \in W^\perp$. □

3 Adjoints.

Definition 3.1 An adjoint of a linear map $T : V \rightarrow V$ is a linear map T^* such that $\langle T(u), v \rangle = \langle u, T^*(v) \rangle$ for all $u, v \in V$.

Theorem 3.1 (existence and uniqueness) Every $T : V \rightarrow V$ has a unique adjoint. If T is represented by A (w.r.t. an orthonormal basis) then T^* is represented by \bar{A}^t .

Proof. (Existence) Let T^* be the linear map represented by \bar{A}^t . We'll prove that it is an adjoint of A .

$$\langle Tv, w \rangle = [v]^t A^t [\bar{w}] = [v]^t \bar{A}^t [\bar{w}] = \langle v, T^*w \rangle.$$

Notice that here we have used that the basis is orthonormal : we said that the matrix of \langle, \rangle was the identity. (Uniqueness) Let T^*, T' be two adjoints. Then we have

$$\langle u, (T^* - T')v \rangle = 0.$$

for all $u, v \in V$. In particular, let $u = (T^* - T')v$, then $\|(T^* - T')v\| = 0$ hence $T^*(v) = T'(v)$ for all $v \in V$. Therefore $T^* = T'$. □

Example 3.2 Consider $V = \mathbb{C}^2$ with the standard orthonormal basis and let T be represented by

$$A = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$$

Then $T^* = T$ (such a linear map is called autoadjoint).

Notice that T being self-adjoint is equivalent to the matrix representing it being hermitian

$$A = \begin{pmatrix} 2i & 1+i \\ -1+i & i \end{pmatrix}$$

Then $T^* = -T$

We also see that $T^{**} = T$ (using that T^* is represented by \overline{A}^t).

4 Isometries.

Theorem 4.1 *If $T : V \rightarrow V$ be a linear map of an inner product space V then the following are equivalent.*

- (i) $TT^* = \text{Id}$ (i.e. $T^* = T^{-1}$).
- (ii) $\forall u, v \in V \quad \langle Tu, Tv \rangle = \langle u, v \rangle$. (i.e. T preserves the inner product.)
- (iii) $\forall v \in V \quad \|Tv\| = \|v\|$. (i.e. T preserves the norm.)

Definition 4.1 *If T satisfies any of the above (and so all of them) then T is called an **isometry**.*

We also see that $T^{**} = T$ (using that T^* is represented by \overline{A}^t).

Proof. (i) \implies (ii)

Let $u, v \in V$ then

$$\langle Tu, Tv \rangle = \langle u, T^*Tv \rangle = \langle u, v \rangle,$$

since $T^* = T^{-1}$.

(ii) \implies (iii)

If $v \in V$ then

$$\|Tv\|^2 = \langle Tv, Tv \rangle$$

so by (ii)

$$\|Tv\|^2 = \langle v, v \rangle = \|v\|^2.$$

Hence $\|Tv\| = \|v\|$, so (iii) holds.

(iii) \implies (ii) We just show that the form can be recovered from the norm. We have

$$2\Re\langle u, v \rangle = \|u+v\|^2 - \|u\|^2 - \|v\|^2, \quad \Im\langle v, w \rangle = \Re\langle v, iw \rangle.$$

For the second equality, notice that,

$$\begin{aligned} 2\Re \langle v, iw \rangle &= \langle v, iw \rangle + \overline{\langle v, iw \rangle} = -i \langle v, w \rangle + i \overline{\langle v, w \rangle} = \\ &= -i(\langle v, w \rangle - \overline{\langle v, w \rangle}) = \frac{1}{i}(\langle v, w \rangle - \overline{\langle v, w \rangle}) = 2\Im \langle v, w \rangle \end{aligned}$$

Now suppose $\|Tv\| = \|v\|$ for all v . Take $u, v \in V$. We have $\|T(u+v)\| = \|u+v\|$, $\|T(u)\| = \|u\|$, and $\|T(v)\| = \|v\|$. It follows that

$$2\Re \langle T(u), T(v) \rangle = \|T(u)+T(v)\|^2 - \|T(u)\|^2 - \|T(v)\|^2 = \|u+v\|^2 - \|u\|^2 - \|v\|^2 = 2\Re \langle u, v \rangle$$

and the second inequality shows that

$$\Im \langle T(u), T(v) \rangle = \Re \langle T(u), iT(v) \rangle = \Re(\langle T(u), T(iv) \rangle) = \Re(\langle u, iv \rangle) = \Im \langle u, v \rangle$$

. Hence

$$\langle T(u), T(v) \rangle = \langle u, v \rangle$$

.

(ii) implies (i):

$$\begin{aligned} \langle T^*Tu, v \rangle &= \langle Tu, Tv \rangle \\ &= \langle u, v \rangle. \end{aligned}$$

Therefore $\langle (TT^* - I)u, v \rangle = 0$ for all v . In particular, take $v = (TT^* - I)u$, then $\langle (TT^* - I)u, (TT^* - I)u \rangle = 0$. Therefore $TT^* = I$. \square

Notice that in an orthonormal basis (with respect to the standard inner product), an isometry is represented by a matrix such that $\overline{A}^t = A^{-1}$.

We let $O_n(\mathbb{R})$ be the set of $n \times n$ real matrices satisfying $AA^t = I_n$ (in other words $A^t = A^{-1}$). If $A \in O_n(\mathbb{R})$ then $\det A = \pm 1$. If $A \in O_n(\mathbb{R})$ then $A^t = A^{-1}$ so

$$\det A = \det A^t = \det(A^{-1}) = \det A^{-1}.$$

Therefore $\det(A)^2 = 1$ and $\det A = \pm 1$.

Theorem 4.2 *The following are equivalent.*

(i) $A \in O_n(\mathbb{R})$.

(ii) The columns of A form an orthonormal basis for \mathbb{R}^n (for the standard inner product on \mathbb{R}^n).

(iii) The rows of A form an orthonormal basis for \mathbb{R}^n .

Proof. We prove (i) \iff (ii) (the proof of (i) \iff (iii) is identical).

Consider $A^t A$. If $A = [C_1, \dots, C_n]$, so the j th column of A is C_j , then the (i, j) th entry of $A^t A$ is $C_i^t C_j$.

So $A^t A = I_n \iff C_i^t C_j = \delta_{i,j} \iff \langle C_i, C_j \rangle = \delta_{i,j} \iff \{C_1, \dots, C_n\}$ is an orthonormal basis for \mathbb{R}^n . \square

For example take the matrix:

$$\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

This matrix is in $O_2(\mathbb{R})$.

In fact it is the matrix of rotation by angle $-\pi/4$.

Theorem 4.3 Let V be a real inner product space with orthonormal basis $\mathcal{E} = \{e_1, \dots, e_n\}$. If $\mathcal{F} = \{f_1, \dots, f_n\}$ is a basis for V and P is the transition matrix from \mathcal{E} to \mathcal{F} , then

$$P \in O_n(\mathbb{R}) \iff \mathcal{F} \text{ is an orthonormal basis for } V.$$

Proof. The j th column of P is $[f_j]_{\mathcal{E}}$ so

$$f_j = \sum_{k=1}^n p_{k,j} e_k.$$

Hence

$$\langle f_i, f_j \rangle = \left\langle \sum_{k=1}^n p_{k,i} e_k, \sum_{l=1}^n p_{l,j} e_l \right\rangle = \sum_{k=1}^n \sum_{l=1}^n p_{k,i} p_{l,j} \langle e_k, e_l \rangle = \sum_{k=1}^n p_{k,i} p_{k,j} = (P^t P)_{i,j}.$$

So \mathcal{F} is an orthonormal basis for $\mathbb{R}^n \iff \langle f_i, f_j \rangle = \delta_{i,j}$ iff $P^t P = I_n \iff P \in O_n(\mathbb{R})$. \square

Notice that it is NOT true that matrices in $O_n(\mathbb{R})$ are diagonalisable.

Indeed, take

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

where θ is not a multiple of π .

The characteristic polynomial is $x^2 - 2\cos(\theta)x + 1$. Then, as $\cos(\theta)^2 - 1 < 0$, there are no real eigenvalues and the matrix is not diagonalisable.

Notice that for a given matrix A , it is easy to check that columns are orthogonal. If that is the case, then A is in $O_n(\mathbb{R})$ and it is easy to calculate inverse : $A^{-1} = A^t$.

5 Orthogonal Diagonalisation.

Definition 5.1 Let (V, \langle, \rangle) be an inner space. A linear map $T: V \rightarrow V$ is self-adjoint if

$$T^* = T$$

Notice that in an orthonormal basis, T is represented by a matrix A such that $\overline{A}^t = A$. In particular if V is real, then A is symmetric.

Theorem 5.1 If $A \in M_n(\mathbb{C})$ is Hermitian then all the eigenvalues of A are real.

Proof. Recall that Hermitian means that $A = \overline{A}^t$ and that this implies that $\langle Au, v \rangle = \langle u, Av \rangle$ for all u, v . Let λ be an eigenvalue of A and let $v \neq 0$ be a corresponding eigenvector. Then

$$Av = \lambda v$$

It follows that

$$\langle Av, v \rangle = \lambda \langle v, v \rangle = \langle v, Av \rangle = \overline{\lambda} \langle v, v \rangle$$

As $v \neq 0$, we can divide by $\langle v, v \rangle = \|v\|^2 \neq 0$ hence we can divide by it. It follows that $\lambda = \overline{\lambda}$ □

In particular a real symmetric matrix always has an eigenvalue : take a complex eigenvalue (always exists !), then by the above theorem it will be real.

Theorem 5.2 (Spectral theorem) *Let $T : V \rightarrow V$ be a self-adjoint linear map of an inner product space V . Then V has an orthonormal basis of eigenvectors.*

Proof. This is rather similar to Theorem 5.4.

We use induction on $\dim(V) = n$. True for $n = 1$ so suppose the result holds $n - 1$ and let $\dim(V) = n$.

Since T is self-adjoint, if \mathcal{E} is an orthonormal basis for V and A is the matrix representing T in \mathcal{E} then

$$A = \overline{A}^t.$$

So A is Hermitian. Hence by Theorem 6.18 A has a real eigenvalue λ .

So there is a vector $e_1 \in V \setminus \{0\}$ such that $Te_1 = \lambda e_1$. Normalizing (dividing by $\|e_1\|$) we can assume that $\|e_1\| = 1$.

Let $W = \text{Span}\{e_1\}$ then by Theorem 6.9 we have $V = W \oplus W^\perp$. Now

$$n = \dim(V) = \dim(W) + \dim(W^\perp) = 1 + \dim(W^\perp),$$

so $\dim(W^\perp) = n - 1$.

We claim that $T : W^\perp \rightarrow W^\perp$, i.e. $T(W^\perp) \subseteq W^\perp$. Let $w = \mu e_1 \in W$, $\mu \in \mathbb{R}$ and $v \in W^\perp$. Then

$$\langle w, Tv \rangle = \langle T^*w, v \rangle = \langle Tw, v \rangle = \langle T(\mu e_1), v \rangle = \langle \mu Te_1, v \rangle = \langle \mu \lambda e_1, v \rangle = 0,$$

since $\mu \lambda e_1 \in W$. Hence $T : W^\perp \rightarrow W^\perp$.

By induction there exists an orthonormal basis of eigenvectors $\{e_2, \dots, e_n\}$ for W^\perp . But $V = W \oplus W^\perp$ so $\mathcal{E} = \{e_1, \dots, e_n\}$ is a basis for V and $\langle e_1, e_i \rangle = 0$ for $2 \leq i \leq n$ and $\|e_1\| = 1$. Hence \mathcal{E} is an orthonormal basis of eigenvectors for V . \square

Theorem 5.3 *Let $T : V \rightarrow V$ be a self-adjoint linear map of a Euclidean space V . If λ, μ are distinct eigenvalues of T then*

$$\forall u \in V_\lambda \quad \forall v \in V_\mu \quad \langle u, v \rangle = 0.$$

Proof. If $u \in V_\lambda$ and $v \in V_\mu$ then

$$\lambda \langle u, v \rangle = \langle \lambda u, v \rangle = \langle Tu, v \rangle = \langle u, T^*v \rangle = \langle u, Tv \rangle = \langle u, \mu v \rangle = \mu \langle u, v \rangle.$$

So $(\lambda - \mu) \langle u, v \rangle = 0$, with $\lambda \neq \mu$. Hence $\langle u, v \rangle = 0$. \square

Example 5.4 Let

$$A = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$$

This matrix is self-adjoint.

One calculates the characteristic polynomial and finds $t(t-2)$ (in particular the minimal polynomial is the same, hence you know that the matrix is diagonalisable for other reasons than being self-adjoint). For eigenvalue zero, one finds eigenvector

$$\begin{pmatrix} -i \\ 1 \end{pmatrix}$$

For eigenvalue 2, one finds $\begin{pmatrix} i \\ 1 \end{pmatrix}$. Then we normalise the vectors : $v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 1 \end{pmatrix}$ and $v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}$. We let

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}$$

and

$$P^{-1}AP = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

In general the procedure for orthogonal orthonormalisation is as follows.

Let A be an $n \times n$ self-adjoint matrix.

Find eigenvalues λ_i and eigenspaces $V_1(\lambda_i)$. Because it is diagonalisable, you will have:

$$V = V_1(\lambda_1) \oplus \cdots \oplus V_r(\lambda_r)$$

Choose a basis for V as union of bases of $V_1(\lambda_i)$. Apply Gram-Schmidt to it to get an orthonormal basis.

For example :

$$A = \begin{pmatrix} 1 & -2 & 2 \\ -2 & 4 & -4 \\ 2 & -4 & 4 \end{pmatrix}$$

This matrix is symmetric hence self-adjoint.

One calculates the characteristic polynomial and finds $\lambda^2(\lambda - 9)$.

For $V_1(9)$, one finds $v_1 = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$ To make this orthonormal, divide by the norm, i.e replace v_1 by $\frac{1}{3}v_1$.

For $V_1(0)$, one finds $V_1(0) = \text{Span}(v_2, v_3)$ with

$$v_3 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

and

$$v_4 = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

By Gram-Schmidt process we replace v_3 by

$$\frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

and v_4 by

$$\frac{1}{3\sqrt{5}} \begin{pmatrix} -2 \\ 4 \\ 5 \end{pmatrix}$$

Let

$$P = \begin{pmatrix} 1/3 & 2/\sqrt{5} & -2/3\sqrt{5} \\ -2/3 & 1/\sqrt{5} & 4/3\sqrt{5} \\ 2/3 & 0 & 5/3\sqrt{5} \end{pmatrix}$$

We have

$$P^{-1}AP = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$