The minimal polynomial

Michael H. Mertens

October 22, 2015

Introduction

In these short notes we explain some of the important features of the minimal polynomial of a square matrix A and recall some basic techniques to find roots of polynomials of small degree which may be useful.

Should there be any typos or mathematical errors in this manuscript, I'd be glad to hear about them via email (`michael.mertens@emory.edu`) and correct them. Please note that the numbering of theorems and definitions is unfortunately not consistent with the lecture.

Atlanta, October 2015, Michael H. Mertens

1 The minimal polynomial and the characteristic polynomial

1.1 Definition and first properties

Throughout this section, $A \in \mathbb{R}^{n\times n}$ denotes a square matrix with n columns and real entries, if not specified otherwise. We want to study a certain polynomial associated to A, the *minimal polynomial*.

Definition 1.1. *Let $p(X) = c_d X^d + c_{d-1}X^{d-1} + ... + c_1 X + c_0$ be a polynomial of degree d in the undeterminate X with real coefficients.*

(i) For $A \in \mathbb{R}^{n\times n}$ we define

$$p(A) := c_d A^d + c_{d-1}A^{d-1} + ... + c_1 A + c_0 I_n.$$

(ii) The **minimal polynomial** *of A, denoted by $\mu_A(X)$, is the monic (i.e. with leading coefficient 1) polynomial of lowest degree such that*

$$\mu_A(A) = 0 \in \mathbb{R}^{n\times n}.$$

It is maybe not immediately clear that this definition always makes sense.

Lemma 1.2. *The minimal polynomial is always well-defined and we have* $\deg \mu_A(X) \leq n^2$.

Proof. We can write the entries of an $n \times n$-matrix as a column, e.g. reading the matrix row-wise,

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \mapsto \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1n} \\ a_{21} \\ \vdots \\ a_{nn} \end{pmatrix}.$$

With this we can identify $\mathbb{R}^{n\times n}$ with $\mathbb{R}^{n^2}$ and, as is easy to see, addition and scalar multipication of matrices are respected by this map. In particular this means that at most n^2 matrices in $\mathbb{R}^{n\times n}$ can be linearly independent, since they can be thought of as living in $\mathbb{R}^{n^2}$, whose dimension is n^2. This means now that there must be a minimal $d \leq n^2$ such that the matrices

$$I_n, A, A^2, ..., A^d$$

(viewed as vectors as described above) are linearly dependent, i.e. there are numbers $c_0, ..., c_{d-1}$ such that

$$A^d + c_{d-1}A^{d-1} + ... + c_1 A + c_0 I_n = 0 \in \mathbb{R}^{n\times n}$$

If we now replace A in this equation by the undeterminate X, we obtain a monic polynomial $p(X)$ satisfying $p(A) = 0$ and the degree d of p is minimal by construction, hence $p(X) = \mu_A(X)$ by definition. $\square$

Remark 1.3. *In fact, there is a much stronger (sharp) bound on the degree of the minimal polynomial, namely we have that* $\deg \mu_A(X) \leq n$. *This is a consequence of the Theorem of Cayley-Hamilton (Theorem 1.11), which we will encounter later.*

Recall the definition of eigenvalues, eigenvectors, and eigenspaces from the lecture.

Definition 1.4. *A non-zero vector $v \in \mathbb{R}^n$ that satisfies the equation*

$$Av = \lambda v$$

for some scalar λ is called an **eigenvector** *of A to the* **eigenvalue** *λ. The set of all vectors satisfying the above equation is called the* **eigenspace** *of A associated to λ, denoted by $\mathcal{E}_A(\lambda)$.*

Remark 1.5. *Recall that $\mathcal{E}_A(\lambda) = \mathrm{Kern}(A - \lambda I_n)$, so in particular, $\mathcal{E}_A(\lambda)$ is a subspace of $\mathbb{R}^n$.*

One of our goals is to find the eigenvalues of A. The following theorem will tell us how we can use the minimal polynomial of A to find them.

Theorem 1.6. *The number λ is an eigenvalue of A if and only if $\mu_A(\lambda) = 0$, which is the case if and only if we can write $\mu_A(X) = (X - \lambda)q(X)$ for some polynomial $q(X)$ of degree $\deg \mu_A(X) - 1$.*

Proof. First suppose that λ is an eigenvalue of A, hence $\mathcal{E}_A(\lambda)$ is not just the zero space. By definition, the maps $v \mapsto Av$ and $v \mapsto \lambda v$ are the same on the subspace $\mathcal{E}_A(\lambda)$ by definition, so (restricted to this subspace!) the minimal polynomial of this map is $X - \lambda$. But since we know that $\mu_A(A) = 0$, this implies that $(X - \lambda)$ divides $\mu_A(\lambda)$.

Now suppose that $\mu_A(\lambda) = 0$, i.e. we can write $\mu_A(X) = (X - \lambda)q(X)$ for some polynomial $q(X)$ of degree $\deg \mu_A(X) - 1$. If then λ is not an eigenvalue of A, i.e. $\mathcal{E}_A(\lambda) = \{0\}$, so in particular the matrix $A - \lambda I_n$ is invertible. In particular this means that $\mu_A(A) = 0$ if and only if $q(A) = 0$, but this can't be because the degree of $q(X)$ is less than the degree of $\mu_A(X)$. Thus λ must be an eigenvalue of A. $\square$

So once one has the minimal polynomial, one only has to find its zeros in order to find the eigenvalues of A. So it remains to show how to compute the minimal polynomial. Theoretically, one could use the method outlined in the proof of Lemma 1.2, i.e. successively compute powers of A and look for linear dependencies. The problem here is that computing powers of large matrices becomes very tedious and in each step one has to solve a linear system with n^2 equations to find the linear dependencies. A much more efficient method is outlined below.

Algorithm 1.7.

INPUT:	$A \in \mathbb{R}^{n\times n}$
OUTPUT:	$\mu_A(X)$
ALGORITHM:	

(i) *Choose any non-zero* $v \in \mathbb{R}^n$.

(ii) *While the set* $S = \{v, Av, A^2v, ..., A^m v\}$ *is linearly independent, compute* $A^{m+1}v = A(A^m v)$ *and add it to* S.

(iii) *Write down the (normalized) linear depency for* S, *i.e. compute numbers* $c_0, ..., c_{d-1}$ *such that* $A^d v = c_{d-1}A^{d-1}v + ... + c_0 v$ *and use these to define the polynomial* $\tilde{\mu}(X) = X^d - (c_{d-1}X^{d-1} + ... + c_0)$.

(iv) *While the vectors found don't span* $\mathbb{R}^n$, *find a vector not in their span and repeat steps 2. and 3.*

(v) *The minimal polynomial is then the least common multiple of all the polynomials found on the way.*

Example 1.8. *Let*

$$A = \begin{pmatrix} 4 & 0 & -3 \\ 4 & -2 & -2 \\ 4 & 0 & -4 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

Find $\mu_A(X)$.

We first choose the vector $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. *We compute successively*

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

$$Ae_1 = \begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix},$$

$$A^2 e_1 = A \begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix} = 4e_1.$$

Hence we have found a linear dependency $(A^2 - 4I_3)e_1 = 0$*, we have found a divisor of our minimal polynomial,*

$$\tilde{\mu}(X) = X^2 - 4.$$

Since e_1 *and* Ae_1 *do not yet span all of* $\mathbb{R}^3$*, we need to repeat the above steps with a vector not in the span of those, e.g. with* $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$*. Then we have*

$$Ae_2 = \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} = -2e_2,$$

so that we find the linear dependence $Ae_2 + 2e_2 = 0$*, i.e. the divisor* $X + 2$ *of our minimal polynomial. Note that by chance,* e_2 *is indeed an eigenvector of* A*, but this is merely a coincidence. Since the three vectors* e_1, Ae_1, e_2 *span* $\mathbb{R}^3$*, we can stop now and find*

$$\mu_A(X) = \operatorname{lcm}(X^2 - 4, X + 2) = X^2 - 4.$$

Another way to compute eigenvalues of a matrix is through the characteristic polynomial.

Definition 1.9. *For* $A \in \mathbb{R}^{n \times n}$ *we define the* **characteristic polynomial** *of* A *as*

$$\chi_A(X) := \det(XI_n - A).$$

This is a monic polynomial of degree n*.*

The motivation for this definition essentially comes from the invertible matrix theorem, especially Theorem 3.8 of the lecture. More precisely we have that λ is an eigenvalue of A if and only if $\mathcal{E}_A(\lambda) = \operatorname{Kern}(A - \lambda I_n) \neq \{0\}$ which is the case if and only if the matrix $A - \lambda I_n$ is *not* invertible, which happens if and only if $\det(A - \lambda I_n) = 0$. This proves the following theorem.

Theorem 1.10. *The number* λ *is an eigenvalue of the matrix* $A \in \mathbb{R}^{n \times n}$ *if and only if* $\chi_A(\lambda) = 0$*.*

In particular, the minimal polynomial $\mu_A(X)$ and the characteristic polynomial $\chi_A(\lambda)$ have the same roots. In fact even more is true.

Theorem 1.11 (Cayley-Hamilton)**.** *The minimal polynomial divides the characteristic polynomial, or in other words, we have*

$$\chi_A(A) = 0 \in \mathbb{R}^{n\times n}.$$

Proof. By extending the definition of the classical adjoint to matrices with polynomials as entries we can write

$$(XI_n - A)\,\mathrm{adj}(XI_n - A) = \det(XI_n - A)I_n = \chi_A(X)I_n. \tag{1.1}$$

Now by construction the entries of the adjoint are polynomials of degree at most $n-1$, hence we can write

$$\mathrm{adj}(XI_n - A) = C_{n-1}X^{n-1} + C_{n-2}X^{n-2} + ... + C_1X + C_0,$$

where $C_0, ..., C_{n-1} \in \mathbb{R}^{n\times n}$ are suitable matrices. If we now let

$$\chi_A(X) = X^n + a_{n-1}X^{n-1} + ... + a_1X + a_0$$

for some numbers $a_0, ..., a_{n-1}$, it follows from (1.1) that

$$\begin{aligned}
&(X^n + a_{n-1}X^{n-1} + ... + a_1X + a_0)I_n\\
=&(XI_n - A)(C_{n-1}X^{n-1} + C_{n-2}X^{n-2} + ... + C_1X + C_0)\\
=&C_{n-1}X^n + (C_{n-2} - AC_{n-1})X^{n-1} + ... + (C_0 - AC_1)X - AC_0,
\end{aligned}$$

so that by comparing coefficients we see that

$$C_{n-1} = I_n, \quad a_jI_n = C_{j-1} - AC_j \ (j = 1, ..., n-1), \quad a_0I_n = -AC_0.$$

With this we obtain that

$$\begin{aligned}
\chi_A(A) &= A^n + a_{n-1}A^{n-1} + ... + a_1A + a_0I_n\\
&= A^n + A^{n-1}(C_{n-2} - A) + A^{n-2}(C_{n-3} - AC_{n-2}) + ... + (A(C_0 - AC_1) - AC_0\\
&= A^n - A^n + AC_0 - AC_0\\
&= 0
\end{aligned}$$

which is what we wanted to show. □

Remark 1.12. *It is a popular joke to "prove" the theorem of Cayley-Hamilton by the following short line of equations,*

$$\chi_A(A) = \det(AI_n - A) = \det(0) = 0.$$

Why is this "proof" rubbish?

If λ is an eigenvalue of A, then, according to Theorems 1.6, 1.10 and 1.11 we can write

$$\mu_A(X) = (X - \lambda)^\gamma p(X) \tag{1.2}$$

and

$$\chi_A(X) = (X - \lambda)^\alpha q(X), \tag{1.3}$$

where $p(\lambda), q(\lambda) \neq 0$. This yields

Definition 1.13. *For an eigenvalue λ of a matrix $A \in \mathbb{R}^{n\times n}$ we call the number γ defined by* (1.2) *the* **geometric multiplicity** *of λ. The number α defined by* (1.3) *is called its* **algebraic multiplicity***.*

1.2 Similarity and diagonalizability

A very important concept in linear algebra is that of similarity of matrices. This can be seen from the viewpoint of linear maps (later) or from a purely computational viewpoint: Suppose we want to compute the matrix A to a large power. Just by applying the definition of the matrix product, this is not really the best way to go. Suppose for example that we can write

$$A = gDg^{-1}$$

for some invertible matrix $g \in \mathbb{R}^{n\times n}$ and a diagonal matrix D. Then we have for example

$$A^2 = (gDg^{-1})^2 = gDg^{-1}gDg^{-1} = gD^2g^{-1},$$

and similarly in general

$$A^k = gD^kg^{-1}.$$

The good thing about this is of course that it is very easy indeed to compute a power of a diagonal matrix. We want to study this phenomenon a bit more closely.

Definition 1.14. *(i) We call a matrix $A \in \mathbb{R}^{n\times n}$* **similar** *to a matrix $B \in \mathbb{R}^{n\times n}$, in symbols $A \sim B$, if there exists an invertible matrix $g \in \mathbb{R}^{n\times n}$ such that $A = gBg^{-1}$.*

(ii) If A is similar to a diagonal matrix, we call A **diagonalizable***.*

Remark 1.15. *(i) If $A \sim B$, then we also have $B \sim A$, since*

$$A = gBg^{-1} \quad \Leftrightarrow \quad B = g^{-1}Ag.$$

(ii) *If $A \sim B$ and $B \sim C$ then we also have $A \sim C$, since it follows from $A = gBg^{-1}$ and $B = hCh^{-1}$ that $A = ghCh^{-1}g^{-1} = (gh)C(gh)^{-1}$.*

In general it is quite hard to decide whether two given matrices are similar or not. But there are several more or less easy to compute data that may give some indication. We give some of them in the following theorem.

Theorem 1.16. *Let $A, B \in \mathbb{R}^{n\times n}$ be similar. Then the following are all true.*

(i) $\det A = \det B$,

(ii) $\operatorname{trace} A = \operatorname{trace} B$,

(iii) $\chi_A(X) = \chi_B(X)$,

(iv) $\mu_A(X) = \mu_B(X)$,

(v) *A and B have the same eigenvalues with the same algebraic and geometric multiplicities.*

Proof. Since $A \sim B$, we can write $A = gBg^{-1}$ for some invertible matrix g.

(i) We have

$$\det A = \det(gBg^{-1}) = \det g \det B(\det g^{-}1) = \det B,$$

since the determinant is multiplicative.

(ii) Exercise 6 on Homework assignment 8.

(iii) We have

$$\begin{aligned}\chi_A(X) &= \det(XI_n - A) = \det(Xgg^{-1} - gBg^{-1}) = \det(g(XI_n - B)g^{-1}) \\ &= \det g \det(XI_n - B)(\det g^{-}1) = \det(XI_n - B) = \chi_B(X).\end{aligned}$$

(iv) Suppose that

$$\mu_A(X) = X^d + a_{d-1}X^{d-1} + ... + a_1X + a_0 \quad \text{and} \quad \mu_B(X) = X^{d'} + b_{d'-1}X^{d'-1} + ... + b_1X + b_0.$$

We show that $\mu_A(B) = \mu_B(A) = 0$, which implies that both minimal polynomials mutually divide each other, which means, since they have the same leading coefficient, that they must be equal. We compute

$$\begin{aligned}
\mu_A(B) &= B^d + a_{d-1}B^{d-1} + ... + a_1B + a_0I_n \\
&= (g^{-1}Ag)^d + a_{d-1}(g^{-1}Ag)^{d-1} + ... + a_1(g^{-1}Ag) + a_0g^{-1}g \\
&= (g^{-1}A^dg) + a_{d-1}(g^{-1}A^{d-1}g) + ... + a_1(g^{-1}Ag) + a_0g^{-1}g \\
&= g^{-1}(A^d + a_{d-1}A^{d-1} + ... + a_1A + a_0I_n)g \\
&= g^{-1}\mu_A(A)g \\
&= 0.
\end{aligned}$$

With a similar computation one also shows $\mu_B(A) = 0$ which completes the proof.

(v) Follows from (iii) and (iv).

□

Remark 1.17. *(i) Theorem 1.16 does* **NOT** *say that if the indicated quantities above agree for two matrices, then they are equal. It just states that if they do not agree for two given matrices, then they cannot be similar. It is not even so that two matrices are similar if* all *the above quantities agree. For example for*

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad \textit{and} \quad B = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

we have $\det A = \det B = 16$, $\operatorname{trace} A = \operatorname{trace} B = 8$, $\chi_A(X) = \chi_B(X) = (X-2)^4$, *and* $\mu_A(X) = \mu_B(X) = (X-2)^2$, *but the two matrices are not similar (which is not so easy to see).*

(ii) On the other hand one can show that two 2×2*-matrices are similar if and only if their minimal polynomials agree (this is false for the characteristic polynomial!).*

(iii) As one can also show, two 3×3*-matrices are similar if and only if they have the same minimal* and *characteristic polynomial.*

2 Roots of polynomials

2.1 Quadratic and biquadratic equations

Even though it is probably the most well-known topic among those discussed in these notes, we begin by recalling how to solve *quadratic equations*. Suppose we have an equation of the form

$$ax^2 + bx + c = 0$$

where a, b, c are given real numbers (with $a \neq 0$, otherwise we would in fact be talking about linear equations) and we want to solve for x. Since $a \neq 0$, we can divide the whole equation by it and add a clever zero on the left hand side, giving

$$x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2 + \frac{c}{a} = 0.$$

The first three summands can easily be recognized to equal $(x + \frac{b}{2a})^2$. This procedure of adding this particular zero is called *completing the square*. Now we reorder the equation to obtain

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}.$$

Now there are three cases to distinguish,

(*i*) If $\Delta := b^2 - 4ac > 0$, then we obtain two distinct real solutions by taking the square-root, namely

$$x = \frac{-b + \sqrt{\Delta}}{2a} \quad \text{or} \quad x = \frac{-b - \sqrt{\Delta}}{2a}. \tag{2.1}$$

Note that the square-root of a positive real number a is also positive by definition and therefore unique, while the equation $x^2 = a$ has two solutions, $\sqrt{a}$ and $-\sqrt{a}$.

(*ii*) If $\Delta = 0$, then there is precisely one zero,

$$x = -\frac{b}{2a}.$$

In this case we speak of a *double zero*, since the derivative of the function $f(x) = ax^2 + bx + c$ would also vanish in this case. The zeros in the first case are called *simple zeros*.

(*iii*) If $\Delta < 0$, then there is no real solution, since the square of a real number is always non-negative.

Because the behaviour of the quadratic equation is determined entirely by the quantity $\Delta = b^2 - 4ac$, it is called the *discriminant* of the equation (from Latin *discriminare* – to distinguish).

In the case where $a = 1$ (we say that the polynomial is *monic* in this case) and b and c are integers, there is an easy alternative to the above formula, which gives the solutions quicker if they are integers (and in many examples, they will be). It is based on the following observation which goes back to the French-Italian mathematician François Viète (1540–1603).

Lemma 2.1. *Let α_1 and α_2 be roots of the polynomial $x^2 + bx + c$. Then we have $b = -(\alpha_1 + \alpha_2)$ and $c = \alpha_1\alpha_2$.*

Proof. If α_1 and α_2 are the two zeros of our polynomial, then we must have

$$x^2 + bx + c = (x - \alpha_1)(x - \alpha_2) = x^2 - (\alpha_1 + \alpha_2) + \alpha_1\alpha_2.$$

Thus a comparison of the coefficients gives the lemma. □

So if one can factor the number c and combine the divisors so that they sum to $-b$, one also has found the solutions to the equation. This may be easier to do without a calculator than taking square-roots, especially if c has very few divisors. If this factoring method doesn't work, then we also know that our solutions will not be integers.

Sometimes it happens that one has to deal with so-called *biquadratic equations*. Those have the general form

$$ax^4 + bx^2 + c = 0.$$

It is not very complicated to solve these as well, one just substitutes $z = x^2$ to obtain a quadratic equation in z, which one can solve by either one of the above methods. Afterwards, we take the positive and negative square-root of the solutions which are non-negative (the others don't yield any real solutions to the biquadratic equation).

Example 2.2. *Let's solve the biquadratic equation*

$$x^4 + x^2 - 20 = 0.$$

Substituting $z = x^2$ gives us the quadratic equation

$$z^2 + z - 20 = 0.$$

The discriminant of this quadratic equation is $\Delta = 1^2 - 4 \cdot 1 \cdot (-20) = 81 > 0$, *therefore, we have two real solutions for* z, *according to our formula* (2.1), *namely*

$$z = \frac{-1 + \sqrt{81}}{2} = 4 \quad \textit{or} \quad z = \frac{-1 - \sqrt{81}}{2} = -5.$$

Since $z = x^2$, *it must be non-negative if* x *is a real number, so the solution* $z = -5$ *is irrelevant for us and we obtain the two real solutions*

$$x = 2 \quad \textit{or} \quad x = -2.$$

2.1.1 Basics on complex numbers

Let us go back to the third case about solving quadratic equations, when the discriminant of a quadratic equation is negative. The argument why there is no solution is that the square of a real number is always non-negative. But often it is necessary to have a solution to such an equation, even if it is not real. So one *imagines* that there is a "number" which we call i with the property $i^2 = -1$. This number i is **not** a real number, but it is indeed called *imaginary*. We state now that essentially all the rules of arithmetic one is used to from working with real numbers can also be used for this number i. With this, we can in fact solve our quadratic equation

$$ax^2 + bx + c = 0$$

even in the case when the discriminant Δ is negative, namely by writing $\Delta = (-1) \cdot (-\Delta)$, keeping in mind that $-\Delta$ is positive, we can write our formula (2.1) as

$$x = \frac{-b + i\sqrt{-\Delta}}{2a} \quad \text{or} \quad x = \frac{-b - i\sqrt{-\Delta}}{2a}, \tag{2.2}$$

an expression we can now make sense of. In general we call an object of the form $\alpha = a + bi$ with real numbers a and b a *complex number*. The number a is called the *real part* of α, the number b is called the *imaginary part* of α (in particular, the imaginary part of a complex number is always a real number). Every complex number can be simplified to be of this form. We collect a few facts about complex numbers here. We note that one can basically calculate with complex numbers exactly as with real numbers, but since it is not really relevant in the course of the lecture, we won't go into this here. The only thing that we will need is the *exponential* of a complex number. We will just give it as a definition, although it is possible to derive it properly.

Definition 2.3. *For a complex number $\alpha = a + bi$ with real numbers a, b we have*

$$\exp(\alpha) := e^{\alpha} := e^{a}\cos(b) + ie^{a}\sin(b).$$

2.2 Cubic equations

2.2.1 Polynomial division

There is an easy way to divide polynomials by one another which works basically like the long division algorithm for integers. This comes in handy if one has guessed one zero, say α, of a polynomial because one can then divide the polynomial by $x - \alpha$ to obtain a polynomial of lower degree one has to deal with (see Lemma 2.4). We want to divide $x^3 - x^2 - 3x - 9$ by $x - 3$. One only looks at the leading terms of the polynomials and divides those. In this case, we obtain x^2.

$$x^3 \; - x^2 - 3x - 9 = \left(x - 3\right)\left(x^2 \qquad\qquad\right).$$

Then we multiply the divisor $x - 3$ by this result and subtract it from the dividend $x^3 - x^2 - 3x - 9$,

$$\begin{array}{l}
x^3 \; - x^2 - 3x - 9 = \left(x - 3\right)\left(x^2 \qquad\qquad\right). \\
\underline{-\, x^3 + 3x^2}
\end{array}$$

We copy the next lower term downstairs and repeat the procedure with the difference.

$$\begin{array}{l}
x^3 \; - x^2 - 3x - 9 = \left(x - 3\right)\left(x^2 \qquad\qquad\right). \\
\underline{-\, x^3 + 3x^2} \\
\qquad\quad 2x^2 - 3x
\end{array}$$

Again, we only divide the leading terms, and we get $+2x$, which we write next to the x^2 from the previous step,

$$\begin{array}{l}
x^3 \; - x^2 - 3x - 9 = \left(x - 3\right)\left(x^2 + 2x \qquad\right). \\
\underline{-\, x^3 + 3x^2} \\
\qquad\quad 2x^2 - 3x
\end{array}$$

We multiply this $2x$ again by our divisor and subtract the result from the dividend,

$$\begin{array}{rl} x^3 \;\; - x^2 - 3x - 9 = \bigl(x-3\bigr)\bigl(x^2 + 2x \quad\;\bigr)\,. \\ \underline{- x^3 + 3x^2 \quad\quad} \\ 2x^2 - 3x \\ \underline{- 2x^2 + 6x} \end{array}$$

We repeat this procedure again and finsh up,

$$\begin{array}{rl} x^3 \;\; - x^2 - 3x - 9 = \bigl(x-3\bigr)\bigl(x^2 + 2x + 3\bigr). \\ \underline{- x^3 + 3x^2 \quad\quad} \\ 2x^2 - 3x \\ \underline{- 2x^2 + 6x} \\ 3x - 9 \\ \underline{- 3x + 9} \\ 0 \end{array}$$

Hence we have found our result of the division, namely $x^2 + 2x + 3$.

In general, it is not necessary that the polynomial division goes through without a remainder. If there is one, it will have a degree less than the divisor. We want to divide $x^4 - 3x^3 + 2x^2 - 5x + 7$ by $x^2 - x + 1$. Up to the last step everything is the same as before,

$$\begin{array}{rl} x^4 - 3x^3 + 2x^2 - 5x + 7 = \bigl(x^2 - x + 1\bigr)\bigl(x^2 - 2x - 1\bigr) \qquad\qquad . \\ \underline{- x^4 \;\; + x^3 \;\; - x^2} \\ - 2x^3 \;\; + x^2 - 5x \\ \underline{2x^3 - 2x^2 + 2x} \\ - x^2 - 3x + 7 \\ \underline{x^2 \;\; - x + 1} \\ - 4x + 8 \end{array}$$

We see that the last difference is a polynomial of degree 1, but not 0. This polynomial is our remainder, which we have to add on the right-hand side,

$$
\begin{array}{rrrrrl}
x^4 & -3x^3 & +2x^2 & -5x & +7 & = \left(x^2-x+1\right)\left(x^2-2x-1\right) -4x+8. \\
-x^4 & +x^3 & -x^2 & & & \\
\cline{1-3}
 & -2x^3 & +x^2 & -5x & & \\
 & 2x^3 & -2x^2 & +2x & & \\
\cline{2-4}
 & & -x^2 & -3x & +7 & \\
 & & x^2 & -x & +1 & \\
\cline{3-5}
 & & & -4x & +8 &
\end{array}
$$

2.2.2 Cubic and higher degree polynomials

As indicated in the last section, we can use polynomial division to find roots of polynomials. This is based on two easy facts which we recall in the following lemma.

Lemma 2.4. *(i) The product of two numbers is* 0 *if and only one of the numbers is zero,*

$$a \cdot b = 0 \quad \Rightarrow \quad a = 0 \quad \textit{or} \quad b = 0.$$

(ii) A number α is the root of a polynomial $p(x)$ if and only if $x - \alpha$ divides $p(x)$, i.e. the polynomial division of $p(x)$ by $x - \alpha$ goes through without a remainder.

So if we want to find the roots or zeros of a cubic polynomial, we can somehow guess one zero, and use polynomial division to obtain a quadratic polynomial, whose zeros we can determine using the formulas in Section 2.1. The question now is how we can guess a zero of a polynomial. At least in case of a monic polynomial with integer coefficients, there is a very easy observation with which we can check at least for integer zeros. It is a generalization of the Theorem of Vieta (Lemma 2.1)

Proposition 2.5. *An integer a is a zero of a monic polynomial with integer coefficients if and only if it divides the absolute term of the polynomial. If no divisor of the absolute term is a zero of the polynomial, then all zeros are irrational.*

So if we cannot find an integer root, then we have essetially no chance of "guessing" one. In this case, there are numerical methods to obtain approximations for zeros, the most well-known goes back to Sir Isaac Newton (1642–1727), or (for polynomials of degree 3 and 4) there are even closed formulas. None of these will be relevant in our course.

Example 2.6. *We want to determine all zeros of the cubic polynomial*

$$x^3 + 4x^2 - 7x - 10.$$

By Proposition 2.5, we have to check the divisors of the absolute term, which is 10 *in this case. By trial and error we find that that* 2 *is in fact a zero of this polynomial. Now we use polynomial division,*

$$\begin{array}{rrrrl}
x^3 & +\,4x^2 & -\,7x & -\,10 & = \big(x-2\big)\big(x^2+6x+5\big). \\
-\,x^3 & +\,2x^2 & & & \\ \hline
 & 6x^2 & -\,7x & & \\
 & -\,6x^2 & +\,12x & & \\ \cline{2-3}
 & & 5x & -\,10 & \\
 & & -\,5x & +\,10 & \\ \cline{3-4}
 & & & 0 &
\end{array}$$

The result is $x^2 + 6x + 5$. *The zeros of this polynomial are* -1 *and* -5 *which one can check either using* (2.1) *or Lemma 2.1. Thus we have the three zeros* $x = 2$, $x = -1$, *and* $x = -5$.

2.3 Exercises

Carry out polynomial division for the following pairs of polynomials.

(*i*) $x^2 - 4x + 3$ divided by $x - 3$,

(*ii*) $x^3 - 4x^2 + 3x - 12$ divided by $x - 4$,

(*iii*) $x^3 - 5x^2 + 3x - 4$ divided by $x^2 + 1$,

(*iv*) $x^4 + 7x^2 - 4x + 2$ divided by $x^2 + 3x - 1$.

Find all zeros (real and complex) of the following polynomials.

(*i*) $x^2 - 4x + 4$,

(*ii*) $x^2 - 4x + 13$,

(*iii*) $x^4 - 25x^2 + 144$,

(*iv*) $x^3 - 2x^2 - 5x + 6$,

(*v*) $x^3 - 2x^2 + 9x - 18$.