Chapter 3

Linear systems

3.1 Motivation for the matrix exponent

First I will study linear systems with constant coefficients of the form

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \quad \boldsymbol{x}(t) \in \mathbf{R}^k, \quad \boldsymbol{A} \in M_k(\mathbf{R}), \tag{3.1}$$

where the notation $M_k(\mathbf{R})$ means the real vector space of square real $k \times k$ matrices. Additionally to (3.1) I also consider the initial value problem for (3.1) with the initial condition

$$\boldsymbol{x}(0) = \boldsymbol{x}_0 \in \mathbf{R}^k, \tag{3.2}$$

where the initial time moment can be taken to be zero without loss of generality since the system is autonomous.

I know, from the previous section, that problem (3.1), (3.2) can be replaced with the integral equation

$$\boldsymbol{x}(t) = \boldsymbol{x}_0 + \int_0^t \boldsymbol{A}\boldsymbol{x}(\tau)\,\mathrm{d}\tau, \tag{3.3}$$

which can be used to produce Picard's iterates

$$\begin{aligned}
\boldsymbol{x}_1(t) &= \boldsymbol{x}_0 + \int_0^t \boldsymbol{A}\boldsymbol{x}_0(\tau)\,\mathrm{d}\tau = (\boldsymbol{I} + t\boldsymbol{A})\,\boldsymbol{x}_0,\\
\boldsymbol{x}_2(t) &= \boldsymbol{x}_0 + \int_0^t \boldsymbol{A}\boldsymbol{x}_1(\tau)\,\mathrm{d}\tau = \left(\boldsymbol{I} + t\boldsymbol{A} + \frac{t^2\boldsymbol{A}^2}{2}\right)\boldsymbol{x}_0,\\
&\dots\\
\boldsymbol{x}_n(t) &= \boldsymbol{x}_0 + \int_0^t \boldsymbol{A}\boldsymbol{x}_{n-1}(\tau)\,\mathrm{d}\tau = \left(\boldsymbol{I} + t\boldsymbol{A} + \ldots + \frac{t^n\boldsymbol{A}^n}{n!}\right)\boldsymbol{x}_0.
\end{aligned}$$

Assuming that I can continue the process indefinitely, I write that the solution to (3.1), (3.2) is given by

$$\boldsymbol{x}(t) = \left(\boldsymbol{I} + t\boldsymbol{A} + \ldots + \frac{t^n\boldsymbol{A}^n}{n!} + \ldots\right)\boldsymbol{x}_0 = \exp(t\boldsymbol{A})\boldsymbol{x}_0,$$

where an almost obvious notation $\exp(t\boldsymbol{A}) = e^{t\boldsymbol{A}}$ is used for the infinite series of matrix functions. Here is a formal definition.

Definition 3.1. *For the matrix $\boldsymbol{A}$ its matrix exponent is the series*

$$e^{\boldsymbol{A}} = \boldsymbol{I} + \boldsymbol{A} + \ldots + \frac{\boldsymbol{A}^n}{n!} + \ldots.$$

Using the last definition, I formally ("formally" here means that I did not prove yet that the corresponding series converges) can write that the solution to (3.1), (3.2) is given by (I still need to prove that this function gives a solutions to my problem)

$$\boldsymbol{x}(t) = e^{t\boldsymbol{A}}\boldsymbol{x}_0. \tag{3.4}$$

Before proceeding with the analysis of the solution (3.4), I need to make sure that the definition for the matrix exponent makes sense. This will be the goal of the next section. After this I will turn to the question how to actually calculate the matrix exponent given an arbitrary matrix $\boldsymbol{A}$, and, more importantly, which corollaries I can obtain if I know, at least in principle, the entries of the matrix exponent.

3.2 Series and linear operators in normed vector spaces

3.2.1 Series

In the vector space X the operation of addition is determined, therefore I can talk about series in the form

$$x_0 + x_1 + \ldots + x_n + \ldots, \quad x_i \in X.$$

For this series, exactly as in the case of the numerical series, I can form the partial sums

$$s_n = x_0 + \ldots + x_n,$$

and the series is called *convergent* if the sequence (s_n) converges in X, i.e., there exists an $s \in X$ such that $\|s_n - s\| \to 0$.

I assume for the following that I deal with a Banach space (so that I can deal with the fundamental sequences and do not have to know the limit of the partial sums). Then it can be proved (almost trivially) that an infinite series $\sum_{i=0}^{\infty} x_i$ converges if and only if for any $\epsilon > 0$ there exists $N(\epsilon)$ such that

$$\left\|\sum_{i=m}^{n} x_i\right\| \leq \epsilon$$

whenever $n \geq m > N$. (This is just a restatement of the fact that the sequence of the partial sums is fundamental).

Now with the series $\sum_{i=0}^{\infty} x_i$ consider the series of real numbers $\sum_{i=0}^{\infty} \|x_i\|$. If the latter series converges than it is said that the former series *converges absolutely.*

Lemma 3.2. *Let X be a Banach space. If the series $\sum_{i=0}^{\infty} x_i$ converges absolutely then it is convergent.*

Proof.

$$\left\|\sum_{i=m}^{n} x_i\right\| \leq \sum_{i=m}^{n} \|x_i\|.$$

■

3.2.2 Linear operators

Consider a mapping of the normed vector space X into the normed vector space Y:

$$L\colon X \longrightarrow Y.$$

This mapping is called a *linear operator*, if

$$L(\alpha x_1 + x_2) = \alpha L x_1 + L x_2, \quad x_1, x_2 \in X.$$

A linear operator L is *bounded* if $\|Lx\| \leq M\|x\|$ for some real number $M \geq 0$ for all $x \in X$ (note that the norms in the last inequality are from two different spaces).

Exercise 3.1. Prove that linear operator $L\colon X \longrightarrow Y$ between two normed vector spaces is 1) continuous if and only if it is continuous at 0, and 2) it is continuous if and only if it is bounded. Recall that A is continuous at a point $x \in X$ if $x_n \to x$ in X implies $Lx_n \to Lx$ in Y and it is continuous if it is continuous at every point of its definition.

Exercise 3.2. Can you give an example of a linear discontinuous operator?

Linear bounded operators themselves form a vector space $\mathscr{L}(X, Y)$, if the addition and multiplication by scalars is understood pointwise. Therefore, it is natural to consider the smallest possible constant M in the definition of the bounded operator as a norm on $\mathscr{L}(X, Y)$.

Definition 3.3. *Let $L \in \mathscr{L}(X, Y)$, where X, Y are normed vector spaces. I define the (uniform) norm of L to be*

$$\|L\| = \inf\{M\colon \|Lx\| \leq M\|x\| \text{ for all } x \in X\}.$$

Exercise 3.3. Prove that the definitions of the norm of a linear continuous operator

$$\|L\| = \sup_{x\neq 0} \frac{\|Lx\|}{\|x\|} = \sup_{\|x\|=1} \|Lx\| = \sup_{\|x\|\leq 1} \|Lx\|$$

are equivalent.

Since I consider only bounded linear operators then the last definition makes perfect sense. It immediately implies that

$$\|Lx\| \leq \|L\|\|x\|.$$

Let me check the norm axioms. It is obviously nonnegative and equal to zero if and only if L is the zero operator. Furthermore,

$$\|\alpha L\| = \sup_{x\in X,\, x\neq 0} \frac{\|\alpha Lx\|}{\|x\|} = \alpha\|L\|.$$

To prove the triangle inequality, consider

$$\|(A + B)(x)\| \leq \|Ax\| + \|Bx\| \leq \|A\|\|x\| + \|B\|\|x\|,$$

therefore

$$\|A + B\| = \sup_{x\in X,\, x\neq 0} \frac{\|(A + B)(x)\|}{\|x\|} \leq \|A\| + \|B\|.$$

Moreover,

$$\|ABx\| \le \|A\|\|Bx\| \le \|A\|\|B\|\|x\|$$

implies

$$\|AB\| \le \|A\|\|B\|.$$

The last inequality can be used to see that

$$\|A^j\| \le \|A\|^j.$$

The sequence (A_n) of bounded linear operators in $\mathscr{L}(X,Y)$ is said to converge *uniformly* (or in the *operator norm topology*) to A if $\|A_n - A\| \to 0$. As an exercise, prove that the space of bounded linear operators with the uniform norm is a Banach space if Y is a Banach space.

From this point I will concentrate on the special case when $X = Y = \mathbf{R}^k$, and the linear operators are represented (in some bases) by the square matrices $\boldsymbol{A}$, $\boldsymbol{B}$, ... Matrix $\boldsymbol{A}$ represents a bounded linear operator, and its norm is given by

$$\|\boldsymbol{A}\| = \max_{\boldsymbol{x}\in\mathbf{R}^k,\, \boldsymbol{x}\neq 0} \frac{|\boldsymbol{A}\boldsymbol{x}|}{|\boldsymbol{x}|},$$

where again for a norm on Euclidian space $\mathbf{R}^k$ I use the notation $|\cdot|$.

Now consider the series

$$\boldsymbol{I} + \boldsymbol{A} + \frac{\boldsymbol{A}^2}{2!} + \ldots$$

and the corresponding series of norms

$$1 + \|\boldsymbol{A}\| + \frac{\|\boldsymbol{A}^2\|}{2!} + \ldots.$$

Using the fact that $\|\boldsymbol{A}^j\| \le \|\boldsymbol{A}\|^j$ I see that the partial sums are bounded by the partial sums of the series for e^a, $a := \|\boldsymbol{A}\|$:

$$1 + a + \frac{a^2}{2!} + \ldots = e^a,$$

I know that the last series converges to e^a for any $a \in \mathbf{R}$, and therefore the original series converges uniformly and absolutely, and hence there exists the sum of this series which I can denote $e^{\boldsymbol{A}}$. I also proved that $\|e^{\boldsymbol{A}}\| \le e^{\|\boldsymbol{A}\|}$. The validity of the definition of the matrix exponent was justified.

Exercise 3.4. Using the definition, calculate the matrix exponent for

$$(a)\ \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad (b)\ \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad (c)\ \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad (d)\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Exercise 3.5. Consider the set X of all polynomials of degree less than k.

1. Show that this set is a vector space. What is its dimension?

2. Consider the operator $\boldsymbol{A}$, which acts on the vector space X by taking the derivative: $P(x) \to \frac{\mathrm{d}}{\mathrm{d}x}P(x)$. Show that $\boldsymbol{A}$ is a linear operator.

3. Consider also the operator $\boldsymbol{H}^t$ that shifts a polynomial $P(x) \in X$ by t: i.e., $P(x) \to P(x+t)$. Show that $\boldsymbol{H}^t$ is a linear operator.

4. Prove that $e^{t\boldsymbol{A}} = \boldsymbol{H}^t$.

3.3 Properties of the matrix exponent

1. Matrix exponent is a linear bounded operator.

2. Assume that matrix $\boldsymbol{A}$ is diagonal, with diagonal elements $\lambda_1, \ldots, \lambda_k$. Matrix exponent in this case is also diagonal, with diagonal elements $e^{\lambda_1}, \ldots, e^{\lambda_k}$ (since $\boldsymbol{A}^m$ are all diagonal). Therefore, the calculation of $e^{\boldsymbol{A}}$ is simplest in the basis in which $\boldsymbol{A}$ is diagonal.

3. The family of linear operators $e^{t\boldsymbol{A}} \colon \mathbf{R}^k \longrightarrow \mathbf{R}^k$ defines a one-parameter group of linear transformations of $\mathbf{R}^k$ (i.e., in other words, the family $\{e^{t\boldsymbol{A}}\}$ defines a *linear flow*). This actually follows from the general theorem from the previous chapter, but I will show it directly. First, the group property has to be proved
$$e^{(t+s)\boldsymbol{A}} = e^{t\boldsymbol{A}} e^{s\boldsymbol{A}}.$$
To prove it, consider
$$\left(\boldsymbol{I} + t\boldsymbol{A} + \frac{t^2\boldsymbol{A}^2}{2!} + \ldots\right)\left(\boldsymbol{I} + s\boldsymbol{A} + \frac{s^2\boldsymbol{A}^2}{2!} + \ldots\right) = \\ \boldsymbol{I} + (t+s)\boldsymbol{A} + \left(\frac{t^2}{2} + ts + \frac{s^2}{2}\right)\boldsymbol{A}^2 + \ldots,$$
which proves the formula. To justify that we can multiply infinite series remember that these series converge absolutely.

   Second, one needs to show that
$$\frac{\mathrm{d}}{\mathrm{d}t} e^{t\boldsymbol{A}} = \boldsymbol{A} e^{t\boldsymbol{A}},$$
which follows from the formal differentiation of the series for the matrix exponent (again, the absolute convergence of the series allows term-wise differentiation).

4. The previous point actually proves

   Theorem 3.4. *The solution to the system*
$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$$
with the initial condition $\boldsymbol{x}(0) = \boldsymbol{x}_0$, is given by
$$\boldsymbol{x}(t) = e^{t\boldsymbol{A}}\boldsymbol{x}_0, \quad t \in \mathbf{R}.$$

   Proof. Using the formula for the derivative of the matrix exponent we see that $\boldsymbol{x}$ is actually a solution. Since $e^0 = \boldsymbol{I}$ then we also have $\boldsymbol{x}(0) = \boldsymbol{x}_0$, and the theorem of uniqueness yields that any solution to our problem coincides with $e^{t\boldsymbol{A}}\boldsymbol{x}_0$. ∎

5. Let $\boldsymbol{A} \colon \mathbf{R}^k \longrightarrow \mathbf{R}^k$ be a linear operator, and $\epsilon \in \mathbf{R}$.

   Lemma 3.5.
$$\det(\boldsymbol{I} + \epsilon\boldsymbol{A}) = 1 + \epsilon \operatorname{tr} \boldsymbol{A} + \mathcal{O}(\epsilon^2).$$

Proof. The determinant of the operator $\boldsymbol{I} + \epsilon\boldsymbol{A}$ is equal to the product of the corresponding eigenvalues $1 + \epsilon\lambda_j$, where λ_j are the eigenvalues of $\boldsymbol{A}$. Therefore,

$$\det(\boldsymbol{I} + \epsilon\boldsymbol{A}) = \prod_{j=1}^{k}(1 + \epsilon\lambda_j) = 1 + \epsilon\sum_{j=1}^{k}\lambda_j + \mathcal{O}(\epsilon^2).$$

■

Theorem 3.6.

$$\det e^{\boldsymbol{A}} = e^{\operatorname{tr}\boldsymbol{A}}.$$

Proof. I can define the matrix exponent also using the limit

$$e^{\boldsymbol{A}} = \lim_{m\to\infty}\left(\boldsymbol{I} + \frac{\boldsymbol{A}}{m}\right)^m.$$

I have

$$\det e^{\boldsymbol{A}} = \det\left(\lim_{m\to\infty}\left(\boldsymbol{I} + \frac{\boldsymbol{A}}{m}\right)^m\right) = \lim_{m\to\infty}\left(\det\left(\boldsymbol{I} + \frac{\boldsymbol{A}}{m}\right)^m\right),$$

since the determinant is a continuous function (as a polynomial). Next, using the previous lemma,

$$\left(\det\left(\boldsymbol{I} + \frac{\boldsymbol{A}}{m}\right)\right)^m = \left(1 + \frac{1}{m}\operatorname{tr}\boldsymbol{A} + \mathcal{O}(\frac{1}{m^2})\right)^m = e^{\operatorname{tr}\boldsymbol{A}}.$$

■

Exercise 3.6. Prove that the matrix exponent can be equivalently defined as

$$e^{\boldsymbol{A}} = \lim_{m\to\infty}\left(\boldsymbol{I} + \frac{\boldsymbol{A}}{m}\right)^m.$$

Therefore, I proved that the operator $\boldsymbol{A}$ is non-degenerate ($\det e^{\boldsymbol{A}} > 0$) and preserve the orientation of the space (recall that the determinant of a matrix is the oriented volume of a parallelepiped, whose edges are given by the columns of the matrix, and the determinant of a linear operator $\boldsymbol{A}$ is the oriented volume of the image of the unit cube under the mapping $\boldsymbol{A}$).

Corollary 3.7. *The phase flow* $\{\varphi^t\}$ *of the linear equation*

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \quad \boldsymbol{x}(t) \in \mathbf{R}^k$$

changes the volume of any figure in e^{at} *times during the time* t*. Here* $a = \operatorname{tr}\boldsymbol{A}$*.*

Proof. Indeed

$$\det\varphi^t = \det e^{t\boldsymbol{A}} = e^{\operatorname{tr} t\boldsymbol{A}} = e^{t\operatorname{tr}\boldsymbol{A}}.$$

■

If $\operatorname{tr}\boldsymbol{A} = 0$ then the phase flow of the linear system preserves the volume.

6. In general, if $\boldsymbol{AB} \neq \boldsymbol{BA}$ then

$$\exp(\boldsymbol{A}+\boldsymbol{B}) \neq \exp(\boldsymbol{A})\exp(\boldsymbol{B}).$$

Lemma 3.8. *If $\boldsymbol{A}$ and $\boldsymbol{B}$ commute, i.e.,*

$$[\boldsymbol{A},\boldsymbol{B}] := \boldsymbol{AB} - \boldsymbol{BA} = 0,$$

then

$$\exp(\boldsymbol{A}+\boldsymbol{B}) = \exp(\boldsymbol{A})\exp(\boldsymbol{B}).$$

Exercise 3.7. Prove Lemma 3.8.

Exercise 3.8. Show that $\left(e^{t\boldsymbol{A}}\right)^{-1} = e^{-t\boldsymbol{A}}$.

Exercise 3.9. Show that if $\boldsymbol{A}$ is skew-symmetric them $e^{\boldsymbol{A}}$ is orthogonal. Show that if $\boldsymbol{A}$ is skew-Hermitian then $e^{\boldsymbol{A}}$ is unitary.

Exercise 3.10. Consider matrices

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}.$$

Find

$$e^{t\boldsymbol{A}}, \quad e^{t\boldsymbol{B}}, \quad e^{t(\boldsymbol{A}+\boldsymbol{B})}$$

and conclude that in general

$$e^{\boldsymbol{A}}e^{\boldsymbol{B}} \neq e^{\boldsymbol{A}+\boldsymbol{B}}.$$

Exercise 3.11. To prove that $e^{t\boldsymbol{A}}\boldsymbol{x}_0$ gives all solutions to $\dot{\boldsymbol{x}} = \boldsymbol{Ax}$ I appealed to the general uniqueness theorem from the previous section. This can be avoided as follows. Let $\boldsymbol{x}$ be an arbitrary solution to $\dot{\boldsymbol{x}} = \boldsymbol{Ax}$. Consider $\boldsymbol{y}(t) = e^{-t\boldsymbol{A}}\boldsymbol{x}(t)$ and show that $\dot{\boldsymbol{y}}(t) = 0$ therefore, $\boldsymbol{x}(t) = e^{t\boldsymbol{A}}\boldsymbol{x}_0$.

Exercise 3.12. Is there a real 2×2 matrix $\boldsymbol{S}$ such that

$$e^{\boldsymbol{S}} = \begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix}?$$

Exercise 3.13. Show that if an operator $\boldsymbol{A}\colon \mathbf{R}^k \longrightarrow \mathbf{R}^k$ leaves *invariant* a subspace $E \in \mathbf{R}^k$ (that is, $\boldsymbol{Ax} \in E$ for all $\boldsymbol{x} \in E$) then $e^{t\boldsymbol{A}}$ also leaves E invariant.

Exercise 3.14. Suppose that the linear operator $\boldsymbol{A}\colon \mathbf{R}^k \longrightarrow \mathbf{R}^k$ has a real eigenvalue $\lambda < 0$. Show that the equation $\dot{\boldsymbol{x}} = \boldsymbol{Ax}$ has at least one nontrivial solution $\boldsymbol{x}(t)$ such that

$$\lim_{t\to\infty} \boldsymbol{x}(t) = 0.$$

3.4 Computation of the matrix exponent

Lemma 3.9. *Let $\boldsymbol{P}$ be a non-degenerate matrix, and $\boldsymbol{A} = \boldsymbol{PBP}^{-1}$ for some matrix $\boldsymbol{B}$. Then*

$$e^{\boldsymbol{A}} = \boldsymbol{P}e^{\boldsymbol{B}}\boldsymbol{P}^{-1}.$$

Proof.

$$(\boldsymbol{PBP}^{-1})^m = \boldsymbol{PB}^m\boldsymbol{P}^{-1}.$$

∎

3.4.1 The case of real eigenvalues

From the linear algebra course I know that if the matrix $\boldsymbol{A}$ has k eigenvalues (non necessarily distinct) such that the corresponding eigenvectors $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k\}$ form a basis of $\mathbf{R}^k$ then matrix $\boldsymbol{P}$ composed from the eigenvectors satisfies (which can be directly checked)

$$\boldsymbol{A}\boldsymbol{P}=\boldsymbol{P}\boldsymbol{\Lambda},$$

and $\boldsymbol{\Lambda}=\operatorname{diag}(\lambda_1,\ldots,\lambda_k)$. Here

$$\boldsymbol{P}=(\boldsymbol{v}_1|\ldots|\boldsymbol{v}_k),$$

i.e., the k-th eigenvector is the k-th column of $\boldsymbol{P}$. Therefore, the IVP

$$\dot{\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x},\quad \boldsymbol{x}(t)\in\mathbf{R}^k,$$

has the unique solution

$$\boldsymbol{x}(t)=\boldsymbol{P}e^{t\boldsymbol{\Lambda}}\boldsymbol{P}^{-1}\boldsymbol{x}_0.$$

Or, denoting, $\boldsymbol{\xi}=\boldsymbol{P}^{-1}\boldsymbol{x}_0$, I find a more convenient form to represent the general solution

$$\boldsymbol{x}(t)=\sum_{j=1}^{k}\xi_j\boldsymbol{v}_je^{\lambda_jt}.$$

3.4.2 The case of complex eigenvalues

Assume now that $\boldsymbol{A}\colon\mathbf{C}^k\longrightarrow\mathbf{C}^k$ and consider a complex system of ODE

$$\dot{\boldsymbol{z}}=\boldsymbol{A}\boldsymbol{z},$$

whose solution is given by $\boldsymbol{z}(t)=e^{t\boldsymbol{A}}\boldsymbol{z}_0$, which is a complex valued function of a real argument. Assume that the linear operator $\boldsymbol{A}$ is such that $\mathbf{C}^k=\bigoplus_{j=1}^k\mathbf{C}$, i.e. there are k complex eigenvalues whose eigenvectors form a basis of $\mathbf{C}^k$. Then exactly as in the real case I find that

$$\boldsymbol{z}(t)=\boldsymbol{P}e^{t\boldsymbol{\Lambda}}\boldsymbol{P}^{-1}\boldsymbol{z}_0=\sum_{j=1}^{k}\xi_j\boldsymbol{v}_je^{\lambda_jt}$$

is the general solution, however now $\boldsymbol{\xi}\in\mathbf{C}^k$, $\boldsymbol{v}_j\in\mathbf{C}^k$, $\lambda_j\in\mathbf{C}$.

If $\boldsymbol{A}$ is real, then I immediately have that for each complex eigenvalue λ_j its complex conjugate $\overline{\lambda}_j$ is also an eigenvalue, with the corresponding eigenvectors $\boldsymbol{v}_j$ and $\overline{\boldsymbol{v}}_j$. Moreover, if $\boldsymbol{z}$ is a solution then $\overline{\boldsymbol{z}}$ is also a solution. This implies that if $\boldsymbol{z}_0$ is real, then $\boldsymbol{z}$ is also real (due to the uniqueness theorem). The solution can be real if and only if the arbitrary constants ξ_j are such that ξ_j is real if λ_j is real, and ξ_j and $\overline{\xi}_j$ are two arbitrary constants corresponding to λ_j and $\overline{\lambda}_j$. This yields that the real valued solution will be given as

$$\boldsymbol{x}(t)=\sum_{j=1}^{\nu}\xi_j\boldsymbol{v}_je^{\lambda_jt}+\sum_{j=\nu+1}^{\nu+\mu}\left(\xi_j\boldsymbol{v}_je^{\lambda_jt}+\overline{\xi}_j\boldsymbol{v}_je^{\overline{\lambda}_jt}\right)=\sum_{j=1}^{\nu}\xi_j\boldsymbol{v}_je^{\lambda_jt}+2\sum_{j=\nu+1}^{\nu+\mu}\operatorname{Re}\left(\xi_j\boldsymbol{v}_je^{\lambda_jt}\right),$$

where the first ν eigenvalues are real and the rest are complex conjugate (μ pairs). The last formula can be rewritten as

$$\boldsymbol{x}(t)=\sum_{j=1}^{\nu}\xi_j\boldsymbol{v}_je^{\lambda_jt}+\eta_{\nu+1}\operatorname{Re}(\boldsymbol{v}_{\nu+1}e^{\lambda_{\nu+1}t})+\eta_{\nu+2}\operatorname{Im}(\boldsymbol{v}_{\nu+2}e^{\lambda_{\nu+2}t})+\ldots,$$

where now all ξ_j and η_j are all real.

Exercise 3.15. Prove that for the system $\dot{\boldsymbol{z}}=\boldsymbol{A}\boldsymbol{z}$ with the real $\boldsymbol{A}$ if $\boldsymbol{z}$ is a solution then $\overline{\boldsymbol{z}}$ is also a solution.

The last representation allows me to get the following geometric picture. I assume that I have k eigenvalues of $\boldsymbol{A}$ such that ν are real and μ complex conjugate pairs, and the list of eigenvectors forms a basis of $\mathbf{C}^k$. Then $\mathbf{R}^k$ can be represented as a direct sum of invariant with respect to $\boldsymbol{A}$ ν one dimensional and μ two dimensional subspaces. Indeed, if I have a pair of conjugate eigenvalues λ and $\overline{\lambda}$ with the eigenvectors $\boldsymbol{v}$ and $\overline{\boldsymbol{v}}$, then consider their real and imaginary parts:

$$\boldsymbol{x}=\frac{\boldsymbol{v}+\overline{\boldsymbol{v}}}{2}\in\mathbf{R}^k,\quad \boldsymbol{y}=\frac{\boldsymbol{v}-\overline{\boldsymbol{v}}}{2\mathrm{i}}\in\mathbf{R}^k,$$

which are linearly independent. The subspace spanned $\boldsymbol{v}$ and $\overline{\boldsymbol{v}}$ is invariant in $\mathbf{C}^k$, therefore, the subspace spanned by $\boldsymbol{x}$ and $\boldsymbol{y}$ is also invariant in $\mathbf{C}^k$. Their linear combination is real if and only if the coefficients are real, and theretofore forms a two dimensional invariant subspace of $\boldsymbol{A}$ in $\mathbf{R}^k$.

Exercise 3.16. Carefully fill in all the missing details in the reasoning above.

Corollary 3.10. *Let $\boldsymbol{x}=(x_1,\ldots,x_k)^\top$ be the solution of the linear system of real ODE with the matrix $\boldsymbol{A}$. Let all the eigenvalues of $\boldsymbol{A}$ be simple. Then each of the functions x_j is a linear combination of e^{λ_kt} and $e^{\alpha_kt}\sin\omega_kt$, $e^{\alpha_kt}\cos\omega_kt$, where λ_k are the real and $\alpha_k\pm\mathrm{i}\omega_k$ are the complex eigenvalues of $\boldsymbol{A}$.*

Corollary 3.11. *Let $\boldsymbol{A}$ be a real square matrix with simple eigenvalues. Then each of the elements of $e^{t\boldsymbol{A}}$ is a linear combination of e^{λ_kt}, $e^{\alpha_kt}\sin\omega_kt$, $e^{\alpha_kt}\cos\omega_kt$, where λ_k are the real and $\alpha_k\pm\mathrm{i}\omega_k$ are the complex eigenvalues of $\boldsymbol{A}$.*

3.4.3 The case of multiple eigenvalues

In the eigenvalue problem $\boldsymbol{A}\boldsymbol{v}=\lambda\boldsymbol{v}$ the eigenvalues are the roots of the characteristic polynomial

$$\det(\boldsymbol{A}-\lambda\boldsymbol{I})=0=\prod_{j=1}^{l}(\lambda-\lambda_j)^{a_j}.$$

The numbers a_j are called the *algebraic multiplicities* of the eigenvalues λ_j, and

$$b_j=\dim\ker(\boldsymbol{A}-\lambda_j\boldsymbol{I})$$

are the *geometric multiplicities*. In general I always have that $b_j\leq a_j$, and if $b_j<a_j$ then, as it is known from the course of linear algebra, the operator $\boldsymbol{A}$ cannot be written in a diagonal form even

as an operator on $\mathbf{C}^k$. Instead, Jordan's blocks appear, for which I would like to calculate the matrix exponent.

Let $\boldsymbol{A}$ be such that

$$\boldsymbol{A} = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}.$$

To calculate the matrix exponent I will use the fact that

$$\boldsymbol{A} = \lambda \boldsymbol{I} + \boldsymbol{N},$$

where $\boldsymbol{N}$ is *nilpotent* (a matrix $\boldsymbol{B}$ is nilpotent if there exists a natural number n such that $\boldsymbol{B}^n = 0$). Since $\boldsymbol{I}$ commutes with anything, I have

$$e^{\boldsymbol{A}} = e^{\boldsymbol{\lambda I}} e^{\boldsymbol{N}}.$$

And using the fact

$$\boldsymbol{N}^l = \begin{bmatrix} 0 & \dots & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ & & & & \vdots \\ & & & & 0 \end{bmatrix},$$

I obtain

$$e^{t\boldsymbol{N}} = \begin{bmatrix} 1 & t & t^2/2 & \dots & t^{n-1}/(n-1)! \\ & 1 & t & \ddots & \vdots \\ & & 1 & \ddots & t^2/2 \\ & & & \ddots & t \\ & & & & 1 \end{bmatrix},$$

and

$$e^{t\boldsymbol{A}} = e^{\lambda t} e^{t\boldsymbol{N}}.$$

Let λ be a real number. A *quasi-polynomial* with the exponent λ is the product $e^{\lambda t} P(t)$, where $P(t)$ is a polynomial. The degree of P is called the degree of the quasi-polynomial. If λ is fixed then the set of all quasi-polynomials of degree less than k is a vector space (prove it and find its dimension).

Corollary 3.12. *Let $\boldsymbol{A}\colon \mathbf{C}^k \longrightarrow \mathbf{C}^k$ be a linear operator, $\lambda_1, \dots, \lambda_m$ be the eigenvalues with the algebraic multiplicities $a_1, \dots, a_m$, $t \in \mathbf{R}$. Then every element of the matrix $e^{t\boldsymbol{A}}$ is a sum of quasi-polynomials of the variable t with the exponents λ_j of degrees less than a_j.*

Corollary 3.13. *Let $\boldsymbol{x}$ be a solution to $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$. Then each component of the vector $\boldsymbol{x}$ is a sum of quasi-polynomials of variable t with the exponents λ_j of degrees less than a_j:*

$$x_i(t) = \sum_{j=1}^{m} e^{\lambda_j t} P_{ij}(t),$$

where $P_{ij}(t)$ is a polynomial of degree less than a_j.

Corollary 3.14. *Let $\boldsymbol{A}\colon \mathbf{R}^k \longrightarrow \mathbf{R}^k$ be a linear operator, λ_j, $1 \leq j \leq m$, be its real eigenvalues with algebraic multiplicities a_j, $\alpha_j \pm \mathrm{i}\omega_j$, $1 \leq l \leq r$, be complex eigenvalues with algebraic multiplicities d_j. Then each element of the matrix $e^{t\boldsymbol{A}}$ and each component of the solution to $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ is a sum of complex quasi-polynomials with exponents λ_j, $\alpha_j + \mathrm{i}\omega_j$ of degrees less than a_j and d_j respectively.*

Remark 3.15. Contrary to the case when all the eigenvalues of $\boldsymbol{A}$ are simple, I do not provide here the exact form of the general solution to $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ in the case when $\boldsymbol{A}$ has multiple eigenvalues. It is very seldom in applications that someone actually needs to calculate this exact form.

Exercise 3.17. Calculate $e^{t\boldsymbol{A}}$ for

$$(a)\ \begin{bmatrix} 5 & -6 \\ 3 & -4 \end{bmatrix}, \quad (b)\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (c)\ \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 1 & 3 \end{bmatrix}.$$

Exercise 3.18. Find $e^{\boldsymbol{A}}$, where

$$\boldsymbol{A} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}.$$

Exercise 3.19. Let a and b be two distinct eigenvalues of 2×2 matrix $\boldsymbol{A}$. Show that

$$e^{t\boldsymbol{A}} = \frac{e^{at} - e^{bt}}{a - b}\boldsymbol{A} + \frac{ae^{bt} - be^{at}}{a - b}\boldsymbol{I}.$$

3.5 Planar linear ODE systems with constant coefficient

Since in the case of semisimple $\boldsymbol{A}$ (operator $\boldsymbol{A}$ is called *semisimple* if it is diagonalizable over $\mathbf{C}^k$) the phase space $\mathbf{R}^k$ splits into the direct sum of one and two dimensional subspaces, a lot of insight about the behavior of solutions to $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ can be gained by studying two dimensional systems of the form

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \quad \boldsymbol{x}(t) \in \mathbf{R}^2. \tag{3.5}$$

I start with a basic fact from linear algebra:

Theorem 3.16. *Let $\boldsymbol{A}$ be a 2×2 real matrix. Then there exists a real invertible 2×2 matrix $\boldsymbol{P}$ such that*

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J},$$

where matrix $\boldsymbol{J}$ is one of the following three matrices in real Jordan's normal form

$$(a)\quad \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad (b)\quad \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \quad (c)\quad \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

Proof. Since the characteristic polynomial has degree two, it may have either two real roots, two complex conjugate roots, or one real root multiplicity two.

(a)I assume that I have two distinct real roots $\lambda_1 \in \mathbf{R} \neq \lambda_2 \in \mathbf{R}$ with the corresponding eigenvectors $\boldsymbol{v}_1 \in \mathbf{R}^2$ and $\boldsymbol{v}_2 \in \mathbf{R}^2$ or a real root $\lambda \in \mathbf{R}$ multiplicity two which has two linearly independent eigenvectors $\boldsymbol{v}_1 \in \mathbf{R}^2$ and $\boldsymbol{v}_2 \in \mathbf{R}^2$. Matrix $\boldsymbol{P}$ now can be taken simply as

$$\boldsymbol{P} = (\boldsymbol{v}_1 \mid \boldsymbol{v}_2),$$

i.e., the j-th column is the j-th eigenvector. The eigenvectors corresponding to distinct eigenvalues are linearly independent, hence $\boldsymbol{P}$ is invertible. Now

$$\boldsymbol{AP} = (\boldsymbol{Av}_1|\boldsymbol{Av}_2) = (\lambda_1\boldsymbol{v}_1|\lambda_2\boldsymbol{v}_2) = \boldsymbol{PJ}.$$

For the case (b), I assume that there is one real root of the characteristic polynomial with the eigenvector $\boldsymbol{v}_1$. Then there is another vector $\boldsymbol{v}_2$, which satisfies

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v}_2 = \boldsymbol{v}_1,$$

which is linearly independent of $\boldsymbol{v}_1$. Now take $\boldsymbol{P} = (\boldsymbol{v}_1 \mid \boldsymbol{v}_2)$, and

$$\boldsymbol{AP} = (\lambda\boldsymbol{v}_1 \mid \boldsymbol{v}_1 + \lambda\boldsymbol{v}_2) = \boldsymbol{PJ},$$

where $\boldsymbol{J}$ as in (b).

Finally, in the case (c) I have $\lambda_{1,2} = \alpha \pm \mathrm{i}\beta$ as eigenvalues and the corresponding eigenvectors $\boldsymbol{v}_1 \pm \mathrm{i}\boldsymbol{v}_2$, where $\boldsymbol{v}_1, \boldsymbol{v}_2$ are real nonzero vectors. Let me take $\boldsymbol{P} = (\boldsymbol{v}_1 \mid \boldsymbol{v}_2)$. Since

$$\boldsymbol{A}(\boldsymbol{v}_1 + \mathrm{i}\boldsymbol{v}_2) = (\alpha + \mathrm{i}\beta)(\boldsymbol{v}_1 + \mathrm{i}\boldsymbol{v}_2),$$

I have

$$\boldsymbol{Av}_1 = \alpha\boldsymbol{v}_1 - \beta\boldsymbol{v}_2, \quad \boldsymbol{Av}_2 = \alpha\boldsymbol{v}_2 + \beta\boldsymbol{v}_1.$$

Now

$$\boldsymbol{AP} = (\alpha\boldsymbol{v}_1 - \beta\boldsymbol{v}_2 \mid \beta\boldsymbol{v}_1 + \alpha\boldsymbol{v}_2) = \boldsymbol{PJ},$$

where $\boldsymbol{J}$ as in (c). The only missing point is to prove that $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are linearly independent, which is left as an exercise. ■

Now I only need to calculate the matrix exponent for all three cases to solve any planar system of the form (3.5). I have in the case (a) that

$$e^{t\boldsymbol{A}} = \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix},$$

in case (b), using the decomposition into two commuting matrices,

$$e^{t\boldsymbol{A}} = e^{\lambda t}\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix},$$

and finally, in the case (c), using

$$\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix} + \begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix},$$

one can show, using the definition of the matrix exponent, that

$$e^{t\boldsymbol{A}} = e^{\alpha t}\begin{bmatrix} \cos\beta t & \sin\beta t \\ -\sin\beta t & \cos\beta t \end{bmatrix}.$$

Exercise 3.20. Fill in the missing details.

For the two dimensional systems it is convenient to represent solutions graphically as parametrically defines curves $t \mapsto \boldsymbol{x}(t) \in \mathbf{R}^2$, for which the changes of the variable t from smaller to bigger values define the direction along these curves. Using the terminology of the dynamical system theory, $\mathbf{R}^2$ in this case is called the *phase* or *state space*, and the images of solutions $\boldsymbol{x}$ in the phase space parameterized by the time t are called the *phase curves* or *phase orbits*. So my task is, given the matrix $\boldsymbol{A}$, is to understand the structure of the orbits on the phase plane — the *phase portrait*.

First, I assume for simplicity that $\lambda_1\lambda_2 \neq 0$. In this case matrix $\boldsymbol{A}$ is non-degenerate, and hence the only solution to the algebraic system $\boldsymbol{A}\boldsymbol{x} = 0$ is the trivial one, $\hat{\boldsymbol{x}} = (0, 0)$. This point is called an *equilibrium*, note that if I have the initial conditions at this point, I will stay at this point for ever. Second, it is enough to understand the structure of the phase portraits of the systems with the matrices in the Jordan normal form, because all other phase portraits are obtained from these by the application of a non-degenerate linear operator $\boldsymbol{P}$, which corresponds to possible stretching, rotations, and/or reflections.

(a) Case of the two real eigenvalues. The general solution to (3.5) with the matrix (a) is given by

$$\boldsymbol{x}(t;\boldsymbol{x}_0) = \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix} \boldsymbol{x}_0 = \begin{bmatrix} e^{\lambda_1 t} x_1^0 \\ e^{\lambda_2 t} x_2^0 \end{bmatrix}.$$

The phase curves can be found as solutions to the first order ODE

$$\frac{\mathrm{d}x_2}{\mathrm{d}x_1} = \frac{\lambda_2 x_2}{\lambda_1 x_1},$$

which is a separable equation, and the directions on the orbits are easily determined by the signs of λ_1 and λ_2 (i.e., if $\lambda_1 < 0$ then $x_1(t) \to 0$ as $t \to \infty$).

Consider a specific example with $0 < \lambda_1 < \lambda_2$. In this case I have that all the orbits are "parabolas," and the direction on the orbits is *from* the origin because both λ's are positive. The only slightly tricky part here is to determine which axis the orbits approach as $t \to -\infty$. This can be done by looking at the explicit equations for the orbits (you should do it) or by noting that when $t \to -\infty$ $e^{\lambda_1 t} \gg e^{\lambda_2 t}$ and therefore x_1 component dominates x_2 in a small enough neighborhood of $(0, 0)$ (see Fig. 3.1, left). The obtained phase portrait is called *topological node* ("topological" is often dropped), and since the arrows point from the origin, it is unstable (I will come back shortly to the discussion of *stability*).

As another example consider the case when $\lambda_2 < 0 < \lambda_1$. In this case (prove it) the orbits are actually "hyperbolas" on (x_1, x_2) plane, and the directions on them can be identifies by noting that on x_1-axis the movement is *from* the origin, and on x_2-axis it is *to* the origin. Such phase portrait is called *saddle* (see Fig. 3.1, middle). All the orbits leave a neighborhood of the origin for both $t \to \pm\infty$ except for five special orbits: first, this is of course the origin itself, second, two orbits on x_1-axis that actually approach the origin if $t \to -\infty$, and, third, two orbits on x_2-axis, which approach the origin if $t \to \infty$. The two orbits on x_1-axis form the *unstable manifold* of the point $\hat{\boldsymbol{x}} = (0, 0)$, and the orbits on x_2-axis form the *stable manifold* of $\hat{\boldsymbol{x}}$. These orbits are also called the saddle's *separatrices* (singular, *separatrix*).

There are several other cases to consider:

- $0 < \lambda_1 < \lambda_2$: unstable node (shown in the figure)
- $0 < \lambda_2 < \lambda_1$: unstable node

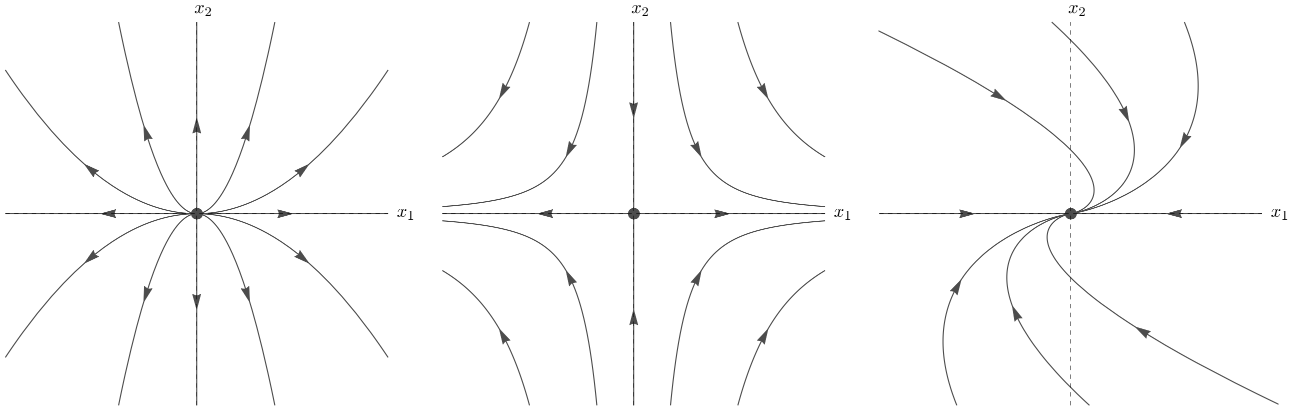

Figure 3.1: Left: Unstable node. Middle: Saddle. Right: Improper stable node

- $0 < \lambda_1 = \lambda_2$: unstable node
- $\lambda_1 < \lambda_2 < 0$: stable node
- $\lambda_2 < \lambda_1 < 0$: stable node
- $\lambda_1 = \lambda_2 < 0$: stable node
- $\lambda_1 < 0 < \lambda_2$: saddle
- $\lambda_2 < 0 < \lambda_1$: saddle (shown in the figure)

You should sketch the phase portraits for each of these cases. Also keep in mind that for now I exclude cases when one or both λ's are zero.

(b) I assume that $\lambda < 0$ (the case $\lambda > 0$ left as an exercises). Now, first, I see from the general solution (write it down!) that $\boldsymbol{x}(t; \boldsymbol{x}_0) \to 0$ as $t \to \infty$, moreover,

$$\frac{\mathrm{d}x_2}{\mathrm{d}x_1} \to 0$$

as $t \to \infty$, therefore the orbits should be tangent to x_1-axis. The phase portrait (Fig. 3.1, right) is sometimes called an *improper stable node.*

(c) The flow of (3.5) is given by

$$\boldsymbol{x}(t; \boldsymbol{x}_0) = e^{\boldsymbol{A}_3 t}\boldsymbol{x}_0 = e^{\alpha t}\begin{bmatrix} \cos\beta t & \sin\beta t \\ -\sin\beta t & \cos\beta t \end{bmatrix}\boldsymbol{x}_0.$$

To determine the phase portrait observe that if $\alpha < 0$ then all the solutions will approach the origin, and if $\alpha > 0$, they will go away from the origin. I also have components of $e^{\boldsymbol{J}t}$, which are periodic functions of t, which finally gives us the whole picture: if $\alpha < 0$ and $\beta > 0$ then the orbits are the spirals approaching the origin clockwise, if $\alpha > 0$ and $\beta > 0$ then the orbits are spiral unwinding from the origin clockwise, and if $\alpha = 0$ then the orbits are closed curves. An example for $\alpha < 0$ and $\beta < 0$ in given in Fig. 3.2, this phase portrait is called the *stable focus* (or *spiral*).

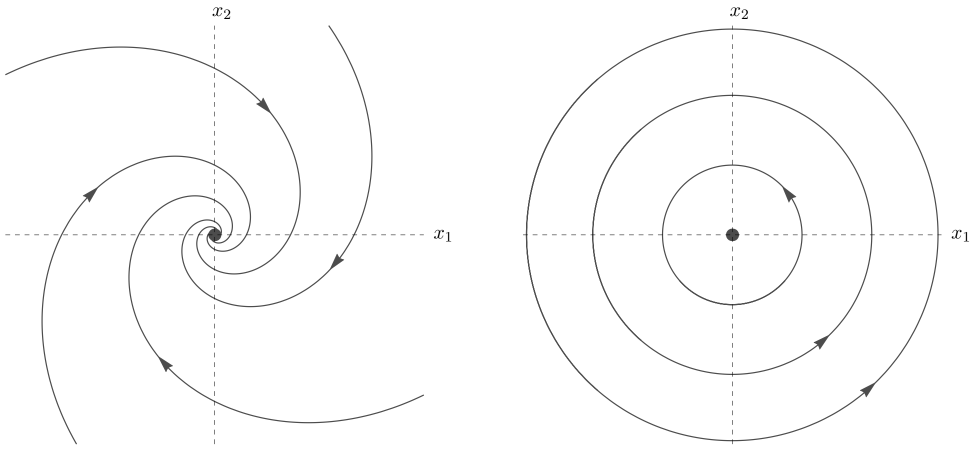

Figure 3.2: Left: Stable focus. Right: Center

If I take $\alpha = 0$ and $\beta < 0$ then the phase portrait is composed of the closed curves and called the *center*. See Fig. 3.2, right.

In the general situation, to determine the direction on the orbits, I can use the original vector field. For example, in the case $\alpha = 0$ $\beta < 0$ I have that for any point $x_1 = 0$ and $x_2 > 0$ the derivative of x_2 is negative, and therefore the direction is counter clockwise.

Example 3.17. Consider system (3.5) with

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}.$$

I find that the eigenvalues and eigenvectors are

$$\lambda_1 = -2, \quad \boldsymbol{v}_1^\top = (-1, 1), \qquad \lambda_2 = 2, \quad \boldsymbol{v}_2^\top = (3, 1).$$

Therefore, the transformation $\boldsymbol{P}$ here is

$$\boldsymbol{P} = \begin{bmatrix} -1 & 3 \\ 1 & 1 \end{bmatrix},$$

and

$$\boldsymbol{J} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix}.$$

The solution to the system

$$\dot{\boldsymbol{y}} = \boldsymbol{J}\boldsymbol{y},$$

where $\boldsymbol{y} = \boldsymbol{P}^{-1}\boldsymbol{x}$, is straightforward and given by

$$\boldsymbol{y}(t; \boldsymbol{x}_0) = \begin{bmatrix} e^{-2t} & 0 \\ 0 & e^{2t} \end{bmatrix} \boldsymbol{y}_0$$

and its phase portrait has the structure of a saddle (see Fig. 3.3, left). To see how actually the phase portrait looks in $\boldsymbol{x}$ coordinates, consider the solution for $\boldsymbol{x}$, which takes the form

$$\boldsymbol{x} = \boldsymbol{P}\boldsymbol{y} = (\boldsymbol{v}_1 e^{\lambda_1 t} \mid \boldsymbol{v}_2 e^{\lambda_2 t})\boldsymbol{y}_0 = C_1 \boldsymbol{v}_1 e^{\lambda_1 t} + C_2 \boldsymbol{v}_2 e^{\lambda_2 t},$$

where I use C_1, C_2 for arbitrary constants. Note that $\boldsymbol{x}$ is changing along the straight line with the direction $\boldsymbol{v}_1$ if $C_2 = 0$, and along the straight line $\boldsymbol{v}_2$ when $C_1 = 0$. The directions of the flow on these lines coincide with the directions of the flow on the axes for the system in $\boldsymbol{y}$ coordinates (see Fig. 3.3).

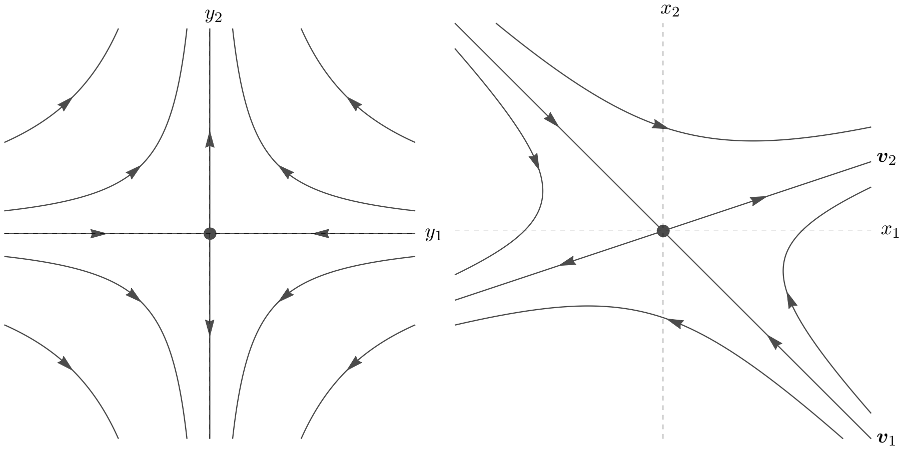

Figure 3.3: Saddle point after the linear transformation (left), and the original phase portraits (right). I have $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{y}$

To summarize, to sketch a phase portrait of a two-dimensional linear ODE system with $\det \boldsymbol{A} \neq 0$ one needs: Find the eigenvalues. If these eigenvalues are complex conjugate, $\lambda_1 = \overline{\lambda}_2 = \alpha + \mathrm{i}\beta$, then if $\alpha < 0$ it is a stable focus, $\alpha > 0$ — unstable focus, and if $\alpha = 0$ it is a center. The direction of rotation (counter- or clockwise) can be determined by determining the direction of the corresponding vector field at any point on the plane. If the eigenvalues are real, then the corresponding eigenvectors have to be found. These eigenvectors define the directions for the straight lines on which the solutions are invariant (if a solution happens to be on this straight line, then it will never leave it). The directions on these straight lines are determined by the signs of the corresponding eigenvalues, if the sign is negative then the direction is to the origin, in the opposite case the direction is from the origin. The actual direction along which other orbits enter the origin (for t to plus or minus infinity) is determined by the absolute values of the eigenvalues. If one finds one eigenvalue multiplicity two with only one linearly independent eigenvector, then this eigenvector determines the direction along which the orbits approach the origin.

I can summarize all the information on the types of linear planar systems in one *parametric portrait* of (3.5). The characteristic polynomial is

$$P(\lambda) = \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = \lambda^2 + \lambda \operatorname{tr} \boldsymbol{A} + \det \boldsymbol{A},$$

I hence have

$$\lambda_{1,2} = \frac{\operatorname{tr} \boldsymbol{A} \pm \sqrt{(\operatorname{tr} \boldsymbol{A})^2 - 4 \det \boldsymbol{A}}}{2}.$$

Using the trace and determinant as the new parameters I can actually present all possible types of planer linear systems in one Fig. 3.4.

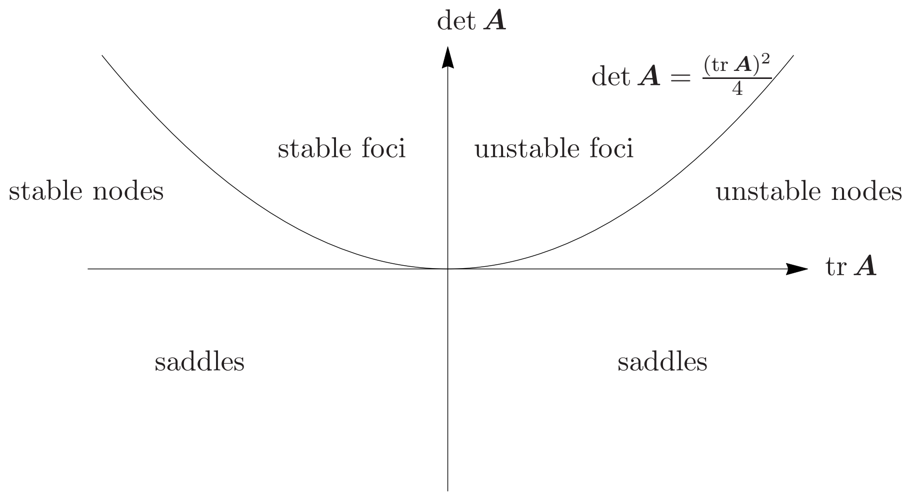

Figure 3.4: The type of the linear system depending on the values of tr $\boldsymbol{A}$ and det $\boldsymbol{A}$. The centers here are situated where det $\boldsymbol{A} > 0$ and tr $\boldsymbol{A} = 0$

Exercise 3.21. Sketch the phase portraits for the linear planar system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is given by

$$(a)\quad \begin{bmatrix} -1 & 0 \\ 2 & -2 \end{bmatrix},\quad (b)\quad \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix},\quad (c)\quad \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix},\quad (d)\quad \begin{bmatrix} -1 & 2 \\ -1 & 1 \end{bmatrix},\quad (e)\quad \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}.$$

Classify the origin for these systems, and identify in each case those vectors $\boldsymbol{u} \in \mathbf{R}^2$ such that $\boldsymbol{x}(t;\boldsymbol{u}) \to 0$.

Exercise 3.22. Which value (if any) of the parameter k in the following matrices makes the origin a sink for the corresponding differential equation $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ with $\boldsymbol{A}$ as follows:

$$(a)\quad \begin{bmatrix} a & -k \\ k & 2 \end{bmatrix},\quad (b)\quad \begin{bmatrix} 3 & 0 \\ k & -4 \end{bmatrix},\quad (c)\quad \begin{bmatrix} k^2 & 1 \\ 0 & k \end{bmatrix},\quad (d)\quad \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & k \end{bmatrix}?$$

Exercise 3.23. Let $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ have non-real eigenvalues. Show that $b \neq 0$. Show that the nontrivial solution curves to $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ are spiral or ellipses that are oriented clockwise if $b > 0$ and counterclockwise if $b < 0$.

Hint: Consider the sign of

$$\frac{\mathrm{d}}{\mathrm{d}t} \arctan \frac{x_2(t)}{x_1(t)}.$$

Exercise 3.24. Classify and sketch the phase portraits of planar differential equation $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ has a zero eigenvalue.

Exercise 3.25. Let $\boldsymbol{A}$ be $k \times k$ real matrix, where k is odd. Show that there exist a nonperiodic solution to $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$.

Exercise 3.26. Let problem $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ with real $\boldsymbol{A} = (a_{ij})_{2\times 2}$ matrix have one periodic solution. Show that all the solutions are periodic.

In a similar way, especially for semisimple operators, I can discuss phase portraits for higher dimensional linear systems of ODE.

Exercise 3.27. Sketch phase portraits for the system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ with eigenvalues $\lambda_1, \lambda_2, \lambda_3$, where $\boldsymbol{A}$ is a 3×3 matrix if

(*a*) $\lambda_1 < \lambda_2 < \lambda_3 < 0$,

(*b*) $\lambda_1 < 0, \lambda_2 = \alpha + \imath\beta$, $\alpha < 0, \beta > 0$,

(*c*) $\lambda_1 < 0, \lambda_2 = \alpha + \imath\beta$, $\alpha > 0, \beta > 0$,

(*d*) $\lambda_1 < 0, \lambda_2 = \lambda_3$, and $\boldsymbol{A}$ is semisimple,

(*e*) $\lambda_1 < \lambda_2 < 0 < \lambda_3$.

What is more important, however, is that the explicit form of the solutions to the linear systems allows important qualitative conclusions. First, let me introduce the term *stability* rigorously for the first time.

Definition 3.18. *The linear system of ODE* (3.5) *is called* (*Lyapunov*) *stable, if all solutions remain bounded for* $t \to \infty$. *It is called* (*globally*) *asymptotically stable if all solutions converge to* 0 *if* $t \to \infty$. *If system is not stable it is called unstable.*

The explicit form of solutions to the linear system with constant coefficients implies

Theorem 3.19. *The linear system* (3.5) *is asymptotically stable if and only if the eigenvalues* λ_j *of* $\boldsymbol{A}$ *satisfy* $\operatorname{Re}\lambda_j < 0$. *Moreover, in this case there exist constants* $C > 0$ *and* $\alpha > 0$ *such that*

$$|e^{t\boldsymbol{A}}\boldsymbol{x}_0| \leq Ce^{-t\alpha}, \quad t \geq 0.$$

The linear system (3.5) *is stable if and only if the eigenvalues of* $\boldsymbol{A}$ *satisfy* $\operatorname{Re}\lambda_j \leq 0$, *and algebraic multiplicities of the eigenvalues with* $\operatorname{Re}\lambda_j = 0$ *coincide with their geometric multiplicities. Moreover, in this case there exists* $C > 0$ *such that*

$$|e^{t\boldsymbol{A}}\boldsymbol{x}_0| \leq C, \quad t \geq 0.$$

The linear system (3.5) *is unstable if and only if there exists an eigenvalue* λ_j *of* $\boldsymbol{A}$ *that satisfies* $\operatorname{Re}\lambda_j > 0$, *or there exists an eigenvalue* λ_j *of* $\boldsymbol{A}$ *that satisfies* $\operatorname{Re}\lambda_j = 0$ *and its algebraic multiplicity is strictly bigger than its geometric multiplicity.*

Exercise 3.28. Prove Theorem 3.19.

Finally, consider a non-homogeneous system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{g}(t), \tag{3.6}$$

where $\boldsymbol{g} \in \mathcal{C}(I; \mathbf{R}^k)$. Exactly as in one dimensional case I can use the variation of the constant method to show that the general solution is given by

$$\boldsymbol{x}(t) = e^{t\boldsymbol{A}}\boldsymbol{x}_0 + \int_0^t e^{(t-\tau)\boldsymbol{A}}\boldsymbol{g}(\tau)\,\mathrm{d}\tau.$$

Note the structure of the general solution, which consists of two parts: The general solution to the homogeneous equation plus a particular solution to the non-homogeneous one.

Exercise 3.29. Prove the formula for the general solution to (3.6).

Exercise 3.30. Using the variation of the constant method solve the following nonhomogeneous system

$$\dot{\boldsymbol{x}} = \begin{bmatrix} 0 & 1 \\ -4 & 0 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0 \\ \sin 2t \end{bmatrix}.$$

Exercise 3.31. Suppose $\boldsymbol{T}\colon \mathbf{R}^k \longrightarrow \mathbf{R}^k$ is an invertible linear operator and $\boldsymbol{c} \in \mathbf{R}^k$ is a nonzero constant vector. Show that there is a change of coordinates of the form

$$\boldsymbol{x} = \boldsymbol{P}\boldsymbol{y} + \boldsymbol{b}, \quad \boldsymbol{b} \in \mathbf{R}^k,$$

transforming the nonhomogeneous equation $\dot{\boldsymbol{x}} = \boldsymbol{T}\boldsymbol{x} + \boldsymbol{c}$ into homogeneous form $\dot{\boldsymbol{y}} = \boldsymbol{S}\boldsymbol{y}$. Find $\boldsymbol{P}$, $\boldsymbol{b}$, $\boldsymbol{S}$.

Exercise 3.32. Solve

$$x' = y, \quad y' = 2 - x.$$

Hint: The previous problem.

3.6 Linear equations of the k-th order

3.6.1 The general theory

The linear ordinary differential equation of the k-th order with constant coefficients takes the form

$$x^{(k)} + a_{k-1}x^{(k-1)} + \ldots + a_1 x' + a_0 x = g(t). \tag{3.7}$$

It requires k initial conditions

$$x(0) = x_0, \quad x'(0) = x_1, \quad \ldots \quad x^{(k-1)}(0) = x_{k-1}.$$

It is called homogeneous if $g(t) = 0$ and non-homogeneous otherwise. If I convert equation (3.7) into a system of the form (3.5), then the matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ -a_0 & -a_1 & \ldots & \ldots & -a_{k-1} \end{bmatrix},$$

which has a very special form (sometimes it is called a *companion matrix*). Using the expansion of the determinant $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$ with respect to the last row, I find that the eigenvalues are the roots of the characteristic polynomial

$$\lambda^k + a_{k-1}\lambda^{k-1} + \ldots + a_1\lambda + a_0 = 0.$$

Moreover, the geometric multiplicity of every eigenvalue is one (can you prove this claim?), which implies that the general solution to the homogeneous equation is given as a linear combination of

$$e^{\lambda_j t},\, te^{\lambda_j}, \ldots, t^a e^{\lambda_j t}, \quad j = 1, \ldots, m, \quad 0 \leq a < c_j,$$

where m is the number of distinct eigenvalues and c_j is the algebraic multiplicity of the j-th eigenvalue. If there is a pair of complex conjugate eigenvalues λ and $\overline{\lambda}$ of the form $\lambda = \alpha \pm \mathrm{i}\beta$, then the above is replaced with

$$e^{\alpha t} \cos \beta t,\ e^{\alpha t} \sin \beta t, \ldots, \quad t^a e^{\alpha t} \cos \beta t,\ t^a e^{\alpha t} \sin \beta t, \quad 0 \leq a < c,$$

where c is the algebraic multiplicity of λ (and hence of $\overline{\lambda}$).

Exercise 3.33. Show that the characteristic polynomial of the companion matrix indeed has the required form.

Exercise 3.34. Solve

$$\sum_{j=0}^{k} \frac{\mathrm{d}^j x}{\mathrm{d}t^j}(t) = 0.$$

Exercise 3.35. For which $a, b \in \mathbf{C}$ all the solutions to

$$\ddot{x} + a\dot{x} + bx = 0$$

are bounded for $-\infty < t < \infty$?

Exercise 3.36. For which a and b all solutions to

$$\ddot{x} + a\dot{x} + bx = 0$$

tend to zero as $t \to +\infty$?

Exercise 3.37. Consider the equation

$$x''' + ax'' + bx' + cx = 0.$$

Find necessary and sufficient conditions on a, b, c to guarantee that the origin is asymptotically stable.

For the nonhomogeneous equation (3.7) it is also possible to write a general formula (using the same method of the variation of the constants), but sometimes it is more convenient to use the so-called method of undetermined coefficients, when a particular solution to (3.7) is first guessed, based on the form of g, in a specific form with arbitrary coefficients, and after this this coefficients are determined. This method works when g is a quasi-polynomial, i.e., a function of the form

$$e^{at} P(t),$$

where a is some constant, and P is a polynomial (note that using Euler's formula $e^{\mathrm{i}\theta} = \cos\theta + \mathrm{i} \sin\theta$ also includes in this expression the possibilities $e^{at}P(t)\cos bt$ and $e^{at}P(t)\sin bt$). What is so special about quasi-polynomials? They are solutions of some homogeneous linear ODE with constant coefficients! So, I write (3.7) in a concise form

$$Lx = g,$$

where L is a linear k-th order differential operator with constant coefficients, and g itself solves ODE $Hg = 0$ (it is said that H *annihilates* g) then the equation

$$HLx = Hg = 0$$

is a homogeneous linear ODE with constant coefficients which I know how to solve! In particular, the knowledge of the form of solutions of the homogeneous ODE yields the following rule of thumb:

Let $g(t) = e^{at}P_n(t)$, where a is some constant and $P_n(t)$ is a polynomial of degree n. Then a particular solution x_p to (3.7) should be looked for in the form

$$x_p(t) = e^{at}t^cQ_n(t),$$

where c is the algebraic multiplicity of a as an eigenvalue of the homogeneous equation (if a is not an eigenvalue then $c = 0$), and $Q_n(t)$ is a polynomial of degree n with undetermined coefficients.

For example, if I need to solve $x'' - 2x' + x = e^t$ then the rule above implies that $x_p(t) = At^2e^t$, and to determine A I need to plug this solution into the equation and analyze the result.

Exercise 3.38. Solve

$$\begin{aligned} &(a) \quad x'' - x = 2e^t - t^2,\\ &(b) \quad x'' - 3x' + 2x = \sin t,\\ &(c) \quad x'' - 4x' + 5x = e^{2t}\sin^2 t. \end{aligned}$$

3.6.2 The harmonic oscillator

Consider a mass hanging on a spring (see the figure). The position of the mass at time t in uniquely

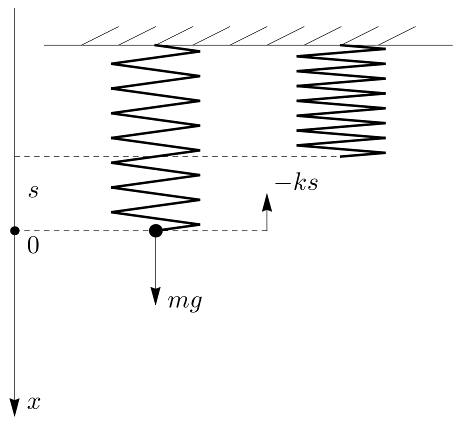

Figure 3.5: A mass on a spring

defined by one coordinate $x(t)$ along the x-axis, whose direction is chosen to be along the direction of the force of gravity. The movement of the mass is determined by the second Newton's law, that can be stated (for this particular one-dimensional case) as

$$ma = \sum F_i,$$

where m is the mass of the object, a is the acceleration, $a = \ddot{x}$, and $\sum F_i$ is the net force applied. The net force includes the gravity $F_1 = mg$, where g is the acceleration due to gravity, ($g \approx 9.8\, m/s^2$ in metric units). The restoring force of the spring is governed by *Hooke's law*, which says that the restoring force in the opposite to the movement direction is proportional to the distance stretched:

$F_2 = -k(x + s)$ if I set the point $x = 0$ at the equilibrium, and s is the length stretched by the mass due to gravity. Note that at the equilibrium ($x = 0$) I must have $mg - ks = 0$. Here minus signifies that the force is acting in the direction opposite to the axis direction. When the mass is not at rest, I can also have damping, which is acting in the direction opposite to the direction of velocity. Observations say that it is reasonable to assume that the damping is proportional to the speed, when x is small enough, hence $F_3 = -c\dot{x}$, where c is a constant of proportionality. Finally, I may have that an external force $F_4 = F(t)$ is applied to the mass. Summing,

$$m\ddot{x} = F_1 + F_2 + F_3 + F_4 \implies m\dot{x} = mg - k(s + x) - c\dot{x} + F(t),$$

and finally, after some simplifications and rearrangements:

$$m\ddot{x} + c\dot{x} + kx = F(t), \tag{3.8}$$

which is a second order linear nonhomogeneous ODE with constant coefficients. The initial conditions — the initial position and initial velocity — are

$$x(0) = x_0, \quad \dot{x}(0) = v_0. \tag{3.9}$$

I will consider cases one by one, starting with the simplest one.

Harmonic oscillations. Here I take $c = 0$ and $F(t) \equiv 0$. Hence,

$$m\ddot{x} + kx = 0,$$

or, after using the new notation $w_0^2 = k/m$

$$\ddot{x} + \omega_0^2 x = 0.$$

This equation has the general solution

$$x_h(t) = C_1 \cos \omega_0 t + C_2 \sin \omega_0 t,$$

where C_1, C_2 are arbitrary constants that are determined by the initial conditions (3.9). For the following it will be convenient to rewrite the last expression in a different form. Assuming the at least one of the arbitrary constants is not zero, I have

$$\begin{aligned}
x_h(t) &= C_1 \cos \omega_0 t + C_2 \sin \omega_0 t \\
&= \sqrt{C_1^2 + C_2^2} \left(\frac{C_1}{\sqrt{C_1^2 + C_2^2}} \cos \omega_0 t + \frac{C_2}{\sqrt{C_1^2 + C_2^2}} \sin \omega_0 t \right) \\
&= A(\cos \omega_0 t \cos \varphi + \sin \omega_0 t \sin \varphi) \\
&= A \cos(\omega_0 t - \varphi),
\end{aligned}$$

where instead of the old constants C_1, C_2 I have new constants A and φ, which can be determined by the initial conditions (3.9) and related to the old constants as

$$A = \sqrt{C_1^2 + C_2^2}, \quad \cos \varphi = \frac{C_1}{\sqrt{C_1^2 + C_2^2}}, \quad \sin \varphi = \frac{C_2}{\sqrt{C_1^2 + C_2^2}}.$$

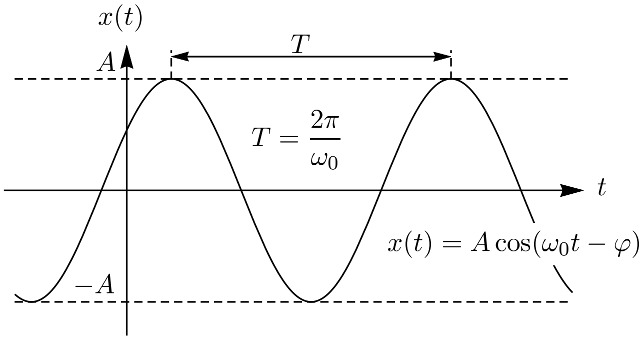

Figure 3.6: Simple harmonic oscillations

The formula

$$x_h(t) = A\cos(\omega_0 t - \varphi)$$

gives a simple way to analyze the displacement $x(t)$ at every time moment t.

Trigonometric functions cos and sin describe periodic oscillations that are called *simple harmonic motion*. Therefore, the original system $\ddot{x} + \omega_0^2 x = 0$ is often called the *simple harmonic oscillator*. Function $\cos\omega_0 t$ has the period

$$T = \frac{2\pi}{\omega_0}.$$

The *frequency* f (number of complete oscillations per time unit, measured usually in hertz ($Hz = 1/s$)) is defined as the reciprocal of the period:

$$f = \frac{1}{T} = \frac{\omega_0}{2\pi},$$

and ω_0 is called the *angular frequency* ($\omega_0 = 2\pi f$, measured in radians per seconds).

Hence I have that the harmonic oscillator produces periodic motion with the angular frequency ω_0. By subtracting φ I simply shift the graph of my function, and this constant is called the *phase*. Finally, the harmonic oscillations are bounded now by A and $-A$, and this constant is called the *amplitude* of oscillations. Therefore, if I am given a simple harmonic oscillator, then its behavior is defined by the angular frequency

$$\omega_0 = \sqrt{\frac{k}{m}},$$

which is the intrinsic property of the system, that is why it is sometimes called the *natural frequency of the system*, and by the amplitude and phase, which can be found given the initial conditions x_0, v_0. Note that the period of oscillations

$$T = 2\pi\sqrt{\frac{m}{k}}$$

does not depend on the initial conditions and hence on the amplitude, which is the property of linear systems. For nonlinear system this does not hold.

Simple harmonic oscillator predicts that the oscillations continue forever, which is not true for the real systems. The reason for this is that I assumed that there was no damping. Now consider the case when $c \neq 0$.

$F(t) \equiv 0$ and $c > 0$. Hence,

$$m\ddot{x} + c\dot{x} + kx = 0.$$

To solve it I write down the characteristic equation

$$m\lambda^2 + c\lambda + k = 0,$$

which can be solved as

$$\lambda_{1,2} = \frac{-c \pm \sqrt{c^2 - 4mk}}{2m}.$$

Here I need to consider 3 cases:

Overdamped motion. Assume that $c^2 - 4mk > 0$, therefore, the characteristic equation has two negative real roots λ_1, λ_2, and the general solution is given by

$$x(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}.$$

Depending on the values of C_1 and C_2 this solution will either never cross zero, or cross it only once. Moreover, since both λ's are negative, the solution approaches zero: $x(t) \to 0$ as $t \to \infty$, which physically means that if the damping is really strong, the mass on a spring will return to its equilibrium position either without or with one oscillation.

Critically damped motion. Let $c^2 - 4mk = 0$, then $\lambda = -c/(2m)$ is the only root of the characteristic polynomial with multiplicity 2. Therefore,

$$x(t) = C_1 e^{\lambda t} + C_2 t e^{\lambda t}.$$

Here the situation is very close to the previous case. Since λ is negative, $x(t) \to 0$ as $t \to \infty$ without oscillations.

Damped oscillations. Let $c^2 - 4mk < 0$, therefore I have two complex conjugate roots $\lambda_1 = \overline{\lambda}_2 = \alpha + \mathrm{i}\beta$, where

$$\alpha = -\frac{c}{2m}, \quad \beta = \frac{\sqrt{4mk - c^2}}{2m} = \sqrt{\omega_0^2 - \left(\frac{c}{2m}\right)^2}.$$

I have

$$x(t) = e^{\alpha t}(C_1 \cos \beta t + C_2 \sin \beta t),$$

or using the approach from the simple harmonic oscillator:

$$x(t) = A e^{\alpha t} \cos(\beta t - \varphi),$$

where A and φ are new arbitrary constants. Note that if I consider $A(t) = Ae^{\alpha t}$ as my "amplitude," then, since $\alpha < 0$, $A(t) \to 0$ as should be expected for damped oscillations. The solution in this case is not periodic, but sometimes called quasiperiodic, because I observe oscillations with decreasing amplitude and the quasi-period given by

$$T = \frac{2\pi}{\beta} = \frac{2\pi}{\sqrt{\omega_0^2 - \left(\frac{c}{2m}\right)^2}},$$

which is larger than the period of simple harmonic oscillations with the angular velocity ω_0, as also should be intuitively expected.

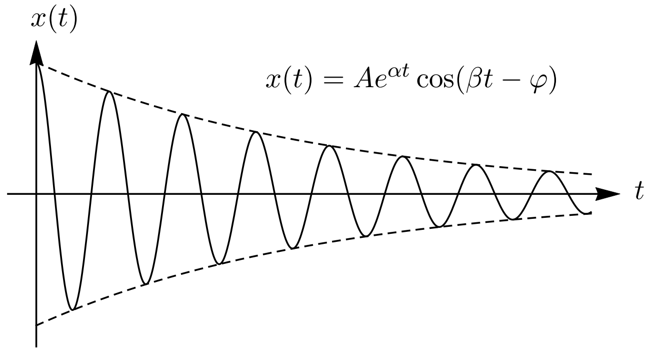

Figure 3.7: Damped oscillations

Exercise 3.39. Sketch the phase portraits in the coordinates $(x, \dot{x})$ for all possible cases for the linear oscillator without an external force.

Now assume that $c = 0$ and $F(t) = F_0 \cos \omega t$, i.e., the external force is a periodic function with amplitude F_0 and angular frequency ω. I have

$$\ddot{x} + \omega_0 x = \frac{F_0}{m} \cos \omega t.$$

The solution to this equation is

$$x(t) = x_h(t) + x_p(t),$$

where $x_h(t)$ is the general solution to the homogeneous equation and $x_p(t)$ is a particular solution to the nonhomogeneous equation. $x_h(t)$ was already found above:

$$x_h(t) = A \cos(\omega_0 t - \varphi).$$

Now, since $\cos \omega t = \operatorname{Re} e^{\mathrm{i}\omega t}$, consider instead the equation (this is an example of *complexification*, moving problem in the complex domain)

$$\ddot{z} + \omega_0 z = \frac{F_0}{m} e^{\mathrm{i}\omega t}.$$

Assume first that $\mathrm{i}\omega$ is not a root of the characteristic polynomial, i.e., $\omega \neq \omega_0$. Then

$$z_p(t) = Ce^{\mathrm{i}\omega t} \implies C = \frac{F_0}{m(\omega_0^2 - \omega^2)}.$$

Therefore,

$$x_p(t) = \operatorname{Re} z_p(t) = \frac{F_0}{m(\omega_0^2 - \omega^2)} \cos \omega t,$$

and the general solution is

$$x(t) = A \cos(\omega_0 t - \varphi) + \frac{F_0}{m(\omega_0^2 - \omega^2)} \cos \omega t,$$

where A and φ are determined by the initial conditions. Note that the general solution is the sum of two periodic functions with different periods. Will the solution be also periodic? The answer is generally "no," for the general solution to be periodic we have to ask that ω_0/ω is a rational number.

Exercise 3.40. Let

$$f(t) = \cos\omega_0 t + \cos\omega_1 t.$$

Show that f is periodic if and only if $\omega_0/\omega_1 \in \mathbf{Q}$ (it is said that the frequencies are commeasurable in this case).

If the angular frequency of the external force approaches the natural frequency of the system, then $|x_p(t)|$ will grow without bounds. To see this, now let $\omega = \omega_0$. In this case,

$$z_p(t) = Cte^{\mathrm{i}\omega_0 t} \implies C = \frac{F_0}{2m\omega_0\mathrm{i}}\,,$$

and hence

$$x_p(t) = \operatorname{Re} z_p(t) = \frac{F_0}{2m\omega_0} t\sin\omega_0 t,$$

which satisfies $x_p(t) \to \infty$ as $t \to \infty$.

In physics the phenomenon when the amplitude grows without bounds if the natural frequency of the system equals the angular frequency of the external force is called *resonance*.

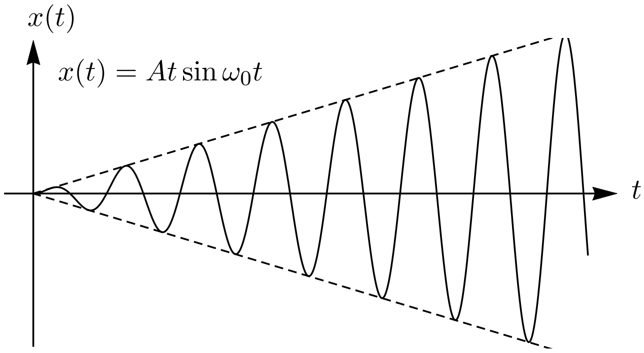

Figure 3.8: Resonance in the system without damping

Consider now

$$m\ddot{x} + c\dot{x} + kx = F_0\cos\omega t.$$

The general solution is given by the sum

$$x(t) = x_h(t) + x_p(t),$$

where $x_h(t)$ was already found above (I assume the damped oscillations occur in the system without external force):

$$x_h(t) = Ae^{\alpha t}\cos(\beta t - \varphi).$$

A particular solution can be found using the same approach as in the case $c = 0$ and is given by

$$x_p(t) = \frac{F_0}{m\big((\omega_0^2 - \omega^2) + (c\omega/m)^2\big)^{1/2}}\cos(\omega t - \phi),$$

where

$$\tan\phi = \frac{c\omega}{k - m\omega^2}\,.$$

Exercise 3.41. Confirm the expression for x_p.

Since $x_h(t) \to 0$ as $t \to \infty$ (this is the *transient* part of the solution), then $x(t) \to x_p(t)$, which is called the *stationary* part of the solution. Hence I conclude that the mass on the spring, when the damping and external periodic force are taken into account, will produce oscillations with the frequency equal the frequency of the external force, and with the amplitude given by

$$\frac{F_0}{m\big((\omega_0^2-\omega^2)+(c\omega/m)^2\big)^{1/2}}\,,$$

which is maximal when (check)

$$\omega^2=\omega_0^2-\frac{c}{2m}\,,$$

provided that $\omega_0^2-c/(2m)>0$. And this value of the angular frequency of the external force is defined to be resonant.

3.7 Non-autonomous linear systems of ODE. General theory

Now I will study the ODE in the form

$$\dot{\boldsymbol{x}}=\boldsymbol{A}(t)\boldsymbol{x}+\boldsymbol{g}(t),\quad \boldsymbol{x}(t)\in\mathbf{R}^k,\quad \boldsymbol{A},\boldsymbol{g}\in\mathcal{C}(I),\tag{3.10}$$

where now the matrix $\boldsymbol{A}$ is time dependent and continuous on some $I\subseteq\mathbf{R}$.

The initial condition is now

$$\boldsymbol{x}(t_0)=\boldsymbol{x}_0,\quad (t_0,\boldsymbol{x}_0)\in I\times\mathbf{R}^k.\tag{3.11}$$

Theorem 3.20. *Let the matrix-function $\boldsymbol{A}$ and the vector-function $\boldsymbol{g}$ be continuous on some interval $I\subseteq\mathbf{R}$. Then the solution to* (3.10), (3.11) *exists, unique and extends to the whole interval I.*

Proof. Problem (3.10), (3.11) satisfies the conditions of the existence and uniqueness theorem. Moreover, since

$$|\boldsymbol{A}(t)\boldsymbol{x}+\boldsymbol{g}(t)|\leq\|\boldsymbol{A}(t)\||\boldsymbol{x}|+|\boldsymbol{g}(t)|\leq L|\boldsymbol{x}|+M,$$

for some $L>0$, $M>0$, therefore, by Corollary 2.39, this solution can be extended to the whole interval I. ■

Note the global character of the theorem.

Together with (3.10) consider the corresponding homogeneous system

$$\dot{\boldsymbol{x}}=\boldsymbol{A}(t)\boldsymbol{x},\quad \boldsymbol{x}(t)\in\mathbf{R}^k,\quad \boldsymbol{A}\in\mathcal{C}(I),\tag{3.12}$$

Exercise 3.42. For the first order linear homogeneous ODE

$$\dot{x}=a(t)x$$

the solution is given by

$$x(t)=x_0e^{\int_{t_0}^t a(\tau)\,\mathrm{d}\tau}.$$

A naive approach would be to solve problem (3.12) by writing

$$\boldsymbol{x}(t) = e^{\int_{t_0}^{t} \boldsymbol{A}(\tau)\,\mathrm{d}\tau}\boldsymbol{x}_0.$$

Consider the matrix

$$\boldsymbol{A}(t) = \begin{bmatrix} 0 & 0 \\ 1 & t \end{bmatrix}$$

and find its solution directly. Also find $e^{\int_0^t \boldsymbol{A}(\tau)\,\mathrm{d}\tau}$ and show that at least in this particular case this formula does not give a solution to the problem.

Explain, what when wrong in this example and give a condition on matrix $\boldsymbol{A}(t)$, $t \in I$ such that the matrix exponent formula would work.

Theorem 3.21 (Principle of superposition)**.**

(a) *If* $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ *solve* (3.12) *then their linear combination* $\alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2$ *also solves* (3.12)*.*

(b) *If* $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ *solve* (3.10) *then their difference* $\boldsymbol{x}_1 - \boldsymbol{x}_2$ *solves* (3.12)*.*

(c) *Any solution to* (3.12) *can be represented as a sum of a particular* (*fixed*) *solution to* (3.10) *and some solution to* (3.12)*.*

Proof. (a) and (b) follow from the linearity of the operator $\frac{\mathrm{d}}{\mathrm{d}t} - \boldsymbol{A}(t)$ acting on the space of continuously differentiable on I vector functions $\boldsymbol{x}\colon I \longrightarrow \mathbf{R}^k$. To show ($c$) fix some solution $\boldsymbol{x}_p$ to (3.10). Assume that arbitrary solution to (3.10) is given by $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_h$ for some function $\boldsymbol{x}_h$. From this, $\boldsymbol{x}_h = \boldsymbol{x} - \boldsymbol{x}_p$ and therefore, due to (b), solves (3.12). ∎

Actually the first point in the last theorem, together with the fact that $\boldsymbol{x} = 0$ solves (3.12), can be restated as: The set of solutions to the homogeneous linear system (3.12) is a vector space. Therefore, it would be nice to figure our what is the dimension of this vector space (in this case any solution can be represented as a linear combination of basis vectors).

Let me first recall the notion of *linear dependence* and *independence* specifically applied to functions and vector functions.

Definition 3.22. *The list of functions* $x_1, \ldots, x_k$ *defined on* $I = (a, b)$ *is called linearly dependent on* I *if there exist scalars* $\alpha_1, \ldots, \alpha_k$*, not equal to zero simultaneously, such that*

$$\alpha_1 x_1(t) + \ldots + \alpha_k x_k(t) \equiv 0, \quad t \in I.$$

If this list of functions is not linear independent then it is called linearly dependent on I*.*

Example 3.23. Consider, e.g., the functions $1, t, t^2, \ldots, t^k$. These functions are linearly independent on any I.

Another example of linearly independent functions on any I is given by $e^{\lambda_1 t}, \ldots, e^{\lambda_k t}$, where all λ_j are distinct.

Exercise 3.43. Prove the statements from the example above.

Exercise 3.44. Decide whether these functions are linearly independent or not:

1.
$$t + 2, \quad t - 2.$$

2.

$$x_1(t) = t^2 - t + 3, \quad x_2(t) = 2t^2 + t, \quad x_3(t) = 2t - 4.$$

3.

$$\log t^2, \quad \log 3t, \quad 7, \quad t \geq 0.$$

4.

$$\sin t, \quad \cos t, \quad \sin 2t.$$

The definition of linear independency verbatim can be used for the vector functions $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ on I (write it down).

Let $(\boldsymbol{x}_j)_{j=1}^k$, $\boldsymbol{x}_j \colon I \longrightarrow \mathbf{R}^k$ be a list of vector functions. The determinant

$$W := \det(\boldsymbol{x}_1 | \ldots | \boldsymbol{x}_k) \colon I \longrightarrow \mathbf{R},$$

is called the *Wronskian*. I have the following important lemma.

Lemma 3.24.

(a) If the Wronskian of $(\boldsymbol{x}_j)_{j=1}^k$ is different from zero at least at one point $t_0 \in I$ then these functions are linearly independent.

(b) If $(\boldsymbol{x}_j)_{j=1}^k$ are linearly dependent then their Wronskian is identically zero on I.

(c) Let $(\boldsymbol{x}_j)_{j=1}^k$ be solutions to linear system (3.12). If their Wronskian is equal to zero at least at one point $t_0 \in I$ then these vector functions are linearly dependent.

Proof. (a) and (b) are the consequences of the standard facts from linear algebra and left as exercises. To show (c), assume that t_0 is such that $W(t_0) = 0$. It means that the linear combination

$$\boldsymbol{x} = \alpha_1 \boldsymbol{x}_1 + \ldots + \alpha_k \boldsymbol{x}_k$$

is such that $\boldsymbol{x}(t_0) = 0$ with not all α_j equal to zero simultaneously. Due to the superposition principle, $\boldsymbol{x}$ solves (3.12) with $\boldsymbol{x}(t_0) = 0$. On the other hand, a vector function $\tilde{\boldsymbol{x}} \equiv 0$ also solves the same problem. Due to the uniqueness theorem $\boldsymbol{x} \equiv \tilde{\boldsymbol{x}}$ and therefore $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ are linearly dependent. ■

Exercise 3.45. Fill in the missed details in the proof above.

Remark 3.25. For arbitrary vector functions statement (c) from the lemma is not true. Consider, e.g.,

$$\boldsymbol{x}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 0 \\ t \end{bmatrix},$$

which are linearly independent; their Wronskian, however, is identically zero.

Lemma 3.26. *Let matrix $\boldsymbol{X} \in \mathcal{C}^{(1)}(I; \mathbf{R}^{k^2})$ be invertible at $t = t_0$. Then at $t = t_0$*

$$\frac{(\det \boldsymbol{X})'}{\det \boldsymbol{X}} = \operatorname{tr}\left(\boldsymbol{X}' \boldsymbol{X}^{-1}\right),$$

where the prime denotes the derivative with respect to t.

Proof. Taylor's formula tells me

$$\boldsymbol{X}(t_0+h)=\boldsymbol{X}(t_0)+h\boldsymbol{X}'(t_0)+o(h),\quad h\to 0.$$

Now calculate the determinant

$$\det \boldsymbol{X}(t_0+h)=\det \boldsymbol{X}(t_0)\det(\boldsymbol{I}+h\boldsymbol{B}+o(h)),$$

where

$$\boldsymbol{B}:=\boldsymbol{X}'(t_0)\boldsymbol{X}^{-1}(t_0).$$

Since, due to Lemma 3.5, $\det(\boldsymbol{I}+h\boldsymbol{B}+o(h))=1+h\operatorname{tr}\boldsymbol{B}+o(h)$, I have

$$\frac{\det \boldsymbol{X}(t_0+h)-\det \boldsymbol{X}(t_0)}{h}=\det \boldsymbol{X}(t_0)(\operatorname{tr}\boldsymbol{B}+o(1)),$$

which proves the lemma. ■

Theorem 3.27 (Liouville's formula or Abel's identity)**.** *Let $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k$ solve* (3.12) *and W be their Wronskian. Then*

$$W(t)=W(t_0)\exp\left(\int_{t_0}^{t}\operatorname{tr}\boldsymbol{A}(\tau)\,\mathrm{d}\tau\right). \tag{3.13}$$

Proof. If $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k$ are linearly dependent, then $W(t)\equiv 0$ and the formula is true. Assume that $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k$ are linearly independent and $\boldsymbol{X}=(\boldsymbol{x}_1|\ldots|\boldsymbol{x}_k)$ be the matrix, whose j-th column is $\boldsymbol{x}_j$. This matrix by construction solves the *matrix* differential equation

$$\dot{\boldsymbol{X}}=\boldsymbol{A}(t)\boldsymbol{X}.$$

From the previous lemma I have

$$\frac{W'(t)}{W(t)}=\operatorname{tr}\left(\boldsymbol{X}'\boldsymbol{X}^{-1}\right)=\operatorname{tr}\left(\boldsymbol{A}(t)\boldsymbol{X}\boldsymbol{X}^{-1}\right)=\operatorname{tr}\left(\boldsymbol{A}(t)\right),$$

which, after integration, implies (3.13). ■

Finally I am ready to prove the main theorem of the theory of linear homogeneous systems of ODE.

Definition 3.28. *A fundamental system of solutions to* (3.12) *is the set of k linearly independent solutions. A fundamental matrix solution is the matrix composed of the fundamental set of solutions:*

$$\boldsymbol{X}=(\boldsymbol{x}_1|\ldots|\boldsymbol{x}_k).$$

Theorem 3.29. *The set of all solutions to* (3.12) *is a vector space of dimension k.*

This theorem basically states that to solve system (3.12) one needs to come up with a fundamental system of solutions, which form the basis of the space of solutions. To find any solution I need to find k (linearly independent) solutions. This is not true for nonlinear systems, and if I know a hundred (or more) of solutions to $\dot{x}=f(t,x)$ it will not help me finding one more solution from those that I have.

Proof. First, I will show that the fundamental system of solutions exists. For this consider k IVPs for (3.12) with

$$\boldsymbol{x}_j(t_0) = \boldsymbol{e}_j, \quad j = 1, \ldots, k,$$

where $\boldsymbol{e}_j \in \mathbf{R}^k$ are the standard unit vectors with 1 at the k-th position and 0 everywhere else. By construction, $W(t_0) \neq 0$ and hence $(\boldsymbol{x}_j)_{j=1}^k$ forms a fundamental system of solutions.

Now consider a solution $\boldsymbol{x}$ to (3.12) with $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$. Since $\boldsymbol{e}_j$ are linearly independent, I have

$$\boldsymbol{x}(t_0) = \alpha_1 \boldsymbol{x}_1(t_0) + \ldots + \alpha_k \boldsymbol{x}_k(t_0).$$

Consider now the function

$$\tilde{\boldsymbol{x}}(t) = \alpha_1 \boldsymbol{x}_1(t) + \ldots + \alpha_k \boldsymbol{x}_k(t),$$

which by the superposition principle solves (3.12) and also satisfies $\tilde{\boldsymbol{x}}(t_0) = \boldsymbol{x}(t_0)$, which, by the uniqueness theorem, implies that $\boldsymbol{x}(t) \equiv \tilde{\boldsymbol{x}}(t)$, which means that any solution can be represented as a linear combination of the solutions in the fundamental system. ■

Corollary 3.30. *If* $\boldsymbol{X}$ *is a fundamental matrix solution, then any solution to* (3.12) *can be represented as*

$$\boldsymbol{x}(t) = \boldsymbol{X}(t)\boldsymbol{\xi}, \quad \boldsymbol{\xi} \in \mathbf{R}^k,$$

where $\boldsymbol{\xi}$ *is an arbitrary constant vector.*

Any two fundamental matrix solutions are related as

$$\boldsymbol{X}(t) = \tilde{\boldsymbol{X}}(t)\boldsymbol{C},$$

where $\boldsymbol{C}$ *is a constant matrix.*

A fundamental matrix solution $\boldsymbol{X}$ satisfying the condition $\boldsymbol{X}(t_0) = \boldsymbol{I}$ is called the *principal matrix solution* (at t_0) and can be found as

$$\boldsymbol{\Phi}(t, t_0) = \boldsymbol{X}(t)\boldsymbol{X}^{-1}(t_0).$$

Using the variation of the constant method, it can be shown that if $\boldsymbol{\Phi}(t, t_0)$ is the principal matrix solution to (3.12) then the general solution to (3.10) with the initial condition (3.11) can be written as

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\boldsymbol{x}_0 + \int_{t_0}^{t} \boldsymbol{\Phi}(t, \tau)\boldsymbol{g}(\tau)\,\mathrm{d}\tau.$$

Exercise 3.46. Prove the last formula.

3.8 Linear k-th order equations with non-constant coefficients

3.8.1 The general theory

Consider a linear k-th order differential equation

$$x^{(k)} + a_{k-1}(t)x^{(k-1)} + \ldots + a_1(t)x' + a_0(t)x = g(t), \tag{3.14}$$

where a_j, g are assumed to be continuous on $I = (a, b)$. Together with (3.14) consider a linear homogeneous equation

$$x^{(k)} + a_{k-1}(t)x^{(k-1)} + \ldots + a_1(t)x' + a_0(t)x = 0, \tag{3.15}$$

and initial conditions

$$x(t_0) = x_0,\, x'(t_0) = x_1, \ldots, x^{(k-1)}(t_0) = x_{k-1}. \tag{3.16}$$

I know that problem (3.14), (3.16) (or (3.15), (3.16)) can be rewritten in the form of a system of k first order equations, and therefore all the previous consideration can be applied. Let me spell them out.

Consider a system of $k-1$ times continuously differentiable functions $x_1, \ldots, x_k$. Their Wronskian is defined as

$$W(t) = \det \begin{bmatrix} x_1(t) & x_2(t) & \ldots & x_k(t) \\ x_1'(t) & x_2'(t) & \ldots & x_k'(t) \\ \vdots & & & \\ x_1^{(k-1)}(t) & x_2^{(k-1)}(t) & \ldots & x_k^{(k-1)}(t) \end{bmatrix}.$$

- If $W(t_0) \neq 0$ then $(x_j)_{j=1}^k$ are linearly independent.
- Let $x_1, \ldots, x_k$ be solutions to (3.15). If $W = 0$ at least at one point then these solutions are linearly dependent.
- Consider vector functions $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ with components $(x_j, x_j', \ldots, x_j^{(k-1)})$, $1 \leq j \leq k$. Then $(x_j)_{j=1}^k$ and $(\boldsymbol{x}_j)_{j=1}^k$ are linearly dependent or independent simultaneously.
- The set of solutions to (3.15) is a vector space of dimension k. The set of k linearly independent solutions to (3.15) is called the fundamental system of solutions.
- If W is the Wronskian of the solutions $x_1, \ldots, x_k$ then I have Liouville's formula

$$W(t) = W(t_0) \exp\left(-\int_{t_0}^{t} a_{k-1}(\tau)\,\mathrm{d}\tau\right).$$

- Using the formula for a particular solution to the nonhomogeneous system, I can write an explicit solution to (3.15), details are left as an exercise.

Exercise 3.47. Provide proofs for all the statements above.

3.8.2 Examples

Here I will discuss a few approaches of analysis of linear ODE, which can be used for specific equations.

Example 3.31 (Second order equation)**.** Consider

$$x'' + a(t)x' + b(t)x = 0.$$

If x_1, x_2 solve this equation then

$$W(t) = \begin{vmatrix} x_1(t) & x_2(t) \\ x_1'(t) & x_2'(t) \end{vmatrix}$$

and Liouville's formula takes the form

$$\begin{vmatrix} x_1(t) & x_2(t) \\ x_1'(t) & x_2'(t) \end{vmatrix} = C \exp\left(-\int_{t_0}^{t} a(\tau)\, \mathrm{d}\tau\right).$$

Sometimes, if one particular solution is known, the second one can be found through the formula above.

For the special case

$$x'' + q(t)x = 0$$

I have

$$\begin{vmatrix} x_1(t) & x_2(t) \\ x_1'(t) & x_2'(t) \end{vmatrix} = C.$$

Or, after simplification,

$$x_2'(t) - \frac{x_2(t)}{x_1(t)} x_1'(t) = \frac{C}{x_1(t)},$$

which gives for x_2 a linear first order ODE, provided I know x_1.

Exercise 3.48. Two particular solutions

$$y_1(t) = t - 1, \quad y_2(t) = \frac{t^2 - t + 1}{t}$$

are known for the differential equation

$$(t^2 - 2t)y'' + 4(t-1)y' + 2y = 6t - 6.$$

Find the general solution.

Example 3.32 (Solving nonhomogeneous equation)**.** Assume that I need to solve

$$x'' + a(t)x' + b(t)x = f(t),$$

and let x_1, x_2 be a fundamental system of solutions to the homogeneous equation. Let me look for a solution to the non-homogeneous equation in the form

$$x(t) = c_1(t)x_1(t) + c_2(t)x_2(t),$$

where c_1, c_2 are unknown functions to be determined.

I have

$$x' = c_1 x_1' + c_2 x_2' + [c_1' x_1 + c_2' x_2].$$

I choose functions c_1, c_2 such that the expression in the square brackets is equal to zero. Then, plugging x into the original equation, I find

$$\begin{aligned} c_1' x_1 + c_2' x_2 &= 0, \\ c_1' x_1' + c_2' x_2' &= f. \end{aligned}$$

Finally, after solving the last system for c_1, c_2, I find a particular solution.

Exercise 3.49. Show that the equation

$$t^2x'' + tx' - x = f(t), \quad t > 0$$

has the general solution

$$x(t) = C_1t + \frac{C_2}{t} + \frac{t}{2}\int_{t_0}^{t}\frac{f(\tau)}{\tau^2}\,\mathrm{d}\tau - \frac{1}{2t}\int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau.$$

Hint: to solve the homogeneous equation use the ansatz $x(t) = t^\lambda$ and find λ.

Exercise 3.50. Show that the equation

$$t^2x'' + tx' + x = f(t), \quad t > 0$$

has the general solution

$$x(t) = C_1 \cos\log t + C_2 \sin\log t + \int_{t_0}^{t}\frac{f(\tau)}{\tau}\sin\log\frac{t}{\tau}\,\mathrm{d}\tau.$$

Example 3.33 (Reduction of order)**.** If one non-trivial solution to the homogeneous linear ODE is known then the order of this equation can be reduced by one.

Consider

$$x^{(k)} + a_{k-1}(t)x^{(k-1)} + \ldots + a_1(t)x' + a_0(t)x = 0,$$

and let $x_1 \neq 0$ solves it. Use the substitution $x(t) = x_1(t)v(t)$, where v is a new unknown function. The equation for v takes the form (fill in the details)

$$b_k(t)v^{(k)} + \ldots + b_1(t)v' = 0,$$

and hence another substitution $w = v'$ reduces its order by one.

Exercise 3.51. Solve the equation

$$(1 + t^2)x'' - 2tx' + 2x = 0,$$

if one solution is given by $x_1(t) = t$.

Exercise 3.52. Solve the equation

$$(2t + 1)x'' + 4tx' - 4x = 0.$$

Hint: Look for a solution in the form $x(t) = e^{pt}$.

Exercise 3.53. Similarly, the same trick (reduction of order) can be used to solve systems of linear equations. Solve the system $\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x}$ with

$$\boldsymbol{A}(t) = \begin{bmatrix} t^2 & -1 \\ 2t & 0 \end{bmatrix},$$

if one of the solutions is $\boldsymbol{\phi}_1(t) = (1, t^2)^\top$. *Hint:* make a substitution $\boldsymbol{x}(t) = \boldsymbol{Q}(t)\boldsymbol{y}(t)$, where $\boldsymbol{Q}(t) = \big(\boldsymbol{\phi}_1(t) \mid \boldsymbol{e}_2\big)$, and $\boldsymbol{e}_2 = (0, 1)^\top$.

Exercise 3.54. Functions

$$x_1 = t, \quad x_2 = t^5, \quad x_3 = |t|^5$$

solve the differential equation

$$t^2 x'' + 5tx' + 5x = 0.$$

Are they linearly independent on $(-1, 1)$?

Exercise 3.55. Let y and z be the solutions to

$$y'' + q(t)y = 0, \quad z'' + Q(t)z = 0$$

with the same initial conditions $y(t_0) = z(t_0)$, $y'(t_0) = z'(t_0)$. Assume that $Q(t) > q(t)$, $y(t) > 0$ and $z(t) > 0$ for all $t \in [t_0, t_1]$. Prove that the function

$$\frac{z(t)}{y(t)}$$

is decreasing in $[t_0, t_1]$.

Exercise 3.56. Prove that two solutions to $x'' + p(t)x' + q(t)x = 0$, where $p, q \in \mathcal{C}(I)$, that achieve maximum at the same value $t_0 \in I$ are linearly dependent on I.

Exercise 3.57. Let $x_1(t) = 1$ and $x_2(t) = \cos t$. Come up with a linear ODE, which has these two functions as particular solutions. Try to find an ODE of the least possible order.

Exercise 3.58. Generalize the previous exercise.

3.9 Linear systems with periodic coefficients

In this section I will consider the systems of the form

$$\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x}, \quad \boldsymbol{x}(t) \in \mathbf{R}^k, \tag{3.17}$$

where $\boldsymbol{A}$ is a continuous periodic matrix function, i.e., there exists $T > 0$ such that $\boldsymbol{A}(t) = \boldsymbol{A}(t+T)$ for all t. The fundamental result about such systems belongs to Floquet and can be formulated in the following form.

Theorem 3.34 (Floquet)**.** *If $\boldsymbol{X}$ is a fundamental matrix solution for* (3.17) *then so is $\boldsymbol{\Xi}$, where*

$$\boldsymbol{\Xi}(t) := \boldsymbol{X}(t+T).$$

Corresponding to each such $\boldsymbol{X}$ there exists a periodic nonsingular matrix $\boldsymbol{P}$ with period T, and a constant matrix $\boldsymbol{B}$ such that

$$\boldsymbol{X}(t) = \boldsymbol{P}(t)e^{t\boldsymbol{B}}. \tag{3.18}$$

Proof. I have

$$\dot{\boldsymbol{\Xi}}(t) = \dot{\boldsymbol{X}}(t+T) = \boldsymbol{A}(t+T)\boldsymbol{X}(t+T) = \boldsymbol{A}(t)\boldsymbol{\Xi}(t),$$

which proves that Ξ is a fundamental matrix solution since $\det \Xi(t) = \det \boldsymbol{X}(t+T) \neq 0$. Therefore, there exists a nonsingular matrix $\boldsymbol{C}$ such that

$$\boldsymbol{X}(t+T) = \boldsymbol{X}(t)\boldsymbol{C},$$

and moreover there exists a constant matrix $\boldsymbol{B}$ such that $\boldsymbol{C} = e^{T\boldsymbol{B}}$ (this matrix is called the *logarithm* of $\boldsymbol{B}$ and does not have to be real).

Now define

$$\boldsymbol{P}(t) := \boldsymbol{X}(t)e^{-t\boldsymbol{B}}.$$

Then

$$\boldsymbol{P}(t+T) = \boldsymbol{X}(t+T)e^{-(t+T)\boldsymbol{B}} = \boldsymbol{X}(t)e^{T\boldsymbol{B}}e^{-(t+T)\boldsymbol{B}} = \boldsymbol{X}(t)e^{-t\boldsymbol{B}} = \boldsymbol{P}(t).$$

Since $\boldsymbol{X}(t)$ and $e^{-t\boldsymbol{B}}$ are nonsingular, then $\boldsymbol{P}(t)$ is nonsingular, which completes the proof. ■

Exercise 3.59. Show that if matrix $\boldsymbol{C}$ is nonsingular then there exists matrix $\boldsymbol{B}$, possibly complex, such that $e^{\boldsymbol{B}} = \boldsymbol{C}$.

Remark 3.35. Actually, if $\boldsymbol{A}(t)$ is real and the system $\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x}$ is considered as $2T$-periodic, then it is possible to find $\boldsymbol{P}_1(t)$ and $\boldsymbol{B}_1$ such that $\boldsymbol{P}_1(t+2T) = \boldsymbol{P}_1(t)$, $\boldsymbol{X}(t) = \boldsymbol{P}_1(t)\exp(\boldsymbol{B}_1 t)$ and $\boldsymbol{B}_1$ is real. I will leave a proof of this fact to the reader.

The matrix $\boldsymbol{C}$, which was introduced in the proof, is called the *monodromy* matrix of equation (3.17), the eigenvalues ρ_j of $\boldsymbol{C}$ are called the *characteristic multipliers*, and the quantities λ_j such that

$$\rho_j = e^{\lambda_j T}$$

are called the *characteristic exponents* (or *Floquet exponents*). The imaginary part of the characteristic exponents is not determined uniquely (recall that the exponent has period $2\pi\mathrm{i}$). I can always choose the characteristic exponents such that they coincide with the eigenvalues of $\boldsymbol{B}$.

Exercise 3.60. Carefully note that for different $\boldsymbol{X}$ one will get different $\boldsymbol{C}$. Explain why this does not influence the conclusions of the theorem and the last paragraph.

Exercise 3.61. Show that the change of variables $\boldsymbol{x} = \boldsymbol{P}(t)\boldsymbol{y}$ for the matrix

$$\boldsymbol{P}(t) = \boldsymbol{X}(t)e^{-t\boldsymbol{B}},$$

where $\boldsymbol{X}(t)$ is the principal matrix solution, turns $\boldsymbol{x} = \boldsymbol{A}(t)\boldsymbol{x}$ in a linear system with constant coefficients.

The notion of stability verbatim translates to the linear systems with non-constant coefficients. In particular, it should be clear that the existence of periodic solutions to (3.17) or the stability of this system are both determined by the eigenvalues of $\boldsymbol{B}$, because the Floquet theorem implies that the solutions are composed of products of polynomials in t, $e^{\lambda_j t}$ and T-periodic functions. I can formulate, leaving the details of the proof to the reader, the following

Theorem 3.36. *Consider system*

$$\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x}, \quad \boldsymbol{x}(t) \in \mathbf{R}^k, \quad \boldsymbol{A}(t) = \boldsymbol{A}(t+T), T > 0, \quad \boldsymbol{A} \in \mathcal{C}(\mathbf{R}_+; \mathbf{R}^k \times \mathbf{R}^k), \quad t > 0.$$

(a) *This system is asymptotically stable if and only if all the characteristic multipliers are in modulus less than one.*

(b) *This system is Lyapunov stable if and only if all the characteristic multipliers are in modulus less than or equal to one, and those with one have equal algebraic and geometric multiplicities.*

(c) *This system is unstable if and only if it has a characteristic multiplier with modulus bigger than one, or it has a characteristic multiplier with modulus equal to one and its algebraic multiplicity is strictly bigger than its geometric multiplicity.*

It is usually a very nontrivial problem to determine the characteristic multipliers. Sometimes the following information can of some use.

Since I have, from the equality $\boldsymbol{X}(t+T) = \boldsymbol{X}(t)e^{T\boldsymbol{B}}$, that

$$\det e^{T\boldsymbol{B}} = \frac{\det \boldsymbol{X}(t+T)}{\det \boldsymbol{X}(t)},$$

therefore, due to Liouville's formula,

$$\det e^{T\boldsymbol{B}} = \exp \int_0^T \operatorname{tr} \boldsymbol{A}(\tau)\, \mathrm{d}\tau = \rho_1 \dots \rho_k,$$

and

$$\lambda_1 + \ldots + \lambda_k = \frac{1}{T} \int_0^T \operatorname{tr} \boldsymbol{A}(\tau)\, \mathrm{d}\tau \quad (\operatorname{mod} \frac{2\pi \mathrm{i}}{T}).$$

Example 3.37. Consider problem (3.17) with

$$\boldsymbol{A}(t) = \begin{bmatrix} \frac{1}{2} - \cos t & b \\ a & \frac{3}{2} + \sin t \end{bmatrix}.$$

Since I have that

$$\int_0^{2\pi} \operatorname{tr} \boldsymbol{A}(\tau)\, \mathrm{d}\tau = 4\pi,$$

therefore

$$\lambda_1 + \lambda_2 = 2 > 0,$$

and hence there exists at least one one-parameter family of solutions to this system which becomes unbounded when $t \to \infty$.

Example 3.38. An important and not obvious fact is that the eigenvalues of $\boldsymbol{A}(t)\,, t \in \mathbf{R}$ cannot be used to infer the stability of the system. Consider

$$\boldsymbol{A}(t) = \begin{bmatrix} -1 + \frac{3}{2}\cos^2 t & 1 - \frac{3}{2}\sin t \cos t \\ -1 - \frac{3}{2}\sin t \cos t & -1 + \frac{3}{2}\sin^2 t \end{bmatrix}.$$

Therefore,

$$\lambda_1 + \lambda_2 = -\frac{1}{2}\,.$$

Hence, no conclusion can be made about the stability. I can calculate the eigenvalues of $\boldsymbol{A}(t)$, which, surprisingly, do not depend on t:

$$\mu_{1,2} = (-1 \pm \mathrm{i}\sqrt{7})/4,$$

which both have negative real part. However, as it can checked directly, the solution

$$t \mapsto \begin{bmatrix} -\cos t \\ \sin t \end{bmatrix} e^{t/2}$$

solves the system, and hence the system is unstable.

Example 3.39. Actually, a converse to the previous example is also true. Consider

$$\boldsymbol{A}(t) = \begin{bmatrix} -\frac{11}{2} + \frac{15}{2}\sin 12t & \frac{15}{2}\cos 12t \\ \frac{15}{2}\cos 12t & -\frac{11}{2} - \frac{15}{2}\sin 12t \end{bmatrix}.$$

The eigenvalues can be calculated as 2 and -13. However, the system with this matrix is asymptotically stable, as can be shown by finding the fundamental matrix solution[1].

Unfortunately there exist no general methods to find matrices $\boldsymbol{P}(t)$ and $\boldsymbol{B}$, and whole books are devoted to the analysis of, e.g., *Hill's equation*

$$\ddot{x} + \big(a + b(t)\big)x = 0,$$

where $b(t) = b(t + \pi)$.

Exercise 3.62. Consider the system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x},$$

where $t \mapsto \boldsymbol{A}(t)$ is a smooth T-periodic matrix function, $\boldsymbol{x}(t) \in \mathbf{R}^k$.

1. $k = 1$, $\boldsymbol{A}(t) = f(t)$. Determine $\boldsymbol{P}(t)$ and $\boldsymbol{B}$ in the Floquet theorem. Give necessary and sufficient conditions for the solutions to be bounded as $t \to \pm\infty$ or to be periodic.
2. $k = 2$ and
$$\boldsymbol{A}(t) = f(t)\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$
Determine $\boldsymbol{P}(t)$ and $\boldsymbol{B}$ in the Floquet theorem. Give necessary and sufficient conditions for the solutions to be bounded as $t \to \pm\infty$ or to be periodic.
3. Consider now
$$\boldsymbol{A}(t) = \begin{bmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{bmatrix}.$$
Note that not only $\operatorname{tr}\boldsymbol{A}(t) = 0$ but also all the terms in $t \mapsto \boldsymbol{A}(t)$ have the average zero value through one period. Are the solutions bounded?

Exercise 3.63. Consider a non-homogeneous problem

$$\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x} + \boldsymbol{f}(t),$$

where both $\boldsymbol{A}$ and $\boldsymbol{f}$ are T-periodic. Prove that if the homogeneous system has no T-periodic vanishing solution then the non-homogeneous system has one and only one T-periodic solution.

[1]Wu, M.Y. A note on stability of linear time-varying systems. IEEE Trans. Automatic Control AC-19 (1974), 162

3.10 Appendix

3.10.1 Jordan's normal form of a matrix

The theory of linear autonomous ODE is essentially a part of the standard linear algebra curriculum. A number of theorems become almost obvious as soon as the systems are written in their *Jordan's normal form*. In my teaching experience I did not meet a lot of first year graduate students who would be able to precisely formulate the main theorem about Jordan's normal form. For this reason I decided to include all the details in these notes. In the following I assume that the student is comfortable with the basic notions of linear algebra, such as vector space, subspace, linear independence, basis, dimension, linear operator, eigenvalues and eigenvectors, kernel and image of a linear operator.

Recall that if I have a linear operator $\mathscr{A}\colon V \longrightarrow V$ on a finite dimensional vector space V over the field $\mathbf{R}$ or $\mathbf{C}$ and fix a basis of V, then I can deal (perform calculations) with matrix $\boldsymbol{A}$, which is *a representation* of my operator in the given basis. If I change my basis, matrix $\boldsymbol{A}$ also changes. For the following I will need the basic fact that if $\boldsymbol{A}, \boldsymbol{A}'$ are two different representations of the same linear operator $\mathscr{A}$ with respect to two different bases, these matrices are *similar*, and they are related as $\boldsymbol{A}' = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$, where $\boldsymbol{T}$ is the matrix of the basis change. The general question is how to find such a basis in which my matrix $\boldsymbol{A}$ of linear operator $\mathscr{A}$ has *the simplest form.* Of course I will need to define precisely what is meant by "the simplest form."

I start working exclusively over $\mathbf{C}$, the main reason for which is that, according to the *fundamental theorem of algebra*, the characteristic polynomial of $\boldsymbol{A}\colon \mathbf{C}^k \longrightarrow \mathbf{C}^k$ has the form

$$p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = (-1)^k(\lambda - \lambda_1)^{\alpha_1} \dots (\lambda - \lambda_m)^{\alpha_m}, \quad \alpha_1 + \dots + \alpha_m = k,$$

where $\lambda_j \in \mathbf{C}$ are the distinct eigenvalues of $\boldsymbol{A}$ and α_j are their corresponding *algebraic* multiplicities. The constants

$$\beta_j = \dim\ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})$$

are called *geometric* multiplicities of λ_j (recall that $\ker(\boldsymbol{A} - \lambda_j\boldsymbol{I})$ is the subspace of $\mathbf{C}^k$ composed of vectors $\boldsymbol{x} \in \mathbf{C}^k$ for which $(\boldsymbol{A} - \lambda_j\boldsymbol{I})\boldsymbol{x} = 0$, this subspace is called *the eigenspace* of λ_j; according to the definition of eigenvectors these subspaces are never *trivial*).

Exercise 3.64. Show that the characteristic polynomial does not depend on a specific representation $\boldsymbol{A}$ and hence it is correct to talk about the characteristic polynomial of operator $\mathscr{A}$.

Exercise 3.65. Show that $1 \leq \beta_j \leq \alpha_j$ for all j.

If I assume that $\beta_j = \alpha_j$ for all j (in particular this is true when $\alpha_j = 1$ for all j) then the answer to my main question is immediate, because I will be able to find a basis in which $\boldsymbol{A}$ is diagonal, and it is hardly questionable that *to be diagonal* is a very convenient property of a matrix to deal with, hence here I assume that "the simplest" means "diagonal." Since this is an important case, I define $\boldsymbol{A}$ to be *semisimple* (from now on I will not distinguish between operator $\mathscr{A}$ and its matrix representation $\boldsymbol{A}$ unless explicitly stated) if $\beta_j = \alpha_j$ for all j.

Theorem 3.40. *Let $\boldsymbol{A}$ be semisimple. Then there exists a basis of $\mathbf{C}^k$ is which $\boldsymbol{A}$ is diagonal.*

Proof. Since $\boldsymbol{A}$ is semisimple and the eigenvectors corresponding to different eigenvalues are linearly independent, as my basis I can take all the linearly independent eigenvectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots$ corresponding to λ_1, λ_2, etc. Since I will end up with exactly k linearly independent eigenvectors $\boldsymbol{T} = [\boldsymbol{u}_1 \mid \ldots \mid \boldsymbol{u}_k]$, they form a basis of $\mathbf{C}^k$ and by construction (check it)

$$\boldsymbol{A}\boldsymbol{T} = \boldsymbol{T}\boldsymbol{J},$$

where $\boldsymbol{J}$ is diagonal, with the eigenvalues of $\boldsymbol{A}$ on the main diagonal, and each eigenvalue is repeated according to its algebraic multiplicity. Since $\boldsymbol{T}$ is invertible, I end up with

$$\boldsymbol{J} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T},$$

i.e., $\boldsymbol{A}$ is similar to the diagonal matrix $\boldsymbol{J}$, which concludes the proof. ■

Theorem above indicates that in the case when $\boldsymbol{A}$ is not semisimple, the things become more involved.

Recall that we say that finite dimensional vector space V is the direct sum of its subspaces $V_1, \ldots, V_m$ if for any $\boldsymbol{v} \in V$ it can be written as the sum of vectors from $V_1, \ldots, V_m$, i.e.,

$$\boldsymbol{v} = \boldsymbol{v}_1 + \ldots + \boldsymbol{v}_m, \quad \boldsymbol{v}_j \in V_j,$$

and this representation is *unique*. The standard notation is

$$V = V_1 \oplus \ldots \oplus V_m.$$

For instance, if all eigenvalues of $\boldsymbol{A}$ are distinct, $\lambda_1, \ldots, \lambda_k$ (which means that $\alpha_j = \beta_j = 1$ for all $j = 1, \ldots, k$), then

$$\mathbf{C}^k = \ker(\boldsymbol{A} - \lambda_1 \boldsymbol{I}) \oplus \ldots \oplus \ker(\boldsymbol{A} - \lambda_k \boldsymbol{I}).$$

Similar equality holds in the semisimple case, but in general ($\beta_j < \alpha_j$ for at least one j) I cannot represent $\mathbf{C}^k$ as a direct sum of eigenspaces. I will need something that I will call *generalized eigenspaces.*

In the following I consider polynomials of matrices. To wit, for complex polynomial

$$p(z) = a_m z^m + \ldots + a_1 z + a_0$$

the expression $p(\boldsymbol{A})$ means $a_m \boldsymbol{A}^m + \ldots + a_1 \boldsymbol{A} + a_0 \boldsymbol{I}$.

Lemma 3.41. *For any $\boldsymbol{A}$ there is non-zero p such that*

$$p(\boldsymbol{A}) = 0.$$

Exercise 3.66. Prove the lemma. *Hint:* Construct p explicitly.

Assume now that $p_{\min}$ is a monic (i.e., its leading coefficient is 1) polynomial of minimum degree such that $p_{min}(\boldsymbol{A}) = 0$. Taking into account the long division algorithm for polynomials, I have

Lemma 3.42. *Let $p(\boldsymbol{A}) = 0$ and let $p_{min}(\boldsymbol{A}) = 0$ be a monic polynomial of minimal degree. Then p_{min} is unique and*

$$p(z) = q(z) p_{min}(z)$$

for some polynomial q for which $q(\boldsymbol{A}) \neq 0$.

Exercise 3.67. Prove this lemma.

As a hint to the previous exercise and to the following discussion I would like to note that polynomials are in many respects similar to integers (of course, they both form *a commutative ring*). In particular, for univariate polynomials over $\mathbf{C}$ it is possible to consider *divisors* (which are simply factors), *greatest common divisor* (gcd) of two non-zero polynomials, and *Bésout's identity*, which I use below.

Due to the proven lemma I define p_{min} to be *the minimal* polynomial of $\boldsymbol{A}$.

Lemma 3.43. *Let $p(z) = p_1(z)p_2(z)$ and $\gcd(p_1, p_2) = 1$ (i.e., polynomials p_1, p_2 are relatively prime). Let $\boldsymbol{A}\colon V \longrightarrow V$. If $p(\boldsymbol{A}) = 0$ then*

$$V = \ker p_1(\boldsymbol{A}) \oplus \ker p_2(\boldsymbol{A}),$$

and each subspace $\ker p_j(\boldsymbol{A}), j = 1, 2$ is invariant under $\boldsymbol{A}$ ($W \subseteq V$ is invariant under $\boldsymbol{A}$ if for all $\boldsymbol{v} \in W$ $\boldsymbol{A}\boldsymbol{v} \in W$).

Proof. First let me prove the invariance. Let $\boldsymbol{v} \in \ker p_j(\boldsymbol{A})$, which means that $p_j(\boldsymbol{A})\boldsymbol{v} = 0$. Now consider $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{v}$. I have $p_j(\boldsymbol{A})\boldsymbol{u} = p_j(\boldsymbol{A})\boldsymbol{A}\boldsymbol{v} = \boldsymbol{A}p_j(\boldsymbol{A})\boldsymbol{v} = \boldsymbol{A}0 = 0$, hence $p_j(\boldsymbol{A})$ is invariant.

Now the assumption $\gcd(p_1, p_2) = 1$ implies (this is *Bésout's identity*) that there exist polynomials q_1, q_2 such that

$$q_1(z)p_1(z) + q_2(z)p_2(z) = 1,$$

or

$$q_1(\boldsymbol{A})p_1(\boldsymbol{A}) + q_2(\boldsymbol{A})p_2(\boldsymbol{A}) = \boldsymbol{I}.$$

Let $\boldsymbol{v} \in \mathbf{C}^k$ be written as (note the order of the vectors in the final sum)

$$\boldsymbol{v} = q_1(\boldsymbol{A})p_1(\boldsymbol{A})\boldsymbol{v} + q_2(\boldsymbol{A})p_2(\boldsymbol{A})\boldsymbol{v} = \boldsymbol{v}_2 + \boldsymbol{v}_1.$$

I claim that $\boldsymbol{v}_j \in \ker p_j(\boldsymbol{A})$ (check it). What is left is to show that there are no other $\boldsymbol{u}_1 \in \ker p_1(\boldsymbol{A}), \boldsymbol{u}_2 \in \ker p_2(\boldsymbol{A})$ such that $\boldsymbol{v} = \boldsymbol{u}_1 + \boldsymbol{u}_2$. I will prove it by contradiction. Indeed, assume that

$$\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{v}_2 = \boldsymbol{u}_1 + \boldsymbol{u}_2,$$

which implies that

$$\boldsymbol{v}_1 - \boldsymbol{u}_1 = \boldsymbol{u}_2 - \boldsymbol{v}_2 = \boldsymbol{w} \in \ker p_1(\boldsymbol{A}) \cap \ker p_2(\boldsymbol{A}),$$

which implies that

$$\boldsymbol{w} = q_1(\boldsymbol{A})p_1(\boldsymbol{A})\boldsymbol{w} + q_2(\boldsymbol{A})p_2(\boldsymbol{A})\boldsymbol{w} = 0 + 0 = 0,$$

that is $\boldsymbol{v}_1 = \boldsymbol{u}_1$ and $\boldsymbol{v}_2 = \boldsymbol{u}_2$. Hence

$$V = \ker p_1(\boldsymbol{A}) \oplus \ker p_2(\boldsymbol{A}),$$

as required. ■

A significant part of all the preliminary work is done, and finally I can see that if

$$p_{min}(z) = (z - \lambda_1)^{l_1} \dots (z - \lambda_m)^{l_m}$$

is the minimal polynomial of $\boldsymbol{A}$, then

Theorem 3.44.
$$V = \ker(\boldsymbol{A} - \lambda_1 \boldsymbol{I})^{l_1} \oplus \ldots \oplus \ker(\boldsymbol{A} - \lambda_m \boldsymbol{I})^{l_m},$$
each $\ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j}$ *is invariant under* $\boldsymbol{A}$*, and* λ_j *are the eigenvalues of* $\boldsymbol{A}$*.*

Proof. The first two statements follow directly from the previous lemma.

Consider $\boldsymbol{A}\colon \ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j} \longrightarrow \ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j}$. Since $\ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j}$ is not trivial (if this was the case p_{min} would not be minimal), $\boldsymbol{A}$ must have an eigenvalue as a linear operator acting on a finite dimensional nontrivial space over $\mathbf{C}$. Let this eigenvalue be λ, i.e., $\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}$, $\boldsymbol{v} \in \ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j}$. I have
$$0 = (\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j} \boldsymbol{v} = (\lambda - \lambda_j)^{l_j} \boldsymbol{v},$$
and hence $\lambda = \lambda_j$, i.e., the roots of the minimal polynomial are the eigenvalues of $\boldsymbol{A}$. In the other direction, assume that λ is an eigenvalue of $\boldsymbol{A}$ with eigenvector $\boldsymbol{v}$. I have
$$0 = p_{min}(\boldsymbol{A})\boldsymbol{v} = (\lambda - \lambda_1)^{l_1} \ldots (\lambda - \lambda_m)^{l_m} \boldsymbol{v},$$
so every eigenvalue must also be a root of p_{min}. ■

The subspace $\ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j}$ is called the *generalized eigenspace* corresponding to λ_j, and its elements are called *generalized eigenvectors*. It is important to understand that constant l_j is the smallest such constant a that $(\boldsymbol{A} - \lambda_j \boldsymbol{I})^a$ vanishes on *any* element of $\ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})$.

Exercise 3.68. Prove the last claim. *Hint:* assume that $(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{a-1}$ vanishes on all the elements and reach a contradiction.

Now, if I select a basis for each generalized eigenspace (carefully note that constants l_j are *not* the numbers of the linearly independent vectors in each *generalized* eigenspace) and put all these bases together, I must get a basis of $V = \mathbf{C}^k$. Since each generalized eigenspace is invariant under $\boldsymbol{A}$, the matrix for it in this basis will have the block diagonal form
$$\begin{bmatrix} \boldsymbol{A}_1 & & & & \\ & \boldsymbol{A}_2 & & & \\ & & \boldsymbol{A}_3 & & \\ & & & \ddots & \\ & & & & \boldsymbol{A}_m \end{bmatrix},$$
where each $\boldsymbol{A}_j$ is a square matrix of dimension $\dim \ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j}$.

Exercise 3.69. Prove the last claim (that the matrix $\boldsymbol{A}$ in this basis is block-diagonal).

What is left is to choose some "simplest" form for each $\boldsymbol{A}_j$.

Now, since each $\ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j}$ is invariant I can consider
$$\boldsymbol{A}\colon \ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j} \longrightarrow \ker(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{l_j},$$
i.e., restrict my operator $\boldsymbol{A}$ on only one generalized subspace. This restriction implies that $\boldsymbol{A}$ has only one eigenvalue λ and the minimal polynomial is $p(z) = (z - \lambda)^l$, where l is the smallest integer that $(\boldsymbol{A} - \lambda \boldsymbol{I})^l = 0$. Denoting $\boldsymbol{N} = \boldsymbol{A} - \lambda \boldsymbol{I}$, I have $\boldsymbol{N}^l = 0$. Such operator $\boldsymbol{N}$ (and its matrix) are called *nilpotent.*

It is possible to have two cases.

Fist, let me deal with the simple case that $l = k$, which means that my minimal polynomial coincides with the characteristic polynomial. By assumption, there is $\boldsymbol{u}$ such that $\boldsymbol{N}^{l-1}\boldsymbol{u} \neq 0$. It follows that $\{\boldsymbol{N}^{l-1}\boldsymbol{u}, \dots, \boldsymbol{N}\boldsymbol{u}, \boldsymbol{u}\}$ are linearly independent and therefore form a basis of V. The matrix for $\boldsymbol{N}$ in this basis is

$$\begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix}.$$

Exercise 3.70. Prove the last statement.

It follows that matrix $\boldsymbol{A}$ has the form in this basis

$$\begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{bmatrix}.$$

Such a matrix is called *Jordan's block.*

In general, set of non-zero vectors $\{\boldsymbol{u}, \boldsymbol{N}\boldsymbol{u}, \dots, \boldsymbol{N}^{l-1}\boldsymbol{u}\}$ with $\boldsymbol{N}^{l}\boldsymbol{u} = 0$ is called *Jordan's chain.*

To deal with the remaining second case $l < k$ I have

Lemma 3.45. *For any finite dimensional vector space V there exists a basis of V consisting of Jordan's chains.*

Idea of a proof. One can use induction on the dimension of V. It is clearly true for $k = 1$. Assume it is true for all vector spaces of dimension $k-1$ and below. We need to show that it is also true for k. Note that since $\boldsymbol{N}$ is nilpotent it is not injective and therefore $\dim \operatorname{im} \boldsymbol{N} < k$ ($\operatorname{im} \boldsymbol{A} = \{\boldsymbol{y} \in V\colon$ there is $\boldsymbol{x} \in V,\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\}$, the *image* of $\boldsymbol{A}$). By the induction hypothesis one therefore can find a basis for $\operatorname{im} \boldsymbol{N}$ consisting of Jordan's chains. Let $\boldsymbol{u}, \boldsymbol{N}\boldsymbol{u}, \dots, \boldsymbol{N}^{l-1}\boldsymbol{u}$ be one such chain. Since $\boldsymbol{u} \in \operatorname{im} \boldsymbol{N}$, there is $\boldsymbol{v} \in V$ such $\boldsymbol{N}\boldsymbol{v} = \boldsymbol{u}$. The same is true for other possible Jordan's chains. In words, I extend each of them by 1. I claim that vectors from all Jordan's chains of the form $\boldsymbol{v}, \boldsymbol{N}\boldsymbol{v}, \dots, \boldsymbol{N}^{l}\boldsymbol{v}$ are linearly independent. I may still have not enough vectors for a basis of V, but I can always add vectors to my set of Jordan's chains to form a basis. Let $\boldsymbol{w}$ be such a vector. Then there is a vector $\boldsymbol{p}$ in the span of vectors from Jordan's chains that $\boldsymbol{N}\boldsymbol{w} = \boldsymbol{N}\boldsymbol{p}$ (because $\boldsymbol{N}\boldsymbol{w}$ is in the span of the image of $\boldsymbol{N}$). It follows that $\boldsymbol{q} = \boldsymbol{w} - \boldsymbol{p} \in \ker \boldsymbol{N}$, and I add all such linearly independent q to my set of Jordan's chains. The resulting collection of vectors is a basis of V. ∎

Exercise 3.71. Fill in all the details missing in the proof above.

We actually can show even more. As before, let me consider only the case of one eigenvalue λ, and nilpotent operator $\boldsymbol{N} = \boldsymbol{A} - \lambda \boldsymbol{I}$. From the theorem above I know that matrix $\boldsymbol{N}$ in the constructed basis has the block-diagonal form (with a number of possible blocks of size 1×1), where each block is a Jordan's block.

Lemma 3.46. *The number and the sizes of the Jordan's block are unique.*

Proof. I have that $\boldsymbol{N}$ is an $k \times k$ matrix. Assume that $\dim\ker \boldsymbol{N} = \beta$, i.e., the geometric multiplicity of eigenvalue λ. Then it implies that I must have exactly β Jordan's blocks. Let $\beta(1)$ be the number of blocks of size 1. Then $\dim\ker \boldsymbol{N}^2$ must differ from $\dim\ker \boldsymbol{N}$ by $\beta - \beta(1)$, and so on, so I get

$$\begin{aligned}\dim\ker \boldsymbol{N} &= \beta,\\ \dim\ker \boldsymbol{N}^2 &= \dim\ker \boldsymbol{N} + \beta - \beta(1) = 2\beta - \beta(1),\\ &\dots\\ \dim\ker \boldsymbol{N}^{l+1} &= \dim\ker \boldsymbol{N}^l + \beta - \beta(1) - \ldots - \beta(l),\end{aligned}$$

where $\beta(l)$ is the number of blocks of size $l \times l$. Hence all $\beta(l)$ are defined uniquely. ∎

Now I see that algebraic multiplicity of eigenvalue λ_j is the sum of the lengthes of all Jordan's chains corresponding to this eigenvalue, its geometric multiplicity is the number of the corresponding Jordan's chains, and finally the power l_j in the minimal polynomial p_{min} corresponding to λ_j is the length of the longest Jordan's chain corresponding to λ_j. Note that even if I know all three constants, I still have multiple choices for the sizes of Jordan's blocks in general, and therefore, in most non-trivial cases I will have to explicitly calculate dimensions of subspaces $\ker \boldsymbol{N}^l$ for different l.

In summary, I have proved

Theorem 3.47 (Jordan's normal form)**.** *Any complex matrix* $\boldsymbol{A}$ *is similar to the matrix in a block diagonal form, each block of which is a Jordan's block, and the number and sizes of these blocks are unique.*

Since in the notation above $l_j \leq \alpha_j$ (where α_j is the algebraic multiplicity of λ_j) then $\deg p_{min} \leq \deg p_{char}$, the minimal polynomial divides the characteristic polynomial, and therefore

Theorem 3.48 (Cayley–Hamilton)**.** *Let* $p_{char}(z) = \det(\boldsymbol{A} - z\boldsymbol{I})$ *be the characteristic polynomial of* $\boldsymbol{A}$. *Then* $p_{char}(\boldsymbol{A}) = 0$.

Example 3.49. Let

$$\boldsymbol{A} = \begin{bmatrix} 2 & 2 & 3\\ 1 & 3 & 3\\ -1 & -2 & -2\end{bmatrix}.$$

Direct calculations yield that $p(z) = (z-1)^3$, that is I have one eigenvalue of algebraic multiplicity 3. Note that it is possible to have three different Jordan's normal forms in this case. I calculate $\boldsymbol{A} - \lambda\boldsymbol{I}$ and find that this matrix has rank 1, that is $\dim\ker(\boldsymbol{A} - \boldsymbol{I}) = 2$, which implies that geometric multiplicity is 2 and I must have two Jordan's blocks. Due to small dimension the only choice for the sizes is 2 and 1 and hence the Jordan's normal form is

$$\begin{bmatrix} 1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}.$$

I invite the students explicitly compute the generalized eigenvectors and find the basis which leads to this form.

Finally, due to the computations above the minimal polynomial is $p_{min}(z) = (z-1)^2$.

Example 3.50. Let

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

The characteristic polynomial is $p(z) = z^4(z-1)$. So I have eigenvalue 1 with algebraic and geometric multiplicity 1 that has 1×1 Jordan's block and eigenvalue 0 of algebraic multiplicity 4. Its geometric multiplicity (check it) is 2, and hence I have two Jordan's chains. It may be the case that one has length 3 and another 1, or 2 and 2. To see it, I calculate $\dim \ker A^2 = 3$, which implies that $\beta(1) = 1$, and hence I have one chain of length 3 and one of length 1. The Jordan's normal form is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Exercise 3.72. Find Jordan's normal form for

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

To finish the discussion on the Jordan's normal form, I recall that I worked over $\mathbf{C}$ and assumed that all my matrices can be complex-valued. While dealing with differential equations, matrix A is often real, but its eigenvalues are not necessarily so. Therefore it would be great to extend the presented results the real realm. The basic fact here is that if λ is a complex eigenvalue of a real matrix A then I must have another complex-conjugate eigenvalue $\bar{\lambda}$. The corresponding eigenvectors are also complex-conjugate.

Instead of going through all the details, I will give a hint how it all works by considering two dimensional example. Specifically,

Lemma 3.51. *Let $A\colon \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ have two complex conjugate eigenvalues $\lambda = \bar{\lambda} = \mu \pm \mathrm{i}\eta$. Then it can always be put in the following real Jordan's form*

$$\begin{bmatrix} \mu & -\eta \\ \eta & \mu \end{bmatrix}.$$

Idea of a proof. By the proven theorem I can always transform my matrix into Jordan's normal form

$$\begin{bmatrix} \mu + \mathrm{i}\eta & 0 \\ 0 & \mu - \mathrm{i}\eta \end{bmatrix}.$$

The last matrix is similar to
$$\begin{bmatrix}\mu & -\eta\\ \eta & \mu\end{bmatrix},$$
since
$$\begin{bmatrix}\mu & -\eta\\ \eta & \mu\end{bmatrix} = \boldsymbol{S}\begin{bmatrix}\mu+\mathrm{i}\eta & 0\\ 0 & \mu-\mathrm{i}\eta\end{bmatrix}\boldsymbol{S}^{-1},$$
for
$$\boldsymbol{S} = \begin{bmatrix}-\mathrm{i} & 1\\ -1 & \mathrm{i}\end{bmatrix},$$
which can be checked directly. Hence the conclusion. ∎

Exercise 3.73. Fill in the missing detail. Specifically you are asked to show that there is *real* invertible matrix $\boldsymbol{P}$ that
$$\begin{bmatrix}\mu & -\eta\\ \eta & \mu\end{bmatrix} = \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}.$$

Finally, I can state

Theorem 3.52. *For any real $\boldsymbol{A}$ there is real $\boldsymbol{P}$ such that*
$$\boldsymbol{J}^{\mathbf{R}} = \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1},$$
where $\boldsymbol{J}^{\mathbf{R}}$ is a block diagonal matrix with Jordan blocks for real eigenvalues $\lambda \in \mathbf{R}$ as before, and with Jordan's blocks
$$\begin{bmatrix}\mu & -\eta & 1 & 0 & & & \\ \eta & \mu & 0 & 1 & & & \\ 0 & 0 & \mu & -\eta & \ddots & & \\ 0 & 0 & \eta & \mu & & 1 & 0\\ & & & & \ddots & 0 & 1\\ & & & & & \mu & -\eta\\ & & & & & \eta & \mu\end{bmatrix}$$
corresponding to the complex eigenvalues $\mu \pm \mathrm{i}\eta \in \mathbf{C}$. The number and the sizes of these blocks are unique.

I will leave the details of the proof as an exercise for a student.

3.10.2 Calculating the matrix exponent

I did not give full details how to compute the matrix exponent for an arbitrary matrix $\boldsymbol{A}$ in the main text. The student can use the previous section to learn how to use Jordan's normal form to calculate $e^{\boldsymbol{A}}$ in the case of multiple eigenvalues. Here I give a brief description of one procedure that is often convenient for the matrices of not very high order and does not require calculating generalized eigenvectors. This procedure (*interpolation method*) actually can be applied to calculate other than exponent functions of matrices[2].

[2] I am borrowing the description and example from Laub, A. J. (2005). Matrix analysis for scientists and engineers. SIAM.

Let $\boldsymbol{A} \in \mathbf{R}^{k\times k}$ and $f(\lambda) = e^{\lambda t}$. My goal is to determine $f(\boldsymbol{A})$. First I need to find the characteristic polynomial p for $\boldsymbol{A}$, $p(\lambda) = \prod_{j=1}^{m}(\lambda - \lambda_j)^{\alpha_j}$, where all λ_j are distinct. Define

$$g(\lambda) = a_0 + a_1\lambda + \ldots + a_{k-1}\lambda^{k-1},$$

where a_j are some constants to be determined. They are, in fact, are the unique solution to k equations:

$$g^{(n)}(\lambda_j) = f^{(n)}(\lambda_j), \quad n = 1, \ldots, \alpha_j, \quad j = 1, \ldots, m.$$

I claim that $f(\boldsymbol{A}) = g(\boldsymbol{A})$, the motivation for this is the *Cayley–Hamilton theorem* (see Theorem 3.48) that says that all powers of $\boldsymbol{A}$ greater than $k-1$ can be expressed as a linear combination of $\boldsymbol{A}^n$, $n = 1, \ldots, k-1$. Thus all the terms of order greater than $k-1$ in the definition of the matrix exponent can be written in terms of these lower powers as well.

Exercise 3.74. Fill in the details in the previous paragraph and prove that g gives the appropriate linear combination (interpolation) for $e^{t\boldsymbol{A}}$.

Example 3.53. Let

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

I find $p(\lambda) = -(\lambda+1)^3$, so $m = 1$ and $\alpha_1 = 3$. I have

$$\begin{aligned} g(-1) = f(-1) &\implies a_0 + a_1 + a_2 = e^{-t}, \\ g'(-1) = f'(-1) &\implies a_1 - 2a_2 = te^{-t}, \\ g''(-1) = f''(-1) &\implies 2a_2 = t^2 e^{-t}. \end{aligned}$$

Solving this system for a_j I get

$$a_2 = \frac{t^2}{2}e^{-t}, \quad a_1 = te^{-t} + t^2e^{-t}, \quad a_0 = e^{-t} + te^{-t} + \frac{t^2}{2}e^{-t},$$

and hence

$$e^{t\boldsymbol{A}} = g(\boldsymbol{A}) = a_0\boldsymbol{I} + a_1\boldsymbol{A} + a_2\boldsymbol{A}^2,$$

which yields

$$\begin{bmatrix} e^{-t} & te^{-t} & 0 \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{-t} \end{bmatrix}.$$

Exercise 3.75. Use the interpolation method to compute

$$e^{t\boldsymbol{A}} = e^{t\begin{bmatrix} -4 & 4 \\ -1 & 0 \end{bmatrix}}.$$

3.10.3 Topological classification of linear flows

3.10.4 More on the implicit function theorem