12 Eigenvectors and eigenvalues. Diagonalization

The basic question remains the same: How to find a basis of our finite dimensional vector space V so that the matrix of a linear operator $\mathscr{A}$, acting on this vector space, takes the *simplest* form. Of course, I need to describe what I deem to be the simplest. At this point of discussion let me use the word "simplest" as a synonym of the word "diagonalizable." Indeed, think of a system of linear equations $\boldsymbol{Ax}=\boldsymbol{b}$ and see that it is very simple to solve this system if $\boldsymbol{A}$ is diagonal.

12.1 Eigenvalues and eigenvectors

The key new notion we will need is the eigenvectors. To motivate their appearance, let me first discuss briefly *invariant subspaces* of a linear operator $\boldsymbol{A}$. (From now on I will stop distinguishing linear operator $\mathscr{A}$ and its matrix $\boldsymbol{A}$ in some fixed bases. This means that I will mostly deal with the vectors spaces $\mathbf{F}^n$, with the field $\mathbf{F}$ being either $\mathbf{R}$ or $\mathbf{C}$.) By definition, a subspace W of the vector space $\mathbf{F}^n$ is called *invariant* with respect to operator $\boldsymbol{A}$ if, abusing notations, $\boldsymbol{A}(W)\subseteq W$. This literally means that for each vector $\boldsymbol{x}\in W$ I have that $\boldsymbol{Ax}\in W$. Now assume that my invariant subspace is one dimensional, with a basis that consists of just one vector $\boldsymbol{v}\in W$. This means that $\boldsymbol{Av}=\lambda\boldsymbol{v}$, for some scalar λ, and for one dimensional invariant subspace the action of my operator is simply "shrinking" or "stretching" along the basis vector $\boldsymbol{v}$. I put quotes because I also consider vector spaces over $\mathbf{C}$ and multiplication by a complex scalar λ is not exactly stretching or shrinking. Now, to describe all one-dimensional invariant subspaces of $\boldsymbol{A}$, I introduce

Definition 12.1. *A non-zero vector $\boldsymbol{v}$ is called an eigenvector of linear operator $\boldsymbol{A}\colon\mathbf{F}^n\longrightarrow\mathbf{F}^n$ if there is a scalar $\lambda\in\mathbf{F}$ such that*

$$\boldsymbol{Av}=\lambda\boldsymbol{v}.$$

This scalar is called the corresponding eigenvalue of $\boldsymbol{A}$.

Note that the eigenvectors are always nonzero, whereas the eigenvalues are allowed to be zero. Also note that the eigenvectors are not defined uniquely, namely, if $\boldsymbol{v}$ is an eigenvector then $\alpha\boldsymbol{v}$ is also an eigenvector for any non-zero scalar α.

Example 12.2. Let

$$\boldsymbol{A}=\begin{bmatrix}3 & 1\\ 0 & 2\end{bmatrix}.$$

I claim that $[1\ \ 0]^\top$ and $[1\ \ -1]^\top$ are two eigenvectors of $\boldsymbol{A}$. Indeed

$$\begin{bmatrix}3 & 1\\ 0 & 2\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}=3\begin{bmatrix}1\\ 0\end{bmatrix},\quad\begin{bmatrix}3 & 1\\ 0 & 2\end{bmatrix}\begin{bmatrix}1\\ -1\end{bmatrix}=2\begin{bmatrix}1\\ -1\end{bmatrix}.$$

Therefore the corresponding eigenvalues are 3 and 2 respectively.

It turns out that if I can come up with a basis of $\mathbf{F}^n$ that consists of eigenvectors then the matrix of my operator in this basis is diagonal. In the language of matrix algebra

Math 329: *Intermediate Linear Algebra* by Artem Novozhilov©
e-mail: artem.novozhilov@ndsu.edu. Fall 2016

Proposition 12.3. *An $n \times n$ matrix $\boldsymbol{A}$ is similar to a diagonal matrix $\boldsymbol{\Lambda}$ if and only if there is a basis of $\mathbf{F}^n$ that consists of eigenvectors of $\boldsymbol{A}$.*

Proof. A proof is almost immediate. Consider matrix $\boldsymbol{P} = [\boldsymbol{v}_1 \mid \ldots \mid \boldsymbol{v}_n]$, where each column is an eigenvector of $\boldsymbol{A}$. Now, by the matrix multiplication, I have

$$\boldsymbol{AP} = [\boldsymbol{Av}_1 \mid \ldots \mid \boldsymbol{Av}_n] = [\lambda_1\boldsymbol{v}_1 \mid \ldots \mid \lambda_n\boldsymbol{v}_n] = \boldsymbol{P\Lambda},$$

where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. ■

The found relation can be written as $\boldsymbol{\Lambda} = \boldsymbol{P}^{-1}\boldsymbol{AP}$, or $\boldsymbol{A} = \boldsymbol{P\Lambda P}^{-1}$.

Please note that the last proposition does not say anything whether such basis exists or how to actually find the eigenvalues and eigenvectors.

I rewrite $\boldsymbol{Av} = \lambda\boldsymbol{v}$ as

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} = 0.$$

That is, the eigenvector $\boldsymbol{v}$ is in the kernel of the operator $\boldsymbol{A} - \lambda\boldsymbol{I}$. Since by definition $\boldsymbol{v} \neq 0$ then the kernel must be different from $\{\boldsymbol{0}\}$. Recall that the kernel is different from $\{\boldsymbol{0}\}$ means that the determinant of this operator is zero (it is said in this case that the operator is *singular*). Therefore, now I can see that the eigenvalues are exactly those scalars that make $\boldsymbol{A} - \lambda\boldsymbol{I}$ singular, and for them I get the condition

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0.$$

To find the corresponding eigenvectors, I need to find a basis of the kernel of $\boldsymbol{A} - \lambda\boldsymbol{I}$ for given λ.

Example 12.4. Let

$$\boldsymbol{A} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}.$$

Then my condition becomes

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \det \begin{bmatrix} 3-\lambda & 2 \\ 1 & 4-\lambda \end{bmatrix} = \lambda^2 - 7\lambda + 10 = (\lambda - 5)(\lambda - 2).$$

Therefore I found two eigenvalues 5 and 2. To find the corresponding eigenvectors, I need to find bases of the kernel of the corresponding operators, or, in somewhat less fancier words, solve two systems $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}$ for $\lambda = 5$ and $\lambda = 2$. For example, for $\lambda = 5$, I have

$$(\boldsymbol{A} - 5\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0} \implies \boldsymbol{v} = [1 \;\; 1]^\top,$$

and for the second eigenvalue I find second eigenvector $\boldsymbol{v} = [-2 \;\; 1]$. It is easy to check that two found eigenvectors linearly independent and hence from a basis of $\mathbf{R}^2$. Therefore, I can diagonalize my operator in this particular case. Do I always need to check that found eigenvectors are linearly independent? Not really.

Proposition 12.5. *Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ be the eigenvectors of $\boldsymbol{A}$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_k \in \mathbf{F}$ of $\boldsymbol{A}$. Then the collection $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is linearly independent.*

Proof. I will prove my proposition by induction on the number of vectors in my collection. First I need to check the base of induction, $k = 1$. And since by definition, the eigenvectors are nonzero, $(\boldsymbol{v}_1)$ is linearly independent.

Now I assume (induction step) that my proposition is true for $k = n - 1$. That is, if I have a linear combination

$$\alpha_1 \boldsymbol{v}_1 + \ldots + \alpha_{n-1} \boldsymbol{v}_{n-1} = 0$$

then I know that all α_j must be zero, by the definition of linear independence.

Take $k = n$ and consider

$$\alpha_1 \boldsymbol{v}_1 + \ldots + \alpha_{n-1} \boldsymbol{v}_{n-1} + \alpha_n \boldsymbol{v}_n = 0.$$

I first multiply this equality by, say, λ_1, and also by $\boldsymbol{A}$. Using the fact that all $\boldsymbol{v}_j$ are eigenvectors I get

$$\begin{aligned} \alpha_1 \lambda_1 \boldsymbol{v}_1 + \ldots + \alpha_{n-1} \lambda_1 \boldsymbol{v}_{n-1} + \alpha_n \lambda_1 \boldsymbol{v}_n &= 0, \\ \alpha_1 \lambda_1 \boldsymbol{v}_1 + \ldots + \alpha_{n-1} \lambda_{n-1} \boldsymbol{v}_{n-1} + \alpha_n \lambda_n \boldsymbol{v}_n &= 0. \end{aligned}$$

Subtracting one from another implies

$$(\lambda_1 - \lambda_2)\alpha_2 \boldsymbol{v}_2 + \ldots + (\lambda_1 - \lambda_n)\alpha_n \boldsymbol{v}_n = 0.$$

Since, by the induction hypotheses, any $n - 1$ eigenvectors are linearly independent, I get that the last equality implies that $(\lambda_1 - \lambda_j)\alpha_j = 0$ for all $j = 2, 3, \ldots, n$. Since by assumption all λ are distinct this means that $\alpha_2 = \ldots = \alpha_n = 0$ and hence I am left with

$$\alpha_1 \boldsymbol{v}_1 = 0,$$

which also implies that $\alpha_1 = 0$ due to $\boldsymbol{v}_1 \neq 0$. Therefore, my eigenvectors are linearly independent. ∎

In the previous example I got that to find the eigenvalues I needed to solve a quadratic equation. Application of any formula for the determinant that we considered implies

Proposition 12.6. *Let $\boldsymbol{A}$ be $n \times n$ matrix. Then its eigenvalues are the roots of the characteristic polynomial*

$$(-1)^n \lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1 \lambda + c_0 := p(\lambda),$$

where $c_j \in \mathbf{F}$.

Moreover, if I define $\operatorname{tr} \boldsymbol{A} = a_{11} + \ldots + a_{nn}$ *be the trace of matrix $\boldsymbol{A}$, then*

$$\begin{aligned} c_{n-1} &= (-1)^{n+1} \operatorname{tr} \boldsymbol{A}, \\ c_0 &= \det \boldsymbol{A}. \end{aligned}$$

If $\boldsymbol{A}$ is a triangular matrix then its diagonal entries are exactly its eigenvalues.

I will leave a proof of this proposition as an exercise.

Now I finally take advantage of the algebraic properties of the field $\mathbf{C}$. I also give *sufficient* (but not necessary!) conditions for a matrix to be diagonalizable.

Proposition 12.7. *Let $\boldsymbol{A}$ be an $n \times n$ matrix with complex entries. Then it has at least one complex eigenvalue. It has exactly n complex eigenvalues if each eigenvalue is counted corresponding to its (algebraic) multiplicity. If the characteristic polynomial of $\boldsymbol{A}$ has n distinct linear factors then $\boldsymbol{A}$ is diagonalizable over $\mathbf{C}$.*

Let $\boldsymbol{A}$ be an $n \times n$ matrix with real entries. If the characteristic polynomial of $\boldsymbol{A}$ has n distinct linear real factors then $\boldsymbol{A}$ is diagonalizable over $\mathbf{R}$.

Remark 12.8. Since in many cases real matrices $\boldsymbol{A}$ do not have any (real) eigenvalues, when one is talking about eigenvalues, they usually mean that the corresponding operator is acting on $\mathbf{C}^n$. Only if it is directly said that we consider only real eigenvalues, you should exclude possible complex roots of the characteristic polynomial from the consideration.

Proof. The proof is immediate. First, since, by the fundamental theorem of algebra, the characteristic polynomial has at least one root, this implies that complex matrix has at least one eigenvalue. The factorization of the characteristic polynomial implies the second claim. Finally, if I have n linear distinct factor it means that I have n distinct eigenvalues, and hence n linearly independent eigenvectors, hence they form a basis, and hence $\boldsymbol{A}$ is diagonalizable. ■

Corollary 12.9. *Let $\lambda_1, \ldots, \lambda_n$ be complex eigenvalues (counting multiplicities) of $\boldsymbol{A}$. Then*

$$\lambda_1 + \ldots + \lambda_n = \operatorname{tr} \boldsymbol{A},$$
$$\lambda_1 \ldots \lambda_n = \det \boldsymbol{A}.$$

Now we can add one more characterization of a square matrix to be not invertible. Recall that $\boldsymbol{A}$ is invertible if and only if $\det \boldsymbol{A} \neq 0$ (and if and only if $\boldsymbol{A}$ is an isomorphism, if and only if $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has only the trivial solution, if and only if $\ker \boldsymbol{A} = \{\boldsymbol{0}\}$, if and only if $\operatorname{im} \boldsymbol{A} = \mathbf{F}^n$, if and only if $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has a unique solution for any $\boldsymbol{b} \in \mathbf{F}^n$). Now I can add that $\det \boldsymbol{A} \neq 0$ if and only if 0 is not an eigenvalue of $\boldsymbol{A}$.

Finally, before proceeding to the examples, I need to check one thing. I talked about operators using matrix $\boldsymbol{A}$ and said that there is no difference. However, the same operator in different bases can have different matrices. So, can we actually talk about eigenvalues of an operator and not simply of eigenvalues of a matrix?

Proposition 12.10. *The characteristic polynomial of a linear operator does not depend on the choice of a basis.*

Proof. Indeed, let $\boldsymbol{A}$ be the matrix of my operator in some basis, and let $p_{\boldsymbol{A}}(\lambda)$ be its characteristic polynomial. Now consider another matrix $\boldsymbol{A}'$ of the same operator. I have that

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

where $\boldsymbol{P}$ is the transition matrix from the new basis to the old one. Consider the characteristic polynomial of $\boldsymbol{A}'$:

$$\begin{aligned} p_{\boldsymbol{A}'}(\lambda) &= \det(\boldsymbol{A}' - \lambda\boldsymbol{I}) = \det(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} - \lambda\boldsymbol{P}^{-1}\boldsymbol{P}) \\ &= \det(\boldsymbol{P}^{-1}(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{P}) = \det \boldsymbol{P}^{-1} \det(\boldsymbol{A} - \lambda\boldsymbol{I}) \det \boldsymbol{P} = \det(\boldsymbol{A} - \lambda\boldsymbol{I}) = p_{\boldsymbol{A}}(\lambda). \end{aligned}$$

■

12.2 Examples

Let me consider a few examples here.

Example 12.11 (Fibonacci numbers)**.** Consider matrix

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

Its eigenvalues are the roots of characteristic polynomial

$$p(\lambda) = \lambda^2 - \lambda - 1,$$

which has the roots

$$\lambda_{1,2} = \frac{1 \pm \sqrt{5}}{2}.$$

Since I have two distinct real eigenvalues, I immediately know that my matrix is diagonalizable over **R**. I find two eigenvectors (the first one corresponds to the plus sign in the eigenvalue, and the second to the minus sign):

$$\boldsymbol{v}_1 = \begin{bmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{bmatrix}, \quad \boldsymbol{P} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{P}^{-1} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & \frac{\sqrt{5}-1}{2} \\ -1 & \frac{1+\sqrt{5}}{2} \end{bmatrix}.$$

My general theory tells me that

$$\boldsymbol{A} = \boldsymbol{P}^{-1} \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \boldsymbol{P}.$$

The reader should actually verify that the last equality is true.

Matrix $\boldsymbol{A}$ is connected to *Fibonacci numbers*, that are defined recursively as

$$F_{n+1} = F_n + F_{n-1}, \quad F_1 = 1,\ F_0 = 0,$$

that is, I set that the first number is zero, the second one is one, and for the rest the n-th number is the sum of $n-1$ and $n-2$. It is easy to see that the first several numbers are

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$$

Can we find a general formula for the n-th number? It turns out that the answer is positive.

First, I introduce new variable $U_n = F_{n-1}$. Now my recursive relation can be written as a sytem

$$\begin{aligned} F_{n+1} &= F_n + U_n, \\ U_{n+1} &= F_n, \end{aligned}$$

or, in the matrix form,

$$\begin{bmatrix} F_{n+1} \\ U_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_n \\ U_n \end{bmatrix}.$$

Now, using notation $\boldsymbol{q}_n = [F_n \ \ U_n]^\top$, I have

$$\boldsymbol{q}_{n+1} = \boldsymbol{A}\boldsymbol{q}_n.$$

Immediately I get that, by iterating the last equality,

$$\boldsymbol{q}_{n+1} = \boldsymbol{A}^{n-1}\boldsymbol{q}_1 = \boldsymbol{A}^{n-1}\begin{bmatrix}1\\0\end{bmatrix}.$$

That is, if I am able to find an arbitrary power of $\boldsymbol{A}$, then I can answer my question. Since $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{\Lambda}\boldsymbol{P}$, I have

$$\boldsymbol{A}^k = (\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1})^k = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}\dots\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{\Lambda}^k\boldsymbol{P}^{-1},$$

and therefore my problem boils down to finding the k-th power of a diagonal matrix. By direct calculations

$$\begin{bmatrix}\frac{1+\sqrt{5}}{2} & 0\\ 0 & \frac{1-\sqrt{5}}{2}\end{bmatrix}^k = \begin{bmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^k & 0\\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^k\end{bmatrix}.$$

Now, if I carefully do all the required computations, I will find that $U_{n+1} = F_n$ is exactly

$$F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n,$$

which gives me a closed formula for the n-th Fibonacci number. (By the way, have you seen the number $\phi = (1+\sqrt{5})/2$ before?)

Example 12.12. Consider now the matrix of the rotation transformation on the plane. Recall that it is

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

The characteristic polynomial is

$$p(\lambda) = \lambda^2 - 2\cos\theta\lambda + 1,$$

and hence

$$\lambda_{1,2} = \frac{2\cos\theta \pm \sqrt{4\cos^2\theta - 4}}{2} = \cos\theta \pm \mathrm{i}\sin\theta = e^{\pm\mathrm{i}\theta},$$

which are complex. That is, for any θ except those that are 0 modulo 2π, the rotation has no real eigenvalues. Therefore, it is not diagonalizable over $\mathbf{R}$. Since it has two (complex conjugate) eigenvalues, it is diagonalizable over $\mathbf{C}$. Let me find its eigenvectors. I start with $\lambda_1 = e^{\mathrm{i}\theta}$. I get

$$\sin\theta\begin{bmatrix}-\mathrm{i} & -1\\ 1 & -\mathrm{i}\end{bmatrix}\boldsymbol{v}_1 = \boldsymbol{0},$$

which has a solution $\boldsymbol{v}_1 = [\mathrm{i}\ \ 1]^\top$ (I assumed that $\sin\theta \neq 0$). Do I need to find another eigenvector? Actually no, since my second eigenvalue is complex conjugate of the first one, it will have a complex conjugate eigenvector $\boldsymbol{v}_2 = [-\mathrm{i}\ \ 1]^\top$. Therefore I have (after some additional calculations) that

$$\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix} = \frac{1}{2\mathrm{i}}\begin{bmatrix}1 & \mathrm{i}\\ -1 & \mathrm{i}\end{bmatrix}\begin{bmatrix}e^{\mathrm{i}\theta} & 0\\ 0 & e^{-\mathrm{i}\theta}\end{bmatrix}\begin{bmatrix}\mathrm{i} & -\mathrm{i}\\ 1 & 1\end{bmatrix}.$$

From the geometric point of view it should be clear why the rotation matrix has no real eigenvalues. It is had a real eigenvalue, it would mean that there is an invariant real one-dimensional subspace subspace, which is geometrically simply a straight line through the origin. By no rotation (except for identity one) leaves any straight line intact, it *rotates* it, hence the conclusion.

Example 12.13. Now consider matrix
$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$
Its characteristic polynomial is $p(\lambda) = (1-\lambda)^2$ and hence there is one real eigenvalue 1 with (algebraic) multiplicity 2. The operator $\boldsymbol{A} - \boldsymbol{I}$ has (check it) one dimensional kernel. That is there is only one linearly independent vector of $\lambda = 1$ and hence our general theory tells us that this matrix is non-diagonalizable.

Example 12.14 (Caley–Hamilton theorem.)**.** Caley–Hamilton theorem is a very important theorem in linear algebra, whose full proof however will require more than we covered in our course. Therefore I will state it but prove only for the case when matrix is diagonalizable.

Theorem 12.15. *Let $\boldsymbol{A}$ be a square matrix. Then it satisfies its own characteristic polynomial, that is*
$$p_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{A}^n + c_{n-1}\boldsymbol{A}^{n-1} + \ldots + c_1\boldsymbol{A} + c_0 = \mathbf{0},$$
where $\mathbf{0}$ is the square zero matrix.

Proof. Once again, while this theorem is true for *any* matrix $\boldsymbol{A}$, I will prove assuming that $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{\Lambda}\boldsymbol{P}$, where $\boldsymbol{\Lambda}$ is the diagonal matrix with the eigenvalues of $\boldsymbol{A}$ on its main diagonal. We have,
$$p_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{P}^{-1}(\boldsymbol{\Lambda}^n + c_{n-1}\boldsymbol{\Lambda}^{n-1} + \ldots + c_1\boldsymbol{\Lambda} + c_0)\boldsymbol{P} = \mathbf{0}$$
as required because inside the parenthesis we have diagonal matrix with the diagonal elements
$$\lambda_k^n + c_{n-1}\lambda_k^{n-1} + \ldots + c_1\lambda_k + c_0,$$
which is zero for any λ_k. ∎

Exercise 1. Give a full proof of Caley–Hamilton theorem for 2×2 matrices.

12.3 Analysis of real 2×2 matrices. Real Jordan normal form

There are two observations which are important to spell out at this point.

First, the theory developed so far works well only if someone finds exactly n distinct (complex) eigenvalues. It turns our that sometimes it is still possible to diagonalize our operator if we have less then n distinct eigenvalues. I decided not to include at this point the exact conditions for this, you can find them in the textbook. I also decided not to answer the question what is the "simplest" possible form of an operator, which cannot be diagonalized. The answer is known, see "Jordan normal form" in any of linear algebra textbooks. The truth is that an honest rigorous coverage of these topics will take the rest of the semester, and there are quite a few important topics left unexplored. Hence, I must sacrifice something.

Second, it should be clear that the theory above works best when we work over $\mathbf{C}$. On the other hand the numbers that surround us are *real*, and hence it would be great to see the "simplest" form of a real operator, which is also real. The general approach to this question still goes through the field $\mathbf{C}$, but after it one needs to go back to the realm of real numbers (for those who likes fancy words: We first complexify our problem and then decomplexify). Again, the general treatment of this question will not be provided.

I decided, however, to cover fully the cases of operators on $\mathbf{R}^2$, to give a glimpse of this general theory.

Theorem 12.16. *Let $\boldsymbol{A}$ be an 2×2 matrix with real entries. Then it is similar to one of the following Jordan real normal forms:*

$$\boldsymbol{J}_1=\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix},\quad \boldsymbol{J}_2=\begin{bmatrix}\lambda & 1\\ 0 & \lambda\end{bmatrix},\quad \boldsymbol{J}_3=\begin{bmatrix}\alpha & \beta\\ -\beta & \alpha\end{bmatrix}.$$

Proof. This is an example of a proof by cases. The characteristic polynomial of my matrix has the form

$$p(\lambda)=\lambda^2-\operatorname{tr}\boldsymbol{A}\,\lambda+\det\boldsymbol{A}.$$

The roots are given by the formula

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

It is possible to have 1) two distinct real roots (if $\big(\operatorname{tr}\boldsymbol{A}\big)^2-4\det\boldsymbol{A}>0$), 2) one real root (algebraic) multiplicity 2 (if $\big(\operatorname{tr}\boldsymbol{A}\big)^2-4\det\boldsymbol{A}=0$), and 3) two distinct complex conjugate roots ($\big(\operatorname{tr}\boldsymbol{A}\big)^2-4\det\boldsymbol{A}<0$). Consider these cases one by one.

1) If I have two distinct real roots it means that I have two linearly independent eigenvectors that form a basis of $\mathbf{R}^2$ and hence $\boldsymbol{A}$ is diagonalizable, and hence similar to $\boldsymbol{J}_1$.

2) Here I must be careful, because two separate cases must be considered. I have only one eigenvalue, but it is still possible to have two linearly independent eigenvectors if $\dim\ker(\boldsymbol{A}-\lambda\boldsymbol{I})=2$. This, however, means that $\boldsymbol{A}-\lambda\boldsymbol{I}=0$ (why?) and therefore $\boldsymbol{A}$ is a diagonal matrix with both diagonal entries equal to λ, hence in this case $\boldsymbol{A}$ is similar (and actually equal to) to $\boldsymbol{J}_1$ with $\lambda_1=\lambda_2$.

Now I assume that $\dim\ker(\boldsymbol{A}-\lambda\boldsymbol{I})=1$ and therefore there is only one linearly independent vector corresponding to λ. In this case my general theory says that it is impossible to diagonalize $\boldsymbol{A}$. Let $\boldsymbol{v}$ be an eigenvector corresponding to λ and consider the problem

$$(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{u}=\boldsymbol{v}$$

for new unknown vector $\boldsymbol{u}$. I claim that this problem always has a nontrivial solution, which, together with $\boldsymbol{v}$, forms a linearly independent collection and hence a basis of $\mathbf{R}^2$ (prove this statement). Now take the matrix $\boldsymbol{P}=[\boldsymbol{v}\mid\boldsymbol{u}]$. It is invertible and, by direct calculations, satisfies

$$\boldsymbol{A}\boldsymbol{P}=\boldsymbol{P}\boldsymbol{J}_2,$$

therefore matrix $\boldsymbol{A}$ is similar to $\boldsymbol{J}_2$.

3) Finally, if I have two complex conjugate eigenvalues $\lambda_1=\bar\lambda_2=\alpha+\beta\mathrm{i}$ with the corresponding eigenvectors $\boldsymbol{v}_1=\bar{\boldsymbol{v}}_2=\boldsymbol{u}_1+\mathrm{i}\boldsymbol{u}_2$, where now $\boldsymbol{u}_1,\boldsymbol{u}_2$ are real vectors. I claim that they are linearly independent and hence form a basis of $\mathbf{R}^2$. Note that

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

implies $\boldsymbol{A}\boldsymbol{u}_1=\alpha\boldsymbol{u}_1-\beta\boldsymbol{u}_2$, $\boldsymbol{A}\boldsymbol{u}_2=\alpha\boldsymbol{u}_2+\beta\boldsymbol{u}_1$.

Consider matrix $\boldsymbol{P}=[\boldsymbol{u}_1\mid\boldsymbol{u}_2]$. Now

$$\boldsymbol{A}\boldsymbol{P}=[\boldsymbol{A}\boldsymbol{u}_1\mid\boldsymbol{A}\boldsymbol{u}_2]=\boldsymbol{P}\boldsymbol{J}_3,$$

which proves the theorem. ∎

Exercise 2. Fill in all the omissions in the above theorem.