Math 4242 Fall 2016 (Darij Grinberg): homework set 8
due: Wed, 14 Dec 2016
[Thanks to Hannah Brand for parts of the solutions.]

Exercise 1. Recall that we defined the multiplication of complex numbers by the rule

$$(a_1, b_1)(a_2, b_2) = (a_1a_2 - b_1b_2, a_1b_2 + a_2b_1).$$

(a) Prove that this multiplication is associative: i.e., that $z_1(z_2z_3) = (z_1z_2)z_3$ for every three complex numbers z_1, z_2, z_3. (Begin by writing z_1 in the form (a_1, b_1), etc.) [5 points]

(b) For any complex number $z = (a, b) = a + bi$, define a **real** matrix W_z by

$$W_z = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Given two complex numbers z_1 and z_2, prove that $W_{z_1z_2} = W_{z_1}W_{z_2}$. [5 points]

Solution to Exercise 1. We begin by proving part **(b)**.

(b) Let z_1 and z_2 be two complex numbers. Write z_1 in the form $z_1 = (a_1, b_1)$. Write z_2 in the form $z_2 = (a_2, b_2)$. Thus,

$$z_1z_2 = (a_1, b_1)(a_2, b_2) = (a_1a_2 - b_1b_2, a_1b_2 + b_1a_2)$$

(by the definition of the product of two complex numbers). Hence, the definition of $W_{z_1z_2}$ yields

$$W_{z_1z_2} = \begin{pmatrix} a_1a_2 - b_1b_2 & a_1b_2 + b_1a_2 \\ -(a_1b_2 + b_1a_2) & a_1a_2 - b_1b_2 \end{pmatrix} = \begin{pmatrix} a_1a_2 - b_1b_2 & a_1b_2 + b_1a_2 \\ -a_1b_2 - b_1a_2 & a_1a_2 - b_1b_2 \end{pmatrix}. \quad (1)$$

On the other hand, we have $z_1 = (a_1, b_1)$. Thus, $W_{z_1} = \begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix}$. Similarly, $W_{z_2} = \begin{pmatrix} a_2 & b_2 \\ -b_2 & a_2 \end{pmatrix}$. Multiplying these two equalities, we obtain

$$W_{z_1}W_{z_2} = \begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix}\begin{pmatrix} a_2 & b_2 \\ -b_2 & a_2 \end{pmatrix} = \begin{pmatrix} a_1a_2 + b_1(-b_2) & a_1b_2 + b_1a_2 \\ (-b_1)a_2 + a_1(-b_2) & (-b_1)b_2 + a_1a_2 \end{pmatrix} = \begin{pmatrix} a_1a_2 - b_1b_2 & a_1b_2 + b_1a_2 \\ -a_1b_2 - b_1a_2 & a_1a_2 - b_1b_2 \end{pmatrix}.$$

Comparing this with (1), we obtain $W_{z_1z_2} = W_{z_1}W_{z_2}$. This solves Exercise 1 **(b)**.

[*Remark:* Of course, we also have $W_{z_1+z_2} = W_{z_1} + W_{z_2}$ and $W_{z_1-z_2} = W_{z_1} - W_{z_2}$ for any two complex numbers z_1 and z_2. These facts, combined, show that the addition, subtraction and multiplication of complex numbers are mirrored by the addition, subtraction and multiplication of their corresponding "W-matrices" (where the *W-matrix* of a complex number z means the 2×2-matrix W_z). In more abstract terms, this says that the map $\mathbb{C} \to \mathbb{R}^{2\times 2}$ sending each complex number z to its W-matrix W_z is a ring homomorphism[1]. This fact is behind our second solution of part **(a)** given below.]

(a) *First solution of part **(a)**:* Here is the straightforward approach:

We defined the multiplication of complex numbers by the rule $(a_1, b_1)(a_2, b_2) = (a_1a_2 - b_1b_2, a_1b_2 + a_2b_1)$.

Given three complex numbers $z_1 = (a_1, b_1)$, $z_2 = (a_2, b_2)$, and $z_3 = (a_3, b_3)$, we can see that $z_1(z_2z_3) = (z_1z_2)z_3$ by explicitly computing both sides of this equation:

$$\begin{aligned} z_1(z_2z_3) &= (a_1, b_1)((a_2, b_2)(a_3, b_3)) \\ &= (a_1, b_1) \cdot (a_2a_3 - b_2b_3, a_2b_3 + a_3b_2) \\ &= (a_1(a_2a_3 - b_2b_3) - b_1(a_2b_3 + a_3b_2), a_1(a_2b_3 + a_3b_2) + (a_2a_3 - b_2b_3)b_1) \\ &= (a_1a_2a_3 - a_1b_2b_3 - a_2b_1b_3 - a_3b_1b_2, a_1a_2b_3 + a_1a_3b_2 + a_2a_3b_1 - b_1b_2b_3) \end{aligned}$$

and

$$\begin{aligned} (z_1z_2)z_3 &= ((a_1, b_1)(a_2, b_2))(a_3, b_3) \\ &= (a_1a_2 - b_1b_2, a_1b_2 + a_2b_1) \cdot (a_3, b_3) \\ &= ((a_1a_2 - b_1b_2)a_3 - (a_1b_2 + a_2b_1)b_3, (a_1a_2 - b_1b_2)b_3 + a_3(a_1b_2 + a_2b_1)) \\ &= (a_1a_2a_3 - a_3b_1b_2 - a_1b_2b_3 - a_2b_1b_3, a_1a_2b_3 - b_1b_2b_3 + a_1a_3b_2 + a_2a_3b_1). \end{aligned}$$

The right hand sides of these two equations are equal (even though the terms appear in slightly different orders in them). Thus, the left hand sides are also equal. In other words, $z_1(z_2z_3) = (z_1z_2)z_3$. This solves part **(a)**.

*Second solution of part **(a)**:* Here is a more elegant proof, using part **(b)**.

Part **(b)** says that $W_{z_1z_2} = W_{z_1}W_{z_2}$ for any two complex numbers z_1 and z_2. Renaming z_1 and z_2 as u and v, we can rewrite this as follows:

$$W_{uv} = W_uW_v \qquad \text{for any two complex numbers } u \text{ and } v. \tag{2}$$

Furthermore, any complex number z can be reconstructed from the 2×2-matrix W_z [2]. Therefore, if u and v are two complex numbers satisfying $W_u = W_v$, then $u = v$.

[1] Strictly speaking, this statement also includes the facts that $W_0 = 0_{2\times 2}$ and $W_1 = I_2$.

[2] *Proof.* Let z be a complex number. Write z in the form (a, b). Then, $W_z = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ (by the definition of W_z). Hence, a and b are the two entries of the first row of the matrix W_z. Therefore, we can reconstruct a and b from W_z. Therefore, we can reconstruct z from W_z (since $z = (a, b)$). Qed.

Now, let z_1, z_2 and z_3 be three complex numbers. Applying (2) to $u = z_1$ and $v = z_2 z_3$, we obtain

$$W_{z_1(z_2z_3)} = W_{z_1} \underbrace{W_{z_2z_3}}_{\substack{=W_{z_2}W_{z_3} \\ \text{(by (2), applied} \\ \text{to } u=z_2 \text{ and } v=z_3)}} = W_{z_1}\left(W_{z_2}W_{z_3}\right). \tag{3}$$

But applying (2) to $u = z_1 z_2$ and $v = z_3$, we obtain

$$W_{(z_1z_2)z_3} = \underbrace{W_{z_1z_2}}_{\substack{=W_{z_1}W_{z_2} \\ \text{(by (2), applied} \\ \text{to } u=z_1 \text{ and } v=z_2)}} W_{z_3} = \left(W_{z_1}W_{z_2}\right)W_{z_3} = W_{z_1}\left(W_{z_2}W_{z_3}\right)$$

(since we know that multiplication of matrices is associative). Comparing this with (3), we obtain $W_{z_1(z_2z_3)} = W_{(z_1z_2)z_3}$.

But recall that if u and v are two complex numbers satisfying $W_u = W_v$, then $u = v$. Applying this to $u = z_1(z_2z_3)$ and $v = (z_1z_2)z_3$, we obtain $z_1(z_2z_3) = (z_1z_2)z_3$ (since $W_{z_1(z_2z_3)} = W_{(z_1z_2)z_3}$). This solves part **(a)** again.

[*Remark:* This second solution illustrates an idea frequently used in algebra: We want to prove that a structure (in our case, the ring $\mathbb{C}$ of complex numbers) satisfies a certain property (in this case, associativity of multiplication). Instead of doing this directly (as was done in the first solution of part **(a)**), we embed the structure in a bigger structure (in our case, the bigger structure is the ring $\mathbb{R}^{2\times 2}$ of 2×2-matrices, and the embedding is the map sending each $z \in \mathbb{C}$ to the matrix W_z) which is already known to possess this property (after all, we know that matrix multiplication is associative); then, we get the property on the smaller structure for free.] □

Here is the algorithm for diagonalizing a matrix we did in class:

Algorithm 0.1. Let $A \in \mathbb{C}^{n\times n}$ be an $n \times n$-matrix. We want to *diagonalize* A; that is, we want to find an invertible $n\times n$-matrix S and a diagonal $n \times n$-matrix Λ such that $A = S\Lambda S^{-1}$. We proceed as follows:

Step 1: We compute the polynomial $\det(A - xI_n)$ (where x is the indeterminate). (This polynomial, or the closely related polynomial $\det(xI_n - A)$, is often called the *characteristic polynomial* of A.)

Step 2: We find the roots of this polynomial $\det(A - xI_n)$. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be these roots **without repetitions** (e.g., multiple roots are **not** listed multiple times), in whatever order you like. For example, if $\det(A - xI_n) = (x-1)^2(x-2)$, then you can set $k = 2$, $\lambda_1 = 1$ and $\lambda_2 = 2$, or you can set $k = 2$, $\lambda_1 = 2$ and $\lambda_2 = 1$, but you must not set $k = 3$.

Step 3: For each $j \in \{1, 2, \ldots, k\}$, we find a basis for $\operatorname{Ker}\left(A - \lambda_j I_n\right)$. (Notice that $\operatorname{Ker}\left(A - \lambda_j I_n\right) \neq \left\{\overrightarrow{0}\right\}$ (because λ_j is a root of $\det\left(A - \lambda_j I_n\right)$, and thus $\det\left(A - \lambda_j I_n\right) = 0$); hence, the basis should consist of at least one vector.)

Step 4: Concatenate these bases into one big list $(s_1, s_2, \ldots, s_m)$ of vectors. If $m < n$, then the matrix A **cannot** be diagonalized, and the algorithm stops here. Otherwise, $m = n$, and we proceed further.

Step 5: Thus, for each $p \in \{1, 2, \ldots, m\}$, the vector s_p belongs to a basis of $\operatorname{Ker}(A - \lambda_j I_n)$ for **some** $j \in \{1, 2, \ldots, k\}$. Denote the corresponding λ_j by μ_p (so that $s_p \in \operatorname{Ker}(A - \mu_p I_n)$). (For example, if s_p belongs to a basis of $\operatorname{Ker}(A - 5I_n)$, then $\mu_p = 5$.) Thus, we have defined m numbers $\mu_1, \mu_2, \ldots, \mu_m$.

Step 6: Let S be the $n \times n$-matrix whose columns are $s_1, s_2, \ldots, s_n$. Let Λ be the diagonal matrix whose diagonal entries (from top-left to bottom-right) are $\mu_1, \mu_2, \ldots, \mu_n$.

(I called these Steps differently in class – the first four steps were called Steps 1.1 to 1.4, while the last two steps were called Part 2. But the above is less confusing.)

Here is an example that is probably too messy for a midterm, but illustrates some things:

Example 0.2. Let $A = \begin{pmatrix} 5 & -1 & 5 \\ 2 & 2 & -4 \\ 1 & -1 & 1 \end{pmatrix}$. Let us diagonalize A. We proceed using Algorithm 0.1:

Step 1: We have $n = 3$ and thus

$$\begin{aligned} \det(A - xI_n) &= \det \begin{pmatrix} 5-x & -1 & 5 \\ 2 & 2-x & -4 \\ 1 & -1 & 1-x \end{pmatrix} \\ &= (5-x)(2-x)(1-x) + (-1)(-4)1 + 5 \cdot 2(-1) \\ &\qquad - (5-x)(-4)(-1) - 5(2-x)1 - (-1)2(1-x) \\ &= -x^3 + 8x^2 - 10x - 24. \end{aligned}$$

Step 2: Now we must find the roots of this polynomial $\det(A - xI_n) = -x^3 + 8x^2 - 10x - 24$.

This is a cubic polynomial, so if it has no rational roots, then finding its roots is quite hopeless (in theory, there is Cardano's formula, but it is so complicated that it is almost useless). Thus, we hope that there is a rational root. To find it, we use the rational root theorem, which says that any rational root of a polynomial with integer coefficients must have the form $\frac{p}{q}$ where p is an integer dividing the constant term and q is a positive integer dividing the leading coefficient. (This is more general than what I quoted in class, and more correct than what I quoted in Section 070.) In our case, the polynomial $-x^3 + 8x^2 - 10x - 24$ has leading coefficient -1 and constant term -24. Thus, any rational root must have the form $\frac{p}{q}$ where p is an integer dividing -24 and q is a positive integer dividing 1. This leaves 16 possibilities for p (namely,

$p \in \{1,2,3,4,6,8,12,24,-1,-2,-3,-4,-6,-8,-12,-24\}$) and 1 possibility for q (namely, $q = 1$). Trying out all of these possibilities, we find that $p = 4$ and $q = 1$ works. Thus, $\dfrac{p}{q} = \dfrac{4}{1} = 4$ is a root.

Hence, we have found one root of our polynomial: namely, $x = 4$. In order to find the others, we divide the polynomial by $x - 4$ (using polynomial long division). We get

$$\frac{-x^3 + 8x^2 - 10x - 24}{x - 4} = -x^2 + 4x + 6.$$

It thus remains to find the roots of $-x^2 + 4x + 6$. This is a quadratic, so we know how to do this. The roots are $2 + \sqrt{10}$ and $2 - \sqrt{10}$.

Thus, altogether, the three roots of $\det(A - xI_n)$ are 4, $2 + \sqrt{10}$ and $2 - \sqrt{10}$. Let me number them $\lambda_1 = 4$, $\lambda_2 = 2 + \sqrt{10}$ and $\lambda_3 = 2 - \sqrt{10}$ (although you can use any numbering you wish).

Step 3: Now, we must find a basis of $\operatorname{Ker}\left(A - \lambda_j I_n\right)$ for each $j \in \{1,2,3\}$. This is a straightforward exercise in Gaussian elimination, and the only complication is that you have to know how to rationalize a denominator (because λ_2 and λ_3 involve square roots). Let me only show the computation for $j = 2$:

Computing $\operatorname{Ker}(A - \lambda_2 I_n)$*:* We have

$$\begin{aligned}\operatorname{Ker}(A - \lambda_2 I_n) &= \operatorname{Ker}\left(A - \left(2 + \sqrt{10}\right) I_n\right) \\ &= \operatorname{Ker}\begin{pmatrix} 3 - \sqrt{10} & -1 & 5 \\ 2 & -\sqrt{10} & -4 \\ 1 & -1 & -1 - \sqrt{10} \end{pmatrix}.\end{aligned}$$

This is the set of all solutions to the system

$$\begin{cases} \left(3 - \sqrt{10}\right) x + (-1)\, y + 5z = 0; \\ 2x + \left(-\sqrt{10}\right) y + (-4)\, z = 0; \\ 1x + (-1)\, y + \left(-1 - \sqrt{10}\right) z = 0 \end{cases}. \qquad (4)$$

So let us solve this system. We divide the first equation by $3 - \sqrt{10}$ (in order to have a simpler pivot entry). This is tantamount to multiplying it by $\dfrac{1}{3 - \sqrt{10}} = -3 - \sqrt{10}$ (this was obtained by rationalizing the denominator, and it is absolutely useful here: you don't want to carry nested fractions around!). It then becomes $x + \left(3 + \sqrt{10}\right) y + \left(-15 - 5\sqrt{10}\right) z = 0$, and the whole system transforms into

$$\begin{cases} x + \left(3 + \sqrt{10}\right) y + \left(-15 - 5\sqrt{10}\right) z = 0; \\ 2x + \left(-\sqrt{10}\right) y + (-4)\, z = 0; \\ 1x + (-1)\, y + \left(-1 - \sqrt{10}\right) z = 0 \end{cases}.$$

Now, subtracting appropriate multiples of the first row from the other two rows, we eliminate x, resulting in the following system:

$$\begin{cases} 1x + \left(3+\sqrt{10}\right) y + \left(-15-5\sqrt{10}\right) z = 0; \\ \left(-6-3\sqrt{10}\right) y + \left(26+10\sqrt{10}\right) z = 0; \\ \left(-4-\sqrt{10}\right) y + \left(14+4\sqrt{10}\right) z = 0 \end{cases}.$$

Next, we divide the second equation by $-6-3\sqrt{10}$ (aka, multiply it by $\dfrac{1}{-6-3\sqrt{10}} = \dfrac{1}{9} - \dfrac{1}{18}\sqrt{10}$), so that it becomes $y + \left(-\dfrac{1}{3}\sqrt{10} - \dfrac{8}{3}\right) z = 0$. Then, subtracting an appropriate multiple of it from the third equation turns the third equation into $0 = 0$. Thus, our system takes the form

$$\begin{cases} x + \left(3+\sqrt{10}\right) y + \left(-15-5\sqrt{10}\right) z = 0; \\ y + \left(-\dfrac{1}{3}\sqrt{10} - \dfrac{8}{3}\right) z = 0; \\ 0 = 0 \end{cases}.$$

In this form, it can be solved by back-substitution (unsurprisingly, there is a free variable, because the kernel is nonzero). The solutions have the form

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \left(\dfrac{4}{3}\sqrt{10} + \dfrac{11}{3}\right) r \\ \left(\dfrac{1}{3}\sqrt{10} + \dfrac{8}{3}\right) r \\ r \end{pmatrix}.$$

Thus,

$$\operatorname{Ker}\left(A - \lambda_2 I_n\right) = \operatorname{span}\left(\begin{pmatrix} \dfrac{4}{3}\sqrt{10} + \dfrac{11}{3} \\ \dfrac{1}{3}\sqrt{10} + \dfrac{8}{3} \\ 1 \end{pmatrix}\right).$$

Hence, $\left(\begin{pmatrix} \dfrac{4}{3}\sqrt{10} + \dfrac{11}{3} \\ \dfrac{1}{3}\sqrt{10} + \dfrac{8}{3} \\ 1 \end{pmatrix}\right)$ is a basis of $\operatorname{Ker}\left(A - \lambda_2 I_n\right)$. (Of course, you can scale the vector by 3 in order to get rid of the denominators.)

Similarly, we can find a basis of $\operatorname{Ker}(A - \lambda_1 I_n)$ (for example, $\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right)$), and a basis of $\operatorname{Ker}(A - \lambda_3 I_n)$ (for example, $\left(\begin{pmatrix} -\frac{4}{3}\sqrt{10} + \frac{11}{3} \\ -\frac{1}{3}\sqrt{10} + \frac{8}{3} \\ 1 \end{pmatrix}\right)$).

Step 4: Now, we concatenate these three bases into one big list $(s_1, s_2, \ldots, s_m)$ of vectors. So this big list is

$$(s_1, s_2, s_3) = \left(\underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}}_{\substack{\text{a basis of} \\ \operatorname{Ker}(A - \lambda_1 I_n)}}, \underbrace{\begin{pmatrix} \frac{4}{3}\sqrt{10} + \frac{11}{3} \\ \frac{1}{3}\sqrt{10} + \frac{8}{3} \\ 1 \end{pmatrix}}_{\substack{\text{a basis of} \\ \operatorname{Ker}(A - \lambda_2 I_n)}}, \underbrace{\begin{pmatrix} -\frac{4}{3}\sqrt{10} + \frac{11}{3} \\ -\frac{1}{3}\sqrt{10} + \frac{8}{3} \\ 1 \end{pmatrix}}_{\substack{\text{a basis of} \\ \operatorname{Ker}(A - \lambda_3 I_n)}} \right).$$

Thus, $m = 3$, so that $m = n$, and thus A can be diagonalized.

Step 5: Since s_1 belongs to a basis of $\operatorname{Ker}(A - \lambda_1 I_n)$, we have $\mu_1 = \lambda_1 = 4$. Similarly, $\mu_2 = \lambda_2 = 2 + \sqrt{10}$ and $\mu_3 = \lambda_3 = 2 - \sqrt{10}$.

Step 6: Now, S is the $n \times n$-matrix whose columns are $s_1, s_2, \ldots, s_n$. In other words,

$$S = \begin{pmatrix} 1 & \frac{4}{3}\sqrt{10} + \frac{11}{3} & -\frac{4}{3}\sqrt{10} + \frac{11}{3} \\ 1 & \frac{1}{3}\sqrt{10} + \frac{8}{3} & -\frac{1}{3}\sqrt{10} + \frac{8}{3} \\ 0 & 1 & 1 \end{pmatrix}.$$

Furthermore, Λ is the diagonal matrix whose diagonal entries (from top-left to bottom-right) are $\mu_1, \mu_2, \ldots, \mu_n$. In other words,

$$\Lambda = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 + \sqrt{10} & 0 \\ 0 & 0 & 2 - \sqrt{10} \end{pmatrix}.$$

These are the S and Λ we were seeking. With some patience, you could check that $A = S \Lambda S^{-1}$ (although it's not necessary to check it).

Remark 0.3. (a) Algorithm 0.1 relies on some nontrivial theorems (for example, Lemma 8.13 in Olver/Shakiban). See §8.3 of Olver/Shakiban for a complete treatment. (Chapter 7 of Lankham/Nachtergaele/Schilling comes close, whereas

Chapter Five.IV of Hefferon is probably overkill.)

(b) What can we do if A is not diagonalizable? The next best thing is the *Jordan normal form* (or *Jordan canonical form*); see §8.6 of Olver/Shakiban.

(c) In Step 4 of Algorithm 0.1, we may sometimes notice that A is not diagonalizable (since $m < n$). Is there a way to notice this earlier, thus saving ourself some useless work?

Yes. For each $j \in \{1, 2, \ldots, k\}$, let α_j be the multiplicity of the root λ_j of the polynomial $\det(A - xI_n)$. (For example, if $\det(A - xI_n) = (x-6)^3 (x+2)$ and $\lambda_1 = 6$, then $\alpha_1 = 3$, because the root $\lambda_1 = 6$ has multiplicity 3.) In Step 3, when computing $\operatorname{Ker}(A - \lambda_j I_n)$, the dimension $\dim(\operatorname{Ker}(A - \lambda_j I_n))$ will be either $= \alpha_j$ or $< \alpha_j$. If it is $< \alpha_j$, then the algorithm is doomed to failure (i.e., you will get $m < n$ in Step 4), and A is not diagonalizable. This can save you some work.

(d) Algorithm 0.1 is more of a theoretical result than an actual workable algorithm; the difficulty of finding exact roots of polynomials, and the instability of Gaussian elimination for non-exact matrices, makes it rather useless. However, for 2×2-matrices it works fine (you can solve quadratics), and it also works nicely for various kinds of "matrices of nice forms" (e.g., you can diagonalize the $n \times n$-matrix $\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ n & n & \cdots & n \end{pmatrix}$ for each n; try it). Practical algorithms for numerical computation are a completely different story. §10.6 of Olver/Shakiban tells the beginnings of the story (namely, how to find eigenvalues, and get something close to diagonalization). Similar to Gaussian elimination, it is wrong to expect diagonalization to work with approximate matrices, because S and Λ can "jump wildly" when A is changed only a little bit; however, certain things can be done that come close to diagonalization.

(e) There is a theorem (called the *spectral theorem*) saying that if A is a **symmetric** matrix with **real** entries, then A is always diagonalizable **over the reals** (i.e., we can find S and Λ with real entries), and moreover you can find an S that is **orthogonal** (i.e., the columns of S are orthonormal). This is a hugely important fact in applications (it is related to the SVD, among many other things), but we will not have the time for it in class. Let me just mention that finding an orthogonal S requires only a simple fix to Algorithm 0.1: In Step 3, you have to choose an **orthonormal** basis of $\operatorname{Ker}(A - \lambda_j I_n)$ (not just some basis). Then, in Step 4, the big list $(s_1, s_2, \ldots, s_m)$ will automatically be an orthonormal basis of $\mathbb{R}^n$. This is one of the miracles of symmetric matrices. See §8.4 in Olver/Shakiban for a proof and more details.

Exercise 2. **(a)** Diagonalize $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$. [5 points]

(b) Diagonalize $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. [5 points]

(c) Diagonalize $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. [10 points]

Solution to Exercise 2. We proceed by using Algorithm 0.1. You have seen this often enough that

(a) We have

$$\det(A - xI_2) = \det \begin{pmatrix} 1-x & 2 \\ 2 & 4-x \end{pmatrix} = (1-x)(4-x) - 2 \cdot 2 = x^2 - 5x.$$

The roots of this polynomial (i.e., the eigenvalues of A) are clearly 0 and 5. We number them as $\lambda_1 = 0$ and $\lambda_2 = 5$.

We now must find bases for $\operatorname{Ker}(A - \lambda_1 I_2)$ and $\operatorname{Ker}(A - \lambda_2 I_2)$. We can do this using the standard Gaussian elimination procedure (you can also see the result directly if you are sufficiently astute), obtaining the basis $\left(\begin{pmatrix} 2 \\ -1 \end{pmatrix} \right)$ for $\operatorname{Ker}(A - \lambda_1 I_2)$ and the basis $\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$ for $\operatorname{Ker}(A - \lambda_2 I_2)$. The big list is therefore $(s_1, s_2) = \left(\begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$. This has size 2, which is our n; hence, the matrix A can be diagonalized. We have $s_1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$, $\mu_1 = \lambda_1 = 0$, $s_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\mu_2 = \lambda_2 = 5$. Therefore, $S = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix}$.

(b) We can take $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$.

One way to solve this is by proceeding exactly as in part **(a)**. Another is to observe that our matrix A is already diagonal, so we can diagonalize it by simply taking $S = I_2$ and $\Lambda = A$.

(c) We can take $S = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

Again, the method is the same as for part **(a)**, but this time we have to solve the cubic equation $x^3 - 3x^2 + 2x = 0$. This is done as follows: The root $x = 0$ is obvious. Leaving this root aside, we can find the other two roots by solving $x^2 - 3x + 2 = 0$; this can be done using the standard formula for the roots of a quadratic. □

Exercise 3. Define a sequence $(g_0, g_1, g_2, \ldots)$ of integers by

$$g_0 = 0, \qquad g_1 = 1, \qquad g_{n+1} = 3g_n + g_{n-1} \qquad \text{for all } n \geq 1.$$

This is similar to the Fibonacci sequence. Here is a partial table of values:

k	0	1	2	3	4	5	6	7	8	9	10
g_k	0	1	3	10	33	109	360	1189	3927		

.

(a) What are g_9 and g_{10} ? [2 points]

(b) Define a 2×2-matrix A by $A = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}$. Find A^2 and A^3. [2 points]

(c) Prove that

$$A^n = \begin{pmatrix} g_{n+1} & g_n \\ g_n & g_{n-1} \end{pmatrix} \qquad (5)$$

for all $n \geq 1$. The proof (or at least the easiest proof) is by induction over n: In the *induction base*, you should check that (5) holds for $n = 1$. In the *induction step*, you assume that (5) holds for $n = m$ for a given positive integer m, and then you have to check that (5) also holds for $n = m + 1$. (Use the fact that $A^{m+1} = AA^m = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix} A^m$.) [10 points]

(d) Diagonalize A. [10 points]

(e) Use this to obtain an explicit formula for g_n. (The formula will involve square roots and n-th powers of numbers, but no recursion and no matrices.) [10 points]

Solution to Exercise 3. **(a)** We have $g_9 = 3g_8 + g_7 = 3 \cdot 3927 + 1189 = 12970$ and $g_{10} = 3g_9 + g_8 = 3 \cdot 12970 + 3927 = 42837$.

(b) We have $A^2 = \begin{pmatrix} 10 & 3 \\ 3 & 1 \end{pmatrix}$ and $A^3 = \begin{pmatrix} 33 & 10 \\ 10 & 3 \end{pmatrix}$. [This is, of course, a particular case of (5).]

(c) We mimic the proof of Proposition 2.48 in the lecture notes:

We shall prove (5) by induction over n:

Induction base: We have $A^1 = A = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}$. Comparing this with

$$\begin{pmatrix} g_{1+1} & g_1 \\ g_1 & g_{1-1} \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix} \qquad (\text{since } g_{1+1} = g_2 = 3,\ g_1 = 1 \text{ and } g_{1-1} = g_0 = 0),$$

we obtain $A^1 = \begin{pmatrix} g_{1+1} & g_1 \\ g_1 & g_{1-1} \end{pmatrix}$. In other words, (5) holds for $n = 1$. This completes the induction base.

Induction step: Let N be a positive integer. (I am calling it N rather than m here, in order to stay closer to the proof of Proposition 2.48 in the lecture notes.) Assume that (5) holds for $n = N$. We must show that (5) also holds for $n = N + 1$.

The definition of the sequence $(g_0, g_1, g_2, \ldots)$ shows that $g_{N+2} = 3g_{N+1} + g_N$ and $g_{N+1} = 3g_N + g_{N-1}$.

We have assumed that (5) holds for $n = N$. In other words,

$$A^N = \begin{pmatrix} g_{N+1} & g_N \\ g_N & g_{N-1} \end{pmatrix}.$$

Now,

$$A^{N+1} = \underbrace{A^N}\quad\underbrace{A} = \begin{pmatrix} g_{N+1} & g_N \\ g_N & g_{N-1} \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} g_{N+1} & g_N \\ g_N & g_{N-1} \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} g_{N+1} \cdot 3 + g_N \cdot 1 & g_{N+1} \cdot 1 + g_N \cdot 0 \\ g_N \cdot 3 + g_{N-1} \cdot 1 & g_N \cdot 1 + g_{N-1} \cdot 0 \end{pmatrix}$$

(by the definition of a product of two matrices)

$$= \begin{pmatrix} 3g_{N+1} + g_N & g_{N+1} \\ 3g_N + g_{N-1} & g_N \end{pmatrix} = \begin{pmatrix} g_{N+2} & g_{N+1} \\ g_{N+1} & g_N \end{pmatrix}$$

(since $3g_{N+1} + g_N = g_{N+2}$ and $3g_N + g_{N-1} = g_{N+1}$). In other words, (5) holds for $n = N + 1$. This completes the induction step; hence, (5) is proven.

(d) We have $A = S\Lambda S^{-1}$, where $S = \begin{pmatrix} 1 & 1 \\ -\frac{1}{2}\sqrt{13} - \frac{3}{2} & \frac{1}{2}\sqrt{13} - \frac{3}{2} \end{pmatrix}$ and $\Lambda = \begin{pmatrix} -\frac{1}{2}\sqrt{13} + \frac{3}{2} & 0 \\ 0 & \frac{1}{2}\sqrt{13} + \frac{3}{2} \end{pmatrix}$. (This can be found using Algorithm 0.1 again.)

(e) Fix $n \in \mathbb{N}$. Let S and Λ be as in the solution to part **(d)**. Then,

$$\Lambda^n = \begin{pmatrix} \left(-\frac{1}{2}\sqrt{13} + \frac{3}{2}\right)^n & 0 \\ 0 & \left(\frac{1}{2}\sqrt{13} + \frac{3}{2}\right)^n \end{pmatrix}$$

(because in order to raise a diagonal matrix to the n-th power, it suffices to raise each diagonal entry to the n-th power). Now, from $A = S\Lambda S^{-1}$, we obtain

$$A^n = S\Lambda^n S^{-1} \qquad \text{(as shown in class)}$$

$$= \begin{pmatrix} 1 & 1 \\ -\frac{1}{2}\sqrt{13} - \frac{3}{2} & \frac{1}{2}\sqrt{13} - \frac{3}{2} \end{pmatrix} \begin{pmatrix} \left(-\frac{1}{2}\sqrt{13} + \frac{3}{2}\right)^n & 0 \\ 0 & \left(\frac{1}{2}\sqrt{13} + \frac{3}{2}\right)^n \end{pmatrix}$$

$$\frac{1}{\sqrt{13}} \begin{pmatrix} \frac{1}{2}\sqrt{13} - \frac{3}{2} & -1 \\ \frac{1}{2}\sqrt{13} + \frac{3}{2} & 1 \end{pmatrix}$$

(since $S = \begin{pmatrix} 1 & 1 \\ -\frac{1}{2}\sqrt{13} - \frac{3}{2} & \frac{1}{2}\sqrt{13} - \frac{3}{2} \end{pmatrix}$, $\Lambda^n = \begin{pmatrix} \left(-\frac{1}{2}\sqrt{13} + \frac{3}{2}\right)^n & 0 \\ 0 & \left(\frac{1}{2}\sqrt{13} + \frac{3}{2}\right)^n \end{pmatrix}$

and $S^{-1} = \dfrac{1}{\sqrt{13}} \begin{pmatrix} \dfrac{1}{2}\sqrt{13} - \dfrac{3}{2} & -1 \\ \dfrac{1}{2}\sqrt{13} + \dfrac{3}{2} & 1 \end{pmatrix}$). If we multiply out this product, we obtain explicit formulas for each of the four entries of A^n. In particular, we obtain the following formula for its $(2,1)$-th entry:

$$(A^n)_{2,1} = \frac{1}{\sqrt{13}} \left(\left(\frac{1}{2}\sqrt{13} + \frac{3}{2} \right)^n - \left(-\frac{1}{2}\sqrt{13} + \frac{3}{2} \right)^n \right).$$

But (5) shows that $(A^n)_{2,1} = g_n$. Hence,

$$g_n = (A^n)_{2,1} = \frac{1}{\sqrt{13}} \left(\left(\frac{1}{2}\sqrt{13} + \frac{3}{2} \right)^n - \left(-\frac{1}{2}\sqrt{13} + \frac{3}{2} \right)^n \right).$$

This is the formula we are looking for. (It is, of course, similar to the Binet formula for the Fibonacci numbers.) □

Exercise 4. Let A be an $n \times n$-matrix. Assume that A can be diagonalized, with $A = S\Lambda S^{-1}$ for an invertible $n \times n$-matrix S and a diagonal $n \times n$-matrix Λ.

(a) Diagonalize A^2. [5 points]

(b) Diagonalize A^{-1}, if A is invertible. (You can use the fact that for an invertible A, the diagonal entries of Λ are nonzero, and so Λ^{-1} is a diagonal matrix again.) [5 points]

(c) Diagonalize A^T (the transpose of A). [10 points]

(The answers should be in terms of S and Λ. For example, $A + I_n$ can be diagonalized as follows: $A + I_n = S(\Lambda + I_n) S^{-1}$. Indeed, S is an invertible matrix, $\Lambda + I_n$ is a diagonal matrix (being the sum of the two diagonal matrices Λ and I_n), and we have

$$S(\Lambda + I_n) S^{-1} = \underbrace{S\Lambda S^{-1}}_{=A} + \underbrace{SI_n}_{=S} S^{-1} = A + \underbrace{SS^{-1}}_{=I_n} = A + I_n.$$

)

Solution to Exercise 4. **(a)** We have $A = S\Lambda S^{-1}$, and thus

$$A^2 = \left(S\Lambda S^{-1} \right)^2 = S\Lambda \underbrace{S^{-1}S}_{=I_n} \Lambda S^{-1} = S \underbrace{\Lambda\Lambda}_{=\Lambda^2} S^{-1} = S\Lambda^2 S^{-1}.$$

The matrix Λ is diagonal. Thus, any power of Λ is diagonal as well (because in order to raise a diagonal matrix Λ to some power, we merely need to raise its diagonal entries to this power). In particular, Λ^2 is diagonal. Therefore, the equality $A^2 = S\Lambda^2 S^{-1}$ provides a diagonalization of A^2.

(b) Assume that A is invertible. Then, it is not hard to see that all diagonal entries of Λ are nonzero[3]. Therefore, the diagonal matrix Λ is invertible, and its inverse Λ^{-1} is obtained by inverting all diagonal entries of Λ. In particular, Λ^{-1} is a diagonal matrix as well. (You can use this fact without proof, but it is helpful to know how it is proven.)

Recall that $(UV)^{-1} = V^{-1}U^{-1}$ for any two invertible $n \times n$-matrices U and V. Applying this to $U = S\Lambda$ and $V = S^{-1}$, we find $\left(S\Lambda S^{-1}\right)^{-1} = \underbrace{\left(S^{-1}\right)^{-1}}_{=S} \underbrace{(S\Lambda)^{-1}}_{=\Lambda^{-1}S^{-1}} = S\Lambda^{-1}S^{-1}$.

We have $A = S\Lambda S^{-1}$, and thus

$$A^{-1} = \left(S\Lambda S^{-1}\right)^{-1} = S\Lambda^{-1}S^{-1}.$$

This equality provides a diagonalization of A^{-1} (since the matrix Λ^{-1} is diagonal).

(c) Proposition 3.18 **(f)** in the lecture notes (applied to S instead of A) shows that the matrix S^T is invertible, and its inverse is $\left(S^T\right)^{-1} = \left(S^{-1}\right)^T$.

Proposition 3.18 **(e)** in the lecture notes shows that any two matrices U and V satisfy $(UV)^T = V^T U^T$ (as long as the product UV is well-defined, i.e., the number of columns of U equals the number of rows of V). Applying this to $U = S\Lambda$ and $V = S^{-1}$, we obtain $\left(S\Lambda S^{-1}\right)^T = \underbrace{\left(S^{-1}\right)^T}_{=\left(S^T\right)^{-1}} \underbrace{(S\Lambda)^T}_{=\Lambda^T S^T} = \left(S^T\right)^{-1} \Lambda^T S^T$.

The matrix Λ its diagonal. Hence, $\Lambda^T = \Lambda$ (because transposing a diagonal matrix does not change it). Now, from $A = S\Lambda S^{-1}$, we obtain

$$A^T = \left(S\Lambda S^{-1}\right)^T = \left(S^T\right)^{-1} \underbrace{\Lambda^T}_{=\Lambda} \underbrace{S^T}_{=\left(\left(S^T\right)^{-1}\right)^{-1}} = \left(S^T\right)^{-1} \Lambda \left(\left(S^T\right)^{-1}\right)^{-1}.$$

This equality provides a diagonalization of A^T (since the matrix Λ is diagonal). □

[3]*Proof.* Assume the contrary. Then, at least one diagonal entry of Λ is zero. But the matrix Λ is diagonal, and thus upper-triangular. Hence, the determinant of Λ equals the product of its diagonal entries, and therefore equals 0 (since at least one diagonal entry is 0, and therefore the whole product must be 0). In other words, $\det \Lambda = 0$. Now, from $A = S\Lambda S^{-1}$, we obtain $\det A = \det\left(S\Lambda S^{-1}\right) = \det S \cdot \underbrace{\det \Lambda}_{=0} \cdot \det\left(S^{-1}\right) = 0$. Hence, A is not invertible (since a square matrix with determinant 0 is not invertible). This contradicts the fact that A is invertible. This contradiction shows that our assumption was false, qed.

[This was not the easiest or most elementary proof, but the shortest one.]