

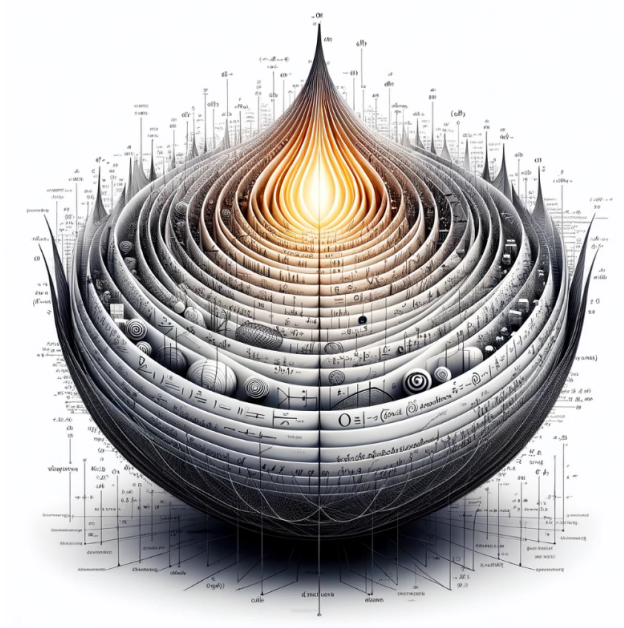
Lecture Notes

Honours Linear Algebra TATA53

Jonathan Nilsson

Spring 2024

This text will be updated with lecture notes throughout the course. It should be regarded as a complement to the course literature. These notes also contain a number of problems intended for the exercise sessions of the course, a list of suggested problems can be found in the syllabus. The problems are designed to prepare you for the assignments. Most problems have a hint in a separate section at the end of this text, these are intended to get you started if you get stuck, but you are encouraged to try and attack the problem without looking at the hints first. There is also a section of answers to the problems.



Contents

1	Vector spaces	4
1.1	Motivation	4
1.2	Definition of vector spaces	4
1.3	Basis and dimension	4
1.4	Subspaces	5
1.5	Direct sum	5
1.6	Affine subsets	7
2	Matrices	10
2.1	A basis for the matrix space	10
2.2	Echelon forms	11
2.3	Elementary matrices	12
2.4	LU-decomposition	13
2.5	Cholesky-factorization	15
2.6	Determinants	16
3	Introductory spectral theory	16
3.1	Eigenvalues and eigenvectors	16
3.2	Complexification and realification	19
3.3	Invariant subspaces	20
3.4	Matrix polynomials	21
4	Jordan normal form	24
4.1	Properties of matrices on Jordan form	24
4.2	Structure theory for nilpotent operators	26
4.3	Jordan chains and Jordanization	30
4.4	Matrix functions and applications	34
5	Inner product spaces	40
5.1	Inner products	40
5.2	Norms	42
5.3	Orthonormal bases and projections	44
5.4	Gram-Schmidt	45
5.5	QR-decomposition	47
5.6	Self-adjoint, unitary, and normal operators	50
5.7	Spectral theorem for normal operators	52
5.8	Positive definite operators	53
6	Problems	57
7	Hints	73
8	Answers	81

Notation

For this course you will need to absorb material from various sources, and authors prefer different notations, something one has to get used to in more advanced math courses. Here is a list of common notations that appear in this and other texts on linear algebra.

$\mathbb{R}, \mathbb{C}, \mathbb{Q}$	The field of real, complex, and rational numbers respectively.
\mathbb{Z}_p or $\mathbb{Z}/p\mathbb{Z}$ or $\mathbb{Z}/(p)$	The field of integers modulo some prime p .
$\mathbb{F}, \mathbb{K}, k$	Common notation for an arbitrary field (like the ones above).
$u, \mathbf{v}, \bar{w}, \vec{x}$	Common notations for vectors.
\mathcal{P}_n or \mathbb{P}_n	Polynomials of degree $\leq n$, with coefficients in some field.
$\mathcal{C}[a, b]$	Set of continuous functions $[a, b] \rightarrow \mathbb{R}$
$\mathcal{C}^n[a, b]$	Set of n times continuously differentiable functions $[a, b] \rightarrow \mathbb{R}$
$A, B, C, D, E, M, N, X, Y$	Capital letters commonly used for matrices.
$\lambda, \mu, \alpha, \beta, a, b, c, s, t, r, z$	Lower-case and Greek letters commonly used for scalars.
$(1, 2, 3)$ or $(1, 2, 3)^T$	Vectors in \mathbb{R}^3 , some authors always prefer writing them as columns ^a .
$\text{Mat}_{m \times n}(\mathbb{C}), M_{m \times n}(\mathbb{C})$	The set of $m \times n$ matrices with complex coefficients.
$\text{Mat}_n(\mathbb{C}), M_n(\mathbb{C})$	Shorthand for the above when $m = n$ (the matrix is square)
$\text{Mat}_{m \times n}, M_{m \times n}, \text{Mat}_n, M_n$	Same as above, the field is supposed to be understood by the context.
$\text{diag}(d_1, d_2, d_3)$	The diagonal matrix with d_1, d_2, d_3 on the diagonal.
$(a_{ij})_{i,j}, (a_{ij})$	The matrix with the number a_{ij} in position (i, j)
A_{ij}	The element of the matrix A at position (i, j)
A_i	The i 'th column of the matrix A
δ_{ij}	The Kronecker delta function, 1 if $i = j$, otherwise 0.
$e_{ij}, e_{i,j}, E_{i,j}, E_{ij}$	The unit-matrix with a single 1 in position (i, j) and zeroes elsewhere.
e_i	Standard basis vector in \mathbb{C}^n (or \mathbb{R}^n) with a single 1 in position i .
$\text{span}(v_1, v_2, v_3), [v_1, v_2, v_3]$	The span of v_1, v_2, v_3 (the set of all linear combinations of the vectors).
$\ker(F), N(F)$	The kernel, or nullspace of the linear map (or matrix) F .
$\text{Im}(F), \text{Ran}(F), V(F)$	The image or range of a linear map (or matrix) F .
A^T, A^t	The transpose of a matrix A .
$A^*, \bar{A}^T, A^H, A^\dagger$	The conjugate transpose of a matrix A .
$\begin{pmatrix} 2 \\ 3 \end{pmatrix}_{\mathbf{e}}$ or $\mathbf{e} \begin{pmatrix} 2 \\ 3 \end{pmatrix}$	$2e_1 + 3e_2$, the vector with coordinates $(2, 3)$ in basis $\mathbf{e} = (e_1, e_2)$
$v_{\mathbf{e}}$	The coordinate vector of v with respect to the basis \mathbf{e}
$F_{\mathbf{e}}$ or $[F]_{\mathbf{e}}$	The matrix of $F : V \rightarrow V$ with respect to a basis \mathbf{e} for V .
$[F]_{\mathbf{f}}$ or $_{\mathbf{f}}[F]_{\mathbf{e}}$	The matrix of $F : V \rightarrow W$ with respect to a bases \mathbf{e} for V , and \mathbf{f} for W .
$[F]$	the matrix of F with respect to some basis understood by the context.
$\sigma(F)$ or $\text{Spec}(F)$	The spectrum of F , the set of eigenvalues.
$p_A(\lambda)$ or just p_A	$\det(A - \lambda I)$, the characteristic polynomial ^b of A .
$m_A(\lambda)$ or just m_A	The minimal polynomial of A
$\text{rk}(F), \text{rank}(F)$	The rank of a matrix (or a linear map) F .
$u \bullet v, u \cdot v, \bar{v}^T u$	The dot product of u , and v , the standard inner product on \mathbb{R}^n or \mathbb{C}^n
$(u, v), (u v), \langle u, v \rangle$	common notations for various inner products of u and v
$\ v\ $ or $ v $	The norm of v or the length ^c of the vector v .
$\ A\ _F$	The Frobenius norm of a matrix A (root of sum of squared absolute values of entries)
U^\perp	The orthogonal complement to a subspace U (with respect to some inner product)
$P_U(v), \text{proj}_U(v), v _U, v_U$	The projection of v onto the subspace U
$A > 0$	The matrix A is positive (all entries are positive ^d)
$\deg p(t), \deg p$	The degree of a polynomial p
m_λ, g_λ	The algebraic and the geometric multiplicity of an eigenvalue λ .

^aThe reason has to do with matrix-vector multiplication, when evaluating a linear map, the matrix of the map should always be multiplied by the vertical coordinate-matrix of the vector. Also, when forming matrices of vectors, writing them as columns is almost always best.

^bAlso called the secular polynomial. Sometimes λ is replaced by another variable. It can also be defined for a linear map. Some books define it as $\det(\lambda I - A)$ instead, this may differ only by a minus sign and has the same set of zeros.

^cNote that typically the length of the vector is not well defined, but depends on what inner product is used.

^dNote that this is different from "positive definite" which means that $x^*Ax > 0$ for all $x \neq 0$.

1 Vector spaces

1.1 Motivation

In a first linear algebra course, we typically think of a **vector** as something that looks like $v = (1, 2, -3)$, a triple of real numbers. However, it turns out that the concepts, tools, and techniques of linear algebra (such as linear systems, linear maps, matrices, eigenvalues, etc) are useful in a much broader context, and they can be used to solve problems seemingly unrelated to the space \mathbb{R}^n . In fact, it turns out that all we need is any type of objects that we can add and multiply by scalars in a coherent way.

1.2 Definition of vector spaces

Definition 1.1. A **vector space** over a field^a \mathbb{F} is a set V together with an "addition" operation $V \times V \rightarrow V$ and a "scalar multiplication" $\mathbb{F} \times V \rightarrow V$ that satisfy the following axioms for all $u, v, w \in V$ and all $\lambda, \mu \in \mathbb{F}$:

1. $u + v = v + u$
2. $(u + v) + w = u + (v + w)$
3. There is an element $0 \in V$ satisfying $0 + v = v$ for all $v \in V$
4. For every $v \in V$ there is an element $-v \in V$ such that $v + (-v) = 0$
5. $\lambda \cdot (\mu \cdot v) = (\lambda\mu) \cdot v$
6. $\lambda \cdot (u + v) = (\lambda \cdot u) + (\lambda \cdot v)$
7. $(\lambda + \mu) \cdot v = (\lambda \cdot v) + (\mu \cdot v)$

^aIn this course the field \mathbb{F} will typically be \mathbb{R} or \mathbb{C} . In general a field is any system of numbers that you can add and multiply, with these operations satisfying the same kind of axioms that the real numbers satisfy.

A vector space over \mathbb{F} is also called an \mathbb{F} -vector space. The elements of \mathbb{F} are called **scalars**, while the elements of the vector space V are called **vectors**. We typically write λv instead of $\lambda \cdot v$. A prototypical example of a real vector space is \mathbb{R}^2 , the set of all pairs (x, y) of real numbers, where addition is defined as $(x_1, y_1) + (x_2, y_2) := (x_1 + x_2, y_1 + y_2)$ and $\lambda \cdot (x, y) = (\lambda x, \lambda y)$. The same construction works if we replace \mathbb{R} by \mathbb{C} or any other field. There are however many other examples:

- \mathbb{C}^n , the set of complex n -tuples with coordinate-wise sum and scalar multiplication is a complex vector space.
- The set \mathcal{P}_n of polynomials with complex coefficients and degree $\leq n$ is a complex vector space.
- The set $\text{Mat}_{m \times n}(\mathbb{R})$ of real $m \times n$ -matrices (with the usual ways of adding matrices and multiplying matrices by real numbers) is a real vector space.
- The set of continuous functions $f : \mathbb{R} \rightarrow \mathbb{R}$ is a real vector space.
- The set of infinite sequences of complex numbers (a_1, a_2, \dots) is a complex vector space (like \mathbb{C}^n where $n = \infty$)
- The set of solutions to the differential equation $y''(x) + y'(x) - 6y(x) = 0$ is a real vector space.
- The field \mathbb{Z}_2 of two element is the set $\{0, 1\}$ where addition and multiplication is defined like on the real numbers except that $1 + 1 := 0$. The set of triples of 0's and 1's form a vector space over \mathbb{Z}_2 .

1.3 Basis and dimension

For reference we recap some of the important concepts familiar from a first linear algebra course. A **linear combination** of vectors $v_1, \dots, v_n \in V$ is a vector of form $\lambda_1 v_1 + \dots + \lambda_n v_n$, where the coefficients λ_i are scalars. The set of all such linear combination is called their **span** and is denoted $\text{span}(v_1, \dots, v_n)$.

If $V = \text{span}(v_1, \dots, v_n)$ we say that the vectors span or generate V , in this case it means that every vector of V can be expressed as a linear combination of v_1, \dots, v_n .

On the other hand, the vectors v_1, \dots, v_n are called **linearly independent** if

$$\lambda_1 v_1 + \dots + \lambda_n v_n = 0 \implies \lambda_i = 0 \forall i.$$

In other words, the only linear combination of the vectors that is zero is the trivial combination.

An ordered set of vectors (v_1, \dots, v_n) that both spans V and is linearly independent is called a **basis** for V , and we define the **dimension** $\dim V = n$, the number of basis vectors. The basis-conditions guarantee (and are equivalent to) the fact that every vector in V has a unique expression as a linear combination of the basis vectors. If $v = \lambda_1 v_1 + \dots + \lambda_n v_n$ we say that $(\lambda_1, \dots, \lambda_n)$ is the **coordinate vector** of v with respect to the basis (v_1, \dots, v_n) . Then summing and rescaling vectors in V corresponds to making the same operations on the respective coordinate-vectors, so with the basis in hand we can forget V and instead do all computations in \mathbb{C}^n (or \mathbb{F}^n if V is an \mathbb{F} -vector space).

Example 1.2. Consider again some of the vector spaces discussed above.

- In \mathbb{C}^3 , a basis is $((1, 0, 0), (0, 1, 0), (0, 0, 1))$. The dimension is 3.
- In \mathcal{P}_3 , a basis is $(1, x, x^2, x^3)$. The dimension is 4.
- A basis for $\text{Mat}_{2 \times 2}(\mathbb{R})$ is $\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$. The dimension is 4.
- The set of continuous functions is infinite-dimensional (and it's hard to write down a basis^a)
- The solution space to $y''(x) + y'(x) - 6y(x) = 0$ is 2-dimensional with basis (e^{-3x}, e^{2x}) .

^aIn fact, the axiom of choice is required.

△

1.4 Subspaces

Definition 1.3. Let V be a vector space. A nonempty subset $U \subset V$ is called **subspace** of V if it is closed under taking sums and products by scalars:

- $u_1, u_2 \in U \implies u_1 + u_2 \in U$
- $u \in U, \lambda \in \mathbb{F} \implies \lambda \cdot u \in U$

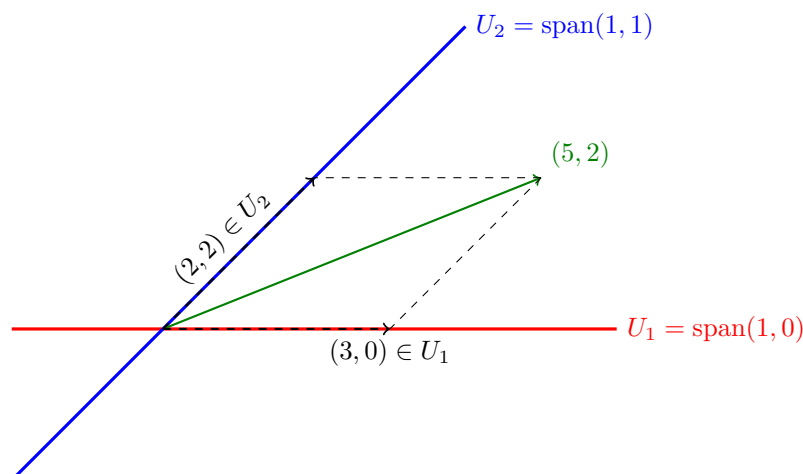
The definition is equivalent to saying that U itself is a vector space with the addition and scalar action inherited from V .

1.5 Direct sum

A basic way to understand an object is to break it down into smaller pieces and analyze them separately.

Definition 1.4. Let U_1 and U_2 be two subspaces of a vector space V . If every vector $v \in V$ has a *unique representation* $v = u_1 + u_2$ where $u_1 \in U_1$ and $u_2 \in U_2$, we say V is the (internal) **direct sum** of U_1 and U_2 , and we write $V = U_1 \oplus U_2$.

For example, let $U_1 = \text{span}(1, 0)$ and $U_2 = \text{span}(0, 1)$ be two lines in \mathbb{R}^2 . Then $U_1 \oplus U_2 = \mathbb{R}^2$; every vector of $v \in \mathbb{R}^2$ has a unique representation as $u_1 + u_2$ with $u_1 \in U_1$ and $u_2 \in U_2$. For example, $(5, 2) = (3, 0) + (2, 2)$, or more generally, $(x, y) = (x - y, 0) + (y, y)$.



The subspaces need to be one-dimensional though. For example, if U_1 is a plane in \mathbb{R}^3 and U_2 is a line (which is not parallel to the plane), then $\mathbb{R}^3 = U_1 \oplus U_2$.

On the other hand, \mathbb{R}^3 can never be the direct sum of two planes. For example, with

$$U_1 : z = 0 \text{ and } U_2 : x + y + z = 0,$$

each vector in \mathbb{R}^3 can be expressed as a sum $u_1 + u_2$ of vectors from the two subspaces, but the expression is not unique, for example there are two ways (and infinitely many more ways) of writing the vector $(1, 2, 3)$ as a sum $u_1 + u_2$ of vectors from each respective subspace:

$$(1, 5, 0) + (0, -3, 3) = (2, 4, 0) + (-1, -2, 3).$$

Note that if $V = U_1 \oplus U_2$ we have a natural way to project a vector onto each subspace:

$$P_{U_1} v = u_1 \text{ and } P_{U_2} v = u_2 \text{ where } v = u_1 + u_2.$$

Because of the uniqueness of the expression $v = u_1 + u_2$ these functions are well defined. Note that this is different from *orthogonal projection*¹, for example, in the graphical example above we have $P_{U_2}(5, 2) = (2, 2)$.

Theorem 1.5. Let U_1 and U_2 be subspaces of a vector space V . Then $V = U_1 \oplus U_2$ if and only if² $U_1 + U_2 = V$ and $U_1 \cap U_2 = \{0\}$.

Proof. The first condition $U_1 + U_2 = V$ is clearly equivalent to being able to express each $v \in V$ as $v = u_1 + u_2$. We show that the second condition is equivalent to uniqueness of the expression: assume that $U_1 \cap U_2$ contains some nonzero vector u_0 . Then for any vector v with representation $v = u_1 + u_2$, we can also write $v = (u_1 + u_0) + (u_2 - u_0)$, so no vector has a unique representation. Conversely, if a vector v has two distinct representations $u_1 + u_2 = v = u'_1 + u'_2$, then subtracting we get $(u_1 - u'_1) = (u'_2 - u_2)$. This vector clearly lies in U_1 since the left hand does, and it lies in U_2 since the right side does. Moreover the vector is nonzero since we assumed the representations were different. Thus we have a nonzero vector that lies in both U_1 and U_2 , so $U \cap U' \neq \{0\}$. \square

Note that if (u_1, \dots, u_n) is a basis for U and if (u'_1, \dots, u'_m) is a basis for U' , then $(u_1, \dots, u_n, u'_1, \dots, u'_m)$ is a basis for $V = U \oplus U'$. In particular $\dim(U \oplus U') = \dim(U) + \dim(U')$.

Example 1.6. Consider the vector space $V = \text{Mat}_n(\mathbb{R})$, and let S be the subspace consisting of symmetric matrices, and let S' be the subspace consisting of skew-symmetric matrices. Let us prove that

$$V = S \oplus S'.$$

According to the theorem above, it suffices to prove that $S + S' = V$ and that $S \cap S' = \{0\}$. Starting with the latter, assume that a matrix A lies in $S \cap S'$. Then A is both symmetric and

¹Indeed, in a general vector space we don't even have the concept of orthogonality before defining an inner product. Note also that it doesn't make sense to project onto a subspace U_1 in this sense without specifying what we choose as U_2 .

²Here $U_1 + U_2 := \{u_1 + u_2 \mid u_1 \in U_1, u_2 \in U_2\}$.

skew symmetric, so $A = A^T = -A$, so $2A = 0$ and therefore $A = 0$. Thus S and S' intersects only in the zero matrix.

For the other part we need to show that any square matrix can be expressed as the sum of a symmetric and a skew-symmetric matrix. For this we note that $B = A + A^T$ is symmetric (since $B^T = (A + A^T)^T = A^T + A = B$), and that $C = (A - A^T)$ is skew symmetric (since $C^T = (A - A^T)^T = A^T - A = -C$). But then we can express A as a sum a symmetric and a skew-symmetric matrix as follows:

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T).$$

△

We can also talk about direct sum of more than two components in an analogous way: If U_1, \dots, U_n are subspaces of V , we say that $V = U_1 \oplus \dots \oplus U_n$ if and only if each vector $v \in V$ has a unique expression

$$v = u_1 + \dots + u_n \text{ where } u_i \in U_i.$$

External direct sum

There is also an analogous construction called the **external direct sum**. Starting with two vector spaces we can construct a larger vector space that has the two vector spaces as direct summands in the above sense (roughly speaking).

Definition 1.7. Let V and W be vector spaces over the same field \mathbb{F} . We define $V \oplus W$ to be the set $V \times W$ which consists of all pairs (v, w) where $v \in V$ and $w \in W$, and we define the sum and scalar action on such pairs in the natural way:

$$(v, w) + (v', w') := (v + v', w + w') \text{ and } \lambda \cdot (v, w) := (\lambda \cdot v, \lambda \cdot w).$$

Under these operations $V \oplus W$ becomes a vector space, called the (external) **direct sum** of V and W .

Note that technically V and W are not subspaces (and not even a subsets) of $V \oplus W$, since the latter object consists of elements of form (v, w) while V and W do not. However, if we identify V with pairs of form $(v, 0)$, and W with pairs $(0, w)$, then $V \oplus W$ is the internal direct sum of V and W . If (v_1, \dots, v_m) is a basis for V and if (w_1, \dots, w_n) is a basis for W then

$$((v_1, 0), \dots, (v_m, 0), (0, w_1), \dots, (0, w_n))$$

is a basis for $V \oplus W$.

Example 1.8. The vector space $V = \text{Mat}_{2 \times 2}(\mathbb{R}) \oplus \mathcal{C}(\mathbb{R})$ consists of objects of form $(A, f(x))$, where A is a 2×2 -matrix and $f : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function. Here's an example of a linear combination in V :

$$\frac{1}{2} \left(\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}, e^x \right) - \frac{1}{2} (I, e^{-x}) = \left(\begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix}, \sinh x \right).$$

△

External direct sums of more than two vector spaces can also be defined in the natural way.

1.6 Affine subsets

If u is a vector of a subspace, we know that $0 \cdot u \in U$. This means that the zero-vector always belongs to a subspace. An affine subset is something that looks like a subspace (think a line or a plane), but it is shifted away from the origin.

Definition 1.9. Let U be a subspace of V . For each $v \in V$ we define

$$v + U := \{v + u \mid u \in U\}$$

and call this an **affine subset** (or affine subspace) parallel to U .

Intuitively, this is just the subspace U shifted by a vector v . For example, the line $\{(2 + 3t, 1 + 5t) \in \mathbb{R}^2 \mid t \in \mathbb{R}\}$ is an affine subset of \mathbb{R}^2 . It can be written as $(2, 1) + U$ where $U = \text{span}(3, 5)$. Similarly, the plane $x + 2y + 3z = 5$ is an affine subspace of \mathbb{R}^3 , it can be written $(0, 1, 1) + U$ where U is the plane $x + 2y + 3z = 0$.

Note that $v \in U$ if and only if $v + U = U$. Also note that an affine subset typically is not closed under addition or scalar multiplication.

It turns out that the *set of all affine subsets* is itself a vector space in a natural way.

Definition 1.10. Let U be a subspace of V . We define the **quotient space**

$$V/U := \{v + U \mid v \in V\}.$$

This is a vector space in the natural way, addition of two affine subsets is defined as

$$(v + U) + (v' + U) := (v + v') + U$$

and multiplication of scalars is defined by

$$\lambda(v + U) := (\lambda v) + U.$$

So intuitively, to add two affine subsets, pick a vector in each one, add them, and then take the affine subset which the sum lies in. It turns out this operation is well defined. For example. Let $U = \text{span}(1, 1)$. Then \mathbb{R}^2/U is the set of lines in \mathbb{R}^2 with slope 1.

If (u_1, \dots, u_m) is a basis for U and if we extend it to a basis (u_1, \dots, u_n) for V , it is easy to see that a basis for V/U is given by the affine subsets

$$(u_{m+1} + U, \dots, u_n + U).$$

In particular $\dim(V/U) = \dim(V) - \dim(U)$. Note however that V/U is not itself a subset of V .

Note also that $0 + U = u + U$ for $u \in U$ - this means that in V/U , we can't tell the difference between different elements of U , so a good way to think about V/U is that we take V and then we "make all the elements of U equal".

Linear maps

Definition 1.11. Let V and W be \mathbb{F} -vector spaces. A map $F : V \rightarrow W$ is called **linear** if

$$F(u + v) = F(u) + F(v) \quad \text{and} \quad F(\lambda v) = \lambda F(v)$$

holds for all $u, v \in V$ and $\lambda \in \mathbb{F}$.

Linear maps are sometimes called linear transformations, endomorphisms, or operators if $V = W$.

A linear map is completely determined by its action on the basis vectors, because if $\mathcal{B} = (v_1, \dots, v_n)$ is a basis for V , a linear map F satisfies

$$F(\lambda_1 v_1 + \dots + \lambda_n v_n) = \lambda_1 F(v_1) + \dots + \lambda_n F(v_n).$$

In fact, if pick a basis $\mathcal{B}' = (w_1, \dots, w_m)$ of W and express the vectors $F(v_i)$ in this basis, and put them as columns in a matrix, we get the matrix of F with respect to the two bases \mathcal{B} and \mathcal{B}' :

$$[F]_{\mathcal{B}', \mathcal{B}} = \begin{pmatrix} | & | & & | \\ F(v_1) & F(v_2) & \cdots & F(v_n) \\ | & | & & | \end{pmatrix}.$$

Example 1.12. Consider the map $D : \mathcal{P}_3 \rightarrow \mathcal{P}_2$ defined by taking derivative, $D(p(x)) = p'(x)$. This map is linear because of the familiar rules proved in a first calculus course:

$$\frac{d}{dx}(f(x) + g(x)) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x) \quad \text{and} \quad \frac{d}{dx}(\lambda f(x)) = \lambda \frac{d}{dx}f(x).$$

Let $\mathcal{B} = (1, x, x^2, x^3)$ and $\mathcal{B}' = (1, x, x^2)$ be the standard bases \mathcal{P}_3 and \mathcal{P}_2 respectively. To find

the matrix for D with respect to these bases, evaluate D on the basis vectors in \mathcal{B} and express them in the basis \mathcal{B}' :

$$D(1) = 0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}_{\mathcal{B}'} \quad D(x) = 1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}_{\mathcal{B}'} \quad D(x^2) = 2x = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}_{\mathcal{B}'} \quad D(x^3) = 3x^2 = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}_{\mathcal{B}'}. \quad .$$

So the matrix for D with respect to \mathcal{B} and \mathcal{B}' is

$$[D]_{\mathcal{B}', \mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

△

There are two important subspaces associated to a linear map:

Definition 1.13. Let $F : V \rightarrow W$ be linear. We define

$$\ker(F) = \{v \in V \mid F(v) = 0\} \quad \text{and} \quad \text{Im}(F) = \{F(v) \mid v \in V\}$$

and call these the **kernel** and **image** of F respectively.

Note that $\ker(F)$ is a subset of V and $\text{Im}(F)$ is a subset of W , in fact it is not hard to prove that they are *subspaces*. The kernel is also called the nullspace of F , and the image is sometimes called the range of F .

We recall the important *dimension theorem*, also called the *rank-nullity theorem*:

Theorem 1.14. Let $F : V \rightarrow W$ be linear. Then

$$\dim \ker(F) + \dim \text{Im}(F) = \dim V.$$

The theorem holds even for infinite-dimensional vector spaces if we define $5 + \infty = \infty$ and so on. The dimension of the image $\dim \text{Im}(F)$ is also called the **rank** of the linear map.

Recall also that if $F : V \rightarrow W$ is a map, an **inverse** of F is a map in the other direction $G : W \rightarrow V$ such that

$$G(F(v)) = v \text{ for all } v \in V \quad \text{and} \quad F(G(w)) = w \text{ for all } w \in W.$$

This is commonly expressed as $G \circ F = \text{id}_V$ and $F \circ G = \text{id}_W$ ³.

Direct sum of linear maps

Let $F : V \rightarrow V$ and $G : W \rightarrow W$ be linear maps. Then we get a corresponding linear map $F \oplus G : V \oplus W \rightarrow V \oplus W$ defined by $(F \oplus G)(v, w) = (F(v), G(w))$. Note that if A is the matrix for F with respect to a given basis of V , and if B is the matrix of G with respect to a given basis of W , then the matrix of $F \oplus G$ with respect to the corresponding basis of $V \oplus W$ is of block form

$$\left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array} \right).$$

Linear maps and quotient spaces

We have a natural linear projection map $\pi : V \rightarrow V/U$ defined by $\pi(v) := v + U$.

On the other hand, suppose $F : V \rightarrow W$ is a linear map, and suppose that U is a subspace of V that lies inside the kernel of F : $F(U) = \{0\}$. Then we can construct a corresponding linear map

$$\tilde{F} : V/U \rightarrow W \text{ defined by } \tilde{F}(v + U) := F(v).$$

The condition $F(U) = \{0\}$ guarantees that the map is well defined.

³Here $\text{id}_V : V \rightarrow V$ is the identity map on V mapping each v to itself.

2 Matrices

2.1 A basis for the matrix space

Let e_{ij} be the $m \times n$ -matrix which has a single 1 in position (i, j) and zeroes elsewhere⁴. The set of all these matrices clearly form a basis for $\text{Mat}_{m \times n}(\mathbb{C})$. Note that if e_{ij} and e_{kl} are such matrices (of compatible sizes), then

$$e_{ij}e_{kl} = \begin{cases} e_{il} & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}.$$

This is typically expressed more compactly as $e_{ij}e_{kl} = \delta_{jk}e_{il}$, where δ_{jk} is the **Kronecker-delta** function, it is 1 if $j = k$ and zero otherwise.

We write $(a_{ij})_{ij}$ or just (a_{ij}) for the matrix that has element a_{ij} in position (i, j) , in other words $(a_{ij})_{ij} = \sum_{i,j} a_{ij}e_{ij}$.

Let $A = (a_{ij})_{ij}$ be a matrix. Recall that the **transpose** of A is $A^T = (a_{ji})_{ij}$. For complex matrices we also define the **Hermitian conjugate** of A as $A^* = (\overline{a_{ji}})$, this is just the conjugate-transpose of A . A matrix $A = (a_{ij})_{ij}$ is called...

- Diagonal if $a_{ij} = 0$ whenever $i \neq j$
- Upper triangular if $a_{ij} = 0$ whenever $i > j$
(strictly upper triangular if $a_{ii} = 0$ also)
- Lower triangular if $a_{ij} = 0$ whenever $j > i$
(strictly lower triangular if $a_{ii} = 0$ also)
- Symmetric if $a_{ij} = a_{ji}$
(skew-symmetric if $a_{ij} = -a_{ji}$)
- Hermitian if $a_{ij} = \overline{a_{ji}}$
(skew-Hermitian if $a_{ij} = -\overline{a_{ji}}$)

Note that a matrix needs to be square in order to be symmetric/skew-symmetric/Hermitian, but the other concepts apply for any size of matrix.⁵ Note also that Hermitian and symmetric has the same meaning when the matrix is real.

From the product-rule for matrices e_{ij} it follows that matrix multiplication can be expressed like this: If $A = (a_{ij})_{ij}$ is an $m \times n$ -matrix and $B = (b_{ij})_{ij}$ is an $n \times k$ -matrix, then

$$AB = \left(\sum_{r=1}^n a_{ir}b_{rj} \right)_{ij},$$

in words this just encodes the familiar rule that the element in position (i, j) in the product is the scalar-product of row number i in A and column number j in B .

Note that $e_{12}e_{23} = e_{13}$ while $e_{23}e_{12} = 0$, so in general $AB \neq BA$ when A and B are matrices. We say that two $n \times n$ -matrices A and B **commute** if $AB = BA$.

Usually we shall not differentiate between square matrices and linear operators. For example, if A is an $m \times n$ -matrix, $\ker(A)$ is the set of vectors $X \in \mathbb{C}^n$ satisfying $AX = 0$ (when X is written as a column), and similarly $\text{Im}(A)$ is the set $\{AX \mid X \in \mathbb{C}^n\}$, which is the same as the span of the columns of A . The rank of A , defined as $\dim \text{Im}(A)$, can therefore be characterized as the maximum number of linear independent columns of A .

Recall that if a linear operator $F : V \rightarrow V$ has matrix A with respect to one basis, and matrix B with respect to a different basis, then $A = SBS^{-1}$, where S has the new basis vectors as columns expressed in the old basis. We say that two square matrices A and B are **similar** if there is a matrix S such that $A = SBS^{-1}$. Then two matrices are similar if they represent the same linear map $V \rightarrow V$ with respect to different choices of basis.

⁴Note that just writing e_{ij} is not completely clear since the size of the matrix is not specified, the 3×3 -matrix e_{12} is different from the 2×2 -matrix e_{12} . However, the format of the matrix is usually obvious from the context.

⁵Although when we say diagonal matrix without qualification, we usually mean a square matrix.

The **trace** of an $n \times n$ -matrix A is the sum of the diagonal entries:

$$\operatorname{tr}(A) := \sum_{k=1}^n a_{kk}.$$

For example we have $\operatorname{tr} \begin{pmatrix} 2+i & 3 \\ 4i & 3-4i \end{pmatrix} = 5 - 3i$.

Now let $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$. Then we have

$$\operatorname{tr}(AB) = \operatorname{tr} \begin{pmatrix} 2 & 3 \\ 2 & 5 \end{pmatrix} = 7 \text{ and } \operatorname{tr}(BA) = \operatorname{tr} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 3 & 2 \\ 1 & 4 & 3 \end{pmatrix} = 7.$$

So $\operatorname{tr}(AB) = 7 = \operatorname{tr}(BA)$. This is no accident:

Theorem 2.1. *If both products AB and BA are defined, we have*

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$

The proof is left as an exercise. An important corollary is that similar matrices have the same trace:

$$\operatorname{tr}((S^{-1}B)S) = \operatorname{tr}(S(S^{-1}B)) = \operatorname{tr}(IB) = \operatorname{tr}(B).$$

This lets us define the trace of a linear operator, as it is independent of the choice of basis. A further consequence of this is that if A is diagonalizable with $A = SDS^{-1}$, where D has the eigenvalues of A on the diagonal, then $\operatorname{tr}(A) = \operatorname{tr}(D)$ which is the *sum of the eigenvalues of A including multiplicities*.⁶

A square matrix⁷ is called **nilpotent** if $N^d = 0$ for some d , the minimal such d is called the nilpotency-degree of N . Nilpotent matrices will be important later when we investigate the Jordan normal form. The prototypical example of such a matrix is:

$$N = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad N^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad N^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad N^4 = 0.$$

If we think of N as a linear operator, it acts on the standard basis like so:

$$e_4 \mapsto e_3 \mapsto e_2 \mapsto e_1 \mapsto 0.$$

The nilpotency-degree of N is 4, and it is clear that $N^m = 0$ for all $m \geq 4$. Note that if A is similar to N and $N^d = 0$, then $A^d = 0$ too, so the nilpotency-degree is basis independent.

2.2 Echelon forms

When solving a linear system by Gaussian elimination, we use row operations to reduce the coefficient matrix to a form suitable for writing down the solutions. The matrix A below is in *row echelon form*. With some further row operations we can reduce it to B which is in *reduced row echelon form*, the encircled elements are called *pivots*.

$$A = \begin{pmatrix} \textcircled{1} & 1 & 9 & 2 & 1 & 8 \\ 0 & \textcircled{2} & 4 & 1 & 2 & 5 \\ 0 & 0 & 0 & \textcircled{4} & 1 & 9 \\ 0 & 0 & 0 & 0 & \textcircled{3} & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad B = \begin{pmatrix} \textcircled{1} & 0 & 7 & 0 & 0 & 2 \\ 0 & \textcircled{1} & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & \textcircled{1} & 0 & 2 \\ 0 & 0 & 0 & 0 & \textcircled{1} & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

⁶This also true for non-diagonalizable operators, we will prove this later.

⁷Nilpotency can be defined the exact same way when $N : V \rightarrow V$ is a linear operator.

Definition 2.2. A matrix is said to be in **row echelon form** (REF) if the first nonzero element of each row is to the *left* of the first nonzero elements in all the rows below, and if all zero-rows are at the bottom. The first nonzero elements of each row in the REF are called **pivots**. The matrix is said to be in **reduced row echelon form** (RREF) if additionally, all pivots are 1, and the pivots have zeros above them.

We know that any matrix can be reduced to REF and RREF by row operations. The RREF of a matrix is unique, and the number of the pivots in the RREF (or REF) is the **rank** of the matrix.

From the RREF we can immediately solve the corresponding linear system. For example, to solve $BX = 0$ for the matrix B above, we introduce parameters $x_3 = s$ and $x_6 = t$ for the variables corresponding to non-pivot columns, then from the RREF we immediately see that the set of solutions⁸ is

$$(x_1, x_2, x_3, x_4, x_5, x_6) = (-7s - 2t, -2s - t, s, -2t, -t, t) \quad s, t \in \mathbb{R}.$$

Since the matrices A and B are in fact row equivalent, the equation $AX = 0$ has the same solutions.

In the above example, we considered a matrix in $\text{Mat}_{5 \times 6}(\mathbb{R})$, but note that matrices, row operations, REF, and RREF makes sense over any field \mathbb{F} .

2.3 Elementary matrices

Row operations on a matrix can be performed by multiplying the matrix from the *left* by an **elementary matrix**. This is best explained by looking at some concrete examples:

$$E_1 A = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & -1 \\ 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Here we note that multiplying A by the matrix E_1 on the left has the same effect as performing the row operation of *adding (-2) times the first row to the second row*.

Another row operation is to multiply one of the rows of A by a nonzero scalar λ , this can be achieved by multiplying by another type of matrix on the left, for example

$$E_2 A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & -1 \\ 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 3 & -3 \\ 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

The last type of row-operation is switching two rows:

$$E_3 A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & -1 \\ 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 0 & 1 \end{pmatrix}.$$

In general, an elementary matrix is an $m \times m$ -matrix of one of the three forms below (empty positions are zeros). Multiplying such a matrix by an $m \times n$ -matrix A from the left has the effect of making a row operation on A .

⁸Here one can ask if we should write $s, t \in \mathbb{C}$ instead, we still get solutions to the system for such s, t . In this case, the matrix was real, so we assumed that we were working over the real numbers.

Matrix	Corresponding row operation
$\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & \lambda & \ddots \\ & & & 1 \end{pmatrix} = I + \lambda e_{ij}$ <p>(λ in position (i, j))</p>	Add λ times row j to row i
$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \lambda & \\ & & & \ddots \\ & & & & 1 \end{pmatrix} = I + (\lambda - 1)e_{ii}$ <p>(identity except $\lambda \neq 0$ on position (i, i))</p>	Multiply row i by a nonzero scalar λ
$\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} = I - e_{ii} - e_{jj} + e_{ij} + e_{ji}$ <p>(I but with rows i and j switched)</p>	Switching rows i and j

Note that multiplying a matrix by an elementary matrix from the *right side* instead has the effect of performing a corresponding column-operation. This however is less useful, for example if we perform a column-operation on a linear system it no longer has the same solutions.

2.4 LU-decomposition

Definition 2.3. An LU-decomposition of an $m \times n$ matrix A is a factorization

$$A = LU$$

where L is a lower-triangular $m \times m$ -matrix, and U is an upper triangular matrix $m \times n$ -matrix.

An LU-decomposition of a matrix A can typically be obtained by reducing A to row echelon form (REF) and keeping track of the elementary matrices corresponding to the row operations.

Example 2.4. We shall find an LU-decomposition $A = LU$ of the matrix A below. We start by row-reducing A to row echelon form U :

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & 3 & -1 \\ -2 & 2 & -1 & -5 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & -2 \\ 0 & 4 & 1 & -3 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & -2 \\ 0 & 0 & -3 & 1 \end{pmatrix} = U.$$

We performed three row operations:

- Add (-1) times the first row to the second row
- Add (2) times the second row to the third row
- Add (-2) times the second row to the the third row

These row operations correspond to left multiplication by these elementary matrices:

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \quad E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix}.$$

Thus we have $E_3(E_2(E_1A)) = (E_3E_2E_1)A = U$ so $A = (E_3E_2E_1)^{-1}U = (E_1^{-1}E_2^{-1}E_3^{-1})U = LU$, where

$$L = E_1^{-1}E_2^{-1}E_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix}$$

We conclude that an LU-decomposition is given by

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & 3 & -1 \\ -2 & 2 & -1 & -5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & -2 \\ 0 & 0 & -3 & 1 \end{pmatrix} = LU.$$

△

Some questions remain - when do LU-decompositions exist, are they unique when they do, and what are they good for?

The reason that the method in the example works is that typically we can reduce a matrix to REF simply by adding multiples of rows to rows below them - the corresponding elementary matrices will be lower triangular, and then their inverses and their product L is also lower triangular. The only problem is if we need to switch two rows to reach an echelon form of A . In this case we can first perform a sequence of row switches in A by left-multiplying by elementary row-switching matrices. Let P be the product of these matrices⁹ Then we proceed with the LU-decomposition as usual to obtain a factorization¹⁰ $PA = LU$. We have proved the following theorem:

Theorem 2.5. *Each $m \times n$ matrix A admits a decomposition $PA = LU$ where*

- L is lower triangular $m \times m$ matrix with ones on the diagonal
- P is a permutation matrix of size $m \times m$
- U is an upper triangular $m \times n$ matrix (a row echelon form of A)

Is such a decomposition unique? Well more or less. First, consider our example above and factor the matrix U as DU' like so:

$$A = LU = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & -2 \\ 0 & 0 & -3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & -\frac{1}{3} \end{pmatrix} = LDU'.$$

This is called an **LDU-decomposition** of A , and it can clearly be found from the LU-decomposition as above.

Definition 2.6. An LDU-decomposition of A is a factorization $A = LDU$ where L is lower triangular $m \times m$ with ones on the diagonal, D is $m \times m$ and diagonal, and U is a row echelon form of A with ones as pivots.

Here we note that $A = (LD)U = L(DU)$ are two LU-decompositions of A (unless $D = I$). So in general the LU-decomposition is not unique, however, with some additional conditions it is.

Proposition 2.7. *If an invertible $n \times n$ matrix A admits an LU-decomposition, then it is unique if we require that L has ones on the diagonal. It follows that it has a unique LDU-decomposition too.*

Proof. Assume $L_1U_1 = L_2U_2$ are two LU-decompositions of A . Then with $L = L_1^{-1}L_2$ we have $U_1 = LU_2$, where U_1 and U_2 are both in row echelon form. But note that since A is invertible, any echelon form must be upper triangular with nonzero elements on the diagonal, but then LU can only itself be upper triangular if $L = I$, which means $L_1 = L_2$, and then since this matrix is invertible we also get $U_1 = U_2$. □

⁹A product of row-switching elementary matrices is called a *permutation matrix* - these can be characterized as having a single 1 in each row and each column, with zeros elsewhere.

¹⁰Equivalently, $A = P^{-1}LU$. Since P^{-1} is also a permutation matrix, some books prefer to write the factorization as $A = PLU$

What are LU-decompositions good for? Imagine that we have a large system of linear equations $Ax = b$ (let's say A is $n \times n$, while $b, x \in \mathbb{R}^n$). Assume we want to solve this system for many different right sides b , and perhaps at different times (or perhaps one b is used to calculate the next recursively). Then if we have a factorization $A = LU$, we have

$$Ax = b \Leftrightarrow L(Ux) = b \Leftrightarrow Ly = b \text{ and } Ux = y.$$

So instead of solving $Ax = b$ directly we can solve the two systems $Ly = b$ and then $Ux = y$. These systems are triangular, so they are fast to solve by back-substitution. If this is done by a computer on a very large matrix (say $n = 10^4$), the speed-increase is significant.¹¹

2.5 Cholesky-factorization

Definition 2.8. A Cholesky-factorization of a square matrix A is of form

$$A = CC^*$$

where C is a lower triangular matrix.

Since CC^* is always Hermitian, so must A be in order to admit such a decomposition. In case an invertible Hermitian matrix A has a decomposition $A = LDU$, we note that

$LDU = A = A^* = U^*D^*L^*$ are two LDU-decompositions, so by uniqueness $L = U^*$ and $D = D^*$ so $A = LDL^*$ and D is real. Moreover, assume that D has positive entries

$$D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix} \text{ and define } \sqrt{D} = \begin{pmatrix} \sqrt{d_1} & & & \\ & \sqrt{d_2} & & \\ & & \ddots & \\ & & & \sqrt{d_n} \end{pmatrix}.$$

The reason for the notation is that $(\sqrt{D})^2 = D$. Now take $C = L\sqrt{D}$. Then C is lower triangular, and

$$CC^* = L\sqrt{D}\sqrt{D}^*L^* = L\sqrt{D}\sqrt{D}L^* = LDL^* = A,$$

so $A = CC^*$ is the¹² Cholesky-factorization of A .

The Cholesky-factorization can be used when solving linear systems with a Hermitian coefficient matrix - algorithmically it is twice as efficient as using the LU-decomposition.

Example 2.9. Let's find the Cholesky-factorization of $A = \begin{pmatrix} 2 & 4 \\ 4 & 12 \end{pmatrix}$. We can reduce it to REF by a single row-operation:

$$EA = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 4 & 12 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 0 & 4 \end{pmatrix} \text{ so } A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = LDU$$

is the LDU-decomposition of A . Now take

$$C = L\sqrt{D} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 \\ 2\sqrt{2} & 2 \end{pmatrix},$$

then it is easy to verify that

$$CC^* = \begin{pmatrix} \sqrt{2} & 0 \\ 2\sqrt{2} & 2 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 4 & 12 \end{pmatrix} = A$$

is the Cholesky-decomposition of A .

△

¹¹But we need to find the LU-decomposition first, so for a single right side b it is not an effective method.

¹²One can show that the Cholesky-factorization exists for Hermitian matrices A that are positive-semidefinite (meaning that all its eigenvalues are ≥ 0). If the matrix is positive-definite the Cholesky-factorization is additionally unique.

2.6 Determinants

Recall from a first linear algebra course that if $A \in \text{Mat}_n(\mathbb{R})$, the **determinant** of A was a real number that could be calculated by several methods:

- Sarrus' rule for 2×2 and 3×3 matrices
- Expansion along rows or columns
- Make row or column operations, then the determinant is the product of the diagonal entries¹³
- As a sum over the permutation group S_n (maybe not in a first course)

The determinant also had a number of important properties:

- $\det(I) = 1$
- $\det(A) = \det(A^T)$
- $\det(AB) = \det(A)\det(B)$
- $\det(A) \neq 0 \Leftrightarrow A^{-1}$ exists and each system $AX = 0$ has a unique solution

It follows from the second point that that $\det(S^{-1}AS) = \det(A)$, so the determinant is basis-independent and can be defined for any linear map. If a linear map is diagonalizable, its determinant is therefore the same as the determinant of $D = \text{diag}(\lambda_1, \dots, \lambda_n)$, in other words, $\det(A) = \lambda_1 \cdots \lambda_n$, the determinant is the product of all eigenvalues counting multiplicities. This also holds for non-diagonalizable maps which we shall see later.

For now we just remark that all these rules and properties above work the same over any field \mathbb{F} . When $\mathbb{F} = \mathbb{C}$ we also note that $\det(\overline{A}) = \overline{\det(A)}$, and therefore $\det(A^*) = \overline{\det(A)}$.

Example 2.10. The determinant of the linear map $\mathbb{C}^2 \rightarrow \mathbb{C}^2$ with standard matrix $A = \begin{pmatrix} 1+i & 1 \\ i & 3 \end{pmatrix}$ is $\det(A) = 3(1+i) - i = 3 + 2i \neq 0$, so the map is invertible.

△

3 Introductory spectral theory

3.1 Eigenvalues and eigenvectors

We recap the theory of eigenvalues and eigenvectors from a first linear algebra course.

Definition 3.1. Let $F : V \rightarrow V$ be a linear map on a vector space over a field \mathbb{F} . If

$$F(v) = \lambda v \text{ for some } \lambda \in \mathbb{F} \text{ and some nonzero } v \in V,$$

we say that λ is an **eigenvalue** for F , and v is a corresponding **eigenvector**.

When $\dim V < \infty$, the map F may be described by a matrix A after a basis is picked, so we shall also speak of eigenvalues and eigenvectors of *matrices*, the condition in the definition then looks like $Av = \lambda v$.

The eigenvalues give important information about the geometric nature of a map as the following example illustrates.

Example 3.2. Let $P : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be given by $[P] = \frac{1}{14} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$ in the standard basis. Then by the standard methods of linear algebra we may find its eigenvalues and eigenvectors. It turns out that P has two eigenvalues: 0 and 1. Every nonzero vector in the plane $x + 2y + 3z = 0$ is an eigenvector for the eigenvalue 0, and every nonzero vector on the line $t(1, 2, 3)$ is an eigenvector for

¹³Remember that changing two rows negates the determinant, and multiplying a row or column by a λ changes the determinant by a factor λ^n

the eigenvalue 1. From this information we can deduce that geometrically, F is a projection onto the line $t(1, 2, 3)$.

△

Example 3.3. Let $C^\infty(\mathbb{R})$ be the real vector space of infinitely differentiable functions $\mathbb{R} \rightarrow \mathbb{R}$, and let D be the linear operator on this space that takes the derivative, $D(f(x)) = f'(x)$. Then every $\lambda \in \mathbb{R}$ is an eigenvalue for D , the set of corresponding eigenvectors are nonzero multiples of $e^{\lambda x}$.

△

The **spectrum** of a linear operator F is the set of eigenvalues, and it is written $\sigma(F)$. The spectrum gives important qualitative information about the operator. In the above examples we have $\sigma(P) = \{0, 1\}$ and $\sigma(D) = \mathbb{R}$. We shall focus on the case when the dimension of the vector space is finite, then there is a concrete standard method for finding the eigenvalues and eigenvectors: fix a basis for V and consider the matrix A of the linear operator. Then λ is an eigenvalue and a nonzero v is an eigenvector if and only if $Av = \lambda v$, or equivalently $(A - \lambda I)v = 0$. This matrix-equation has a nontrivial solution v if and only if $\det(A - \lambda I) = 0$. Solving this equation¹⁴ gives us the eigenvalues, and for each eigenvalue λ we can then solve $(A - \lambda I)v = 0$ to find the corresponding eigenvectors.

For $\lambda \in \mathbb{F}$, we define the corresponding **eigenspace**, to be $E_\lambda := \ker(A - \lambda I)$. This is the subspace of V consisting of all vectors¹⁵ v satisfying $Av = \lambda v$. Note that $\ker(A - \lambda I) = \{0\}$ when λ is not an eigenvalue.

Now let A be a matrix representing a linear operator on a finite-dimensional vector space over a field \mathbb{F} . The **characteristic polynomial**¹⁶ for A is defined as

$$p_A(\lambda) = \det(A - \lambda I).$$

Then $p_A(\lambda) = 0$ if and only if $\lambda \in \mathbb{F}$ is an eigenvalue of A . Note that the coefficients of p_A lie in \mathbb{F} , and that $\deg p_A = \dim V$. A quick calculation

$$\det(SAS^{-1} - \lambda I) = \det(S(A - \lambda I)S^{-1}) = \det(S) \det(A - \lambda I) \det(S^{-1}) = \det(SS^{-1}) \det(A - \lambda I) = \det(A - \lambda I)$$

shows that the characteristic polynomial is the same regardless of the choice of basis in our matrix-representation and we can therefore speak about $p_F(\lambda)$, the characteristic polynomial of a linear map $F : V \rightarrow V$ without specifying a basis.

The **algebraic multiplicity** m_λ of an eigenvalue λ is the multiplicity of λ as a zero in $p_A(t)$, in other words we can factorize $p_A(t) = (t - \lambda)^{m_\lambda} q(t)$ where $q(\lambda) \neq 0$. On the other hand, the **geometric multiplicity** of λ is defined as $\dim \ker(A - \lambda I)$, the dimension of the λ -eigenspace.

Typically we expect the geometric and algebraic multiplicities to coincide for the eigenvalues of an operator. This was the case for the operator $P : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ from Example 3.2 above where we had $g_0 = 2 = m_0$ and $g_1 = 1 = m_1$. But there are exceptions:

Example 3.4. Let $A = \begin{pmatrix} 3 & 5 \\ 0 & 3 \end{pmatrix}$. The characteristic polynomial is $p_A(t) = (t - 3)^2$, so $\lambda = 3$ is an eigenvalue with algebraic multiplicity 2 since $t = 3$ is a double zero for $p_A(t)$. But solving $Av = 3v$ we notice that the only eigenvectors are multiples of $(1, 0)$, so $\ker(A - 3I) = \text{span}(1, 0)$ is only 1-dimensional, and the geometric multiplicity of $\lambda = 3$ is 1. In other words, $1 = g_3 < m_3 = 2$.

△

Proposition 3.5. For each eigenvalue λ of a linear map $F : V \rightarrow V$, its geometric multiplicity is less or equal than its algebraic multiplicity:

$$g_\lambda \leq m_\lambda.$$

¹⁴Solving a polynomial equation of degree ≥ 5 algebraically is not feasible in general, but there are good numeric methods for finding a good approximations, so this is not a problem in applications.

¹⁵Note that $\ker(A - \lambda I)$ also contains the zero-vector, so technically it consists of all eigenvectors of eigenvalue λ and the zero vector.

¹⁶Also called the secular polynomial. Some books define it instead as $\ker(\lambda I - A)$, this is the same up to a sign change. Sometimes a different variable is used in the polynomial, for example $p_A(t) = \det(A - tI)$. The variable-name is unimportant, $p_A(\lambda)$ and $p_A(t)$ is the same function.

Proof. Let V be n -dimensional and assume the geometric multiplicity of λ is $g_\lambda = m$. Then we can pick a basis of m vectors (v_1, \dots, v_m) in the eigenspace E_λ and extend this to a basis $\mathcal{B} = (v_1, \dots, v_n)$ of V . With respect to this latter basis, the matrix of F has block form

$$A = [F]_{\mathcal{B}} = \left(\begin{array}{ccc|ccc} \lambda & & & & & \\ & \ddots & & & & \\ & & \lambda & & & \\ \hline & & & 0 & & \end{array} \begin{array}{cc} B \\ \\ \\ C \end{array} \right)$$

where the top-left block is of size $m \times m$.

Since the characteristic polynomial is independent of the choice of basis we have

$$p_F(t) = p_A(t) = \det(A - tI) = (\lambda - t)^m \det(C - tI) = (\lambda - t)^m p_C(t),$$

which was obtained by expanding the determinant along each of the first m columns. This shows that $(\lambda - t)^m$ divides $p_F(t)$, so the algebraic multiplicity of λ is at least m . This completes the proof. \square

Note however that if λ is an eigenvalue, g_λ is at least 1 since there is always at least one eigenvector.

If a matrix A can be factored as $A = SDS^{-1}$ where D is a diagonal matrix, we say that A is **diagonalizable**. This is equivalent to saying that there is a basis for V consisting of eigenvectors for A . This means that the characteristic polynomial factors completely into linear factors, and that the algebraic and geometric multiplicities agree for all eigenvalues. Concretely, let D be a diagonal matrix with the eigenvalues $\lambda_1, \dots, \lambda_n$ on the diagonal (repeating according to algebraic multiplicities), and pick a basis v_1, \dots, v_n for V consisting of eigenvectors ordered in the same order as the eigenvalues, such that $Av_i = \lambda_i v_i$. Let $S = (v_1 \cdots v_n)$ be the matrix with these eigenvectors as columns. Then $S^{-1}AS = D$, or equivalently $SDS^{-1} = A$. A factorization like this is called a diagonalization of A , it corresponds to making a change of basis so that the new basis vectors are eigenvectors. This just means that geometrically, any diagonalizable linear map just stretches vectors with different factors along a number of axes.

One main difference between real and complex vector spaces when it comes to spectral theory is:

Theorem 3.6. *Each linear operator on a finite-dimensional complex vector space has an eigenvalue.*

Proof. By the fundamental theorem of algebra, every nonconstant polynomial with complex coefficients has a zero in \mathbb{C} . A zero of the characteristic polynomial is an eigenvalue. \square

Example 3.7. Let's diagonalize the linear map $F : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ with matrix $[F] = A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

Since $p_A(\lambda) = \lambda^2 + 1$, the eigenvalues are $\pm i$. To find the eigenvectors for $\lambda = i$ we solve the system $(A - iI)x = 0$:

$$\begin{cases} -ix_1 & -x_2 = 0 \\ x_1 & -ix_2 = 0 \end{cases} \Leftrightarrow x_1 - ix_2 = 0 \Leftrightarrow (x_1, x_2) = t(i, 1) \text{ where } t \in \mathbb{C}.$$

Similarly we get $(A + iI)x = 0 \Leftrightarrow x = t(1, i)$. We put the eigenvalues in a matrix D and corresponding eigenvectors as columns in a matrix S . We now have a diagonalization of A :

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = A = SDS^{-1} = \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \frac{1}{2} \begin{pmatrix} -i & 1 \\ 1 & -i \end{pmatrix}.$$

Note that the same matrix A can be viewed as the matrix of a linear map $G : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. This map G does *not* have any eigenvalues, since $p_A(\lambda) = \lambda^2 + 1$ has no zeroes in \mathbb{R} . Geometrically the map G corresponds to a rotation a quarter of a turn counter-clockwise in \mathbb{R}^2 .

\triangle

Proposition 3.8. *Eigenvectors corresponding to different eigenvalues are linearly independent.*

Proof. We prove the statement by contradiction. Let A be a matrix representing an operator, and let v_1, \dots, v_n be eigenvectors $Av_i = \lambda_i v_i$ where the eigenvalues λ_i are distinct, which satisfy linear dependence relation $\alpha_1 v_1 + \dots + \alpha_n v_n = 0$, where we may assume that all $\alpha_i \neq 0$ (otherwise just remove the corresponding vectors v_i).

We may also assume that n is minimal, such that there is no dependence relation of eigenvectors of A with fewer vectors involved. Then we apply $(A - \lambda_1 I)$ to the relation and obtain

$$(A - \lambda_1 I)(\alpha_1 v_1 + \dots + \alpha_n v_n) = \alpha_1(A - \lambda_1 I)v_1 + \dots + \alpha_n(A - \lambda_1 I)v_n = \alpha_1(\lambda_1 - \lambda_1)v_1 + \dots + \alpha_n(\lambda_n - \lambda_1)v_n = 0.$$

But this is another nontrivial linear dependence but with one less term since the coefficient of v_1 is now zero. This contradicts the minimality of n , and finishes the proof. \square

3.2 Complexification and realification

In the previous example we saw how a matrix A could represent both a linear map between real vector spaces $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, and a linear map between two complex vector spaces $\mathbb{C}^2 \rightarrow \mathbb{C}^2$. Let's try to formalize this idea.

Every complex number can be written uniquely as $a + bi$ with $a, b \in \mathbb{R}$. In other words, the complex numbers \mathbb{C} is a real vector space of dimension 2 with basis $(1, i)$.

Similarly, every vector in \mathbb{C}^2 can be written uniquely as $v + iv'$ where $v, v' \in \mathbb{R}^2$. For example:

$$\begin{pmatrix} 3 + 2i \\ 5 - i \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix} + i \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Following this idea we construct the *complexification* of a real vector space by considering formal pairs of real vectors.

Definition 3.9. Let V be a real vector space. We define the **complexification** of V to consist of all formal sums $v + iv'$ where $v, v' \in V$, where addition of such objects is defined in the natural way:

$$(v_1 + iv'_1) + (v_2 + iv'_2) := (v_1 + v_2) + i(v'_1 + v'_2),$$

and where multiplication by a complex number $a + bi$ on such an object is defined via:

$$(a + bi) \cdot (v + iv') := (av - bv') + i(av' + bv).$$

Then $V^{\mathbb{C}}$ is a complex vector space, it satisfies all the vector space axioms.

For example, the complexification of \mathbb{R}^2 is \mathbb{C}^2 , but the construction above is more general and works for any real vector space V .

Note that a basis for V is still a basis for $V^{\mathbb{C}}$, but complex coefficients are allowed in the latter space¹⁷.

Now, if $F : V \rightarrow W$ is a linear map between real vector spaces, we can define a map $F^{\mathbb{C}} : V^{\mathbb{C}} \rightarrow W^{\mathbb{C}}$ by

$$F^{\mathbb{C}}(v + iv') := F(v) + iF(v').$$

This construction has the effect witnessed before: if $[F] = A$ with respect to some choice of bases in V and W , then $[F^{\mathbb{C}}] = A$ too with respect to the same bases in $V^{\mathbb{C}}$ and $W^{\mathbb{C}}$.

Is the opposite construction possible? Can we from a complex vector space construct a corresponding real one? Yes, if V is a complex vector space, let $V_{\mathbb{R}}$ be the same set as V , where addition is defined the same way, and where multiplication by a scalar λ is defined the same way as in V when $\lambda \in \mathbb{R}$, but where it is *undefined* when λ is not real. Intuitively, just take the same vector space V but "forget" how to multiply vectors by non-real complex numbers¹⁸. This construction is called the **decomplexification** or **realification** of V . If $F : V \rightarrow W$ is a linear map between complex vector spaces, then we get a corresponding linear map between real vector spaces $F_{\mathbb{R}} : V_{\mathbb{R}} \rightarrow W_{\mathbb{R}}$, where $F_{\mathbb{R}}(v) = F(v)$.

Note that if (e_1, \dots, e_n) is a basis for the complex vector space V , it means that every vector $v \in V$ can be expressed uniquely as

$$v = \lambda_1 e_1 + \lambda_2 e_2 + \dots + \lambda_n e_n = (a_1 + ib_1)e_1 + (a_1 + ib_1)e_2 + \dots + (a_n + ib_n)e_n.$$

¹⁷Indeed, this is an alternative way to define complexification. If (e_1, \dots, e_n) is a basis for a real vector space V , then let $V^{\mathbb{C}}$ be the complex vector space with the same basis. It consists of all *complex* linear combinations of the basis vectors.

¹⁸In category theory, the function that maps each complex vector space V to $V_{\mathbb{R}}$ is called the *forgetful functor*.

This shows that $(e_1, ie_1, e_2, ie_2, \dots, e_n, ie_n)$ is a basis for the real vector space $V_{\mathbb{R}}$ since every v can be expressed uniquely as a linear combination of these vectors with *real* coefficients. Note that $\dim V_{\mathbb{R}} = 2 \cdot \dim V$.

3.3 Invariant subspaces

Let $F : V \rightarrow V$ be a linear operator, and let U be a subspace of V . We say that the subspace U is **F -invariant**, or invariant under F , if

$$u \in U \Rightarrow F(u) \in U.$$

In other words, we "stay in the subspace" if we apply F to a vector in the subspace. Since matrices describe linear maps we shall also talk about invariant subspaces for matrices.

If $U \subset V$ is a subspace, and $F : V \rightarrow W$ is a linear map, the restriction $F|_U : U \rightarrow W$ is the same map but with domain U .

But if $U \subset V$ is a subspace that is *invariant* under a map $F : V \rightarrow V$, we can consider its restriction to be a map from U to itself¹⁹.

An eigenspace for a linear map F is clearly an F -invariant subspace, and so is a sum of eigenspaces. But there are also other examples as illustrated in the following example:

Example 3.10. Let $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be the rotation a quarter of a turn in the positive direction around the axis $(1, 2, 2)$. Then $U_1 = \text{span}(1, 2, 2)$ is a one-dimensional F -invariant subspace, and the restriction $F|_{U_1} : U_1 \rightarrow U_1$ is the identity map since every vector in U_1 is mapped to itself by F . In the basis $((1, 2, 2))$ of U_1 , the matrix of $F|_{U_1}$ is the 1×1 -matrix (1) .

Similarly, the plane $U_2 : x + 2y + 2z = 0$ is also a F -invariant subspace, since every vector of U_2 is mapped to another vector in U_2 when it is rotated around $(1, 2, 2)$. If we pick $((2, -2, 1), (2, 1, -2))$ as a basis for U_2 , the matrix of $F|_{U_2}$ is $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Note that U_2 contains no eigenvectors of F .

△

Example 3.11. The differential operator $D : \mathcal{P} \rightarrow \mathcal{P}$ is given by $D(p(x)) = p'(x)$. Then \mathcal{P}_n , the polynomials of degree $\leq n$, is a D -invariant subspace for each n .

△

We know that many maps can be represented by diagonal matrices. What about maps that can't? Before we prove the Jordan theorem we need the following preliminary result.

Theorem 3.12. *Let $F : V \rightarrow V$ be linear operator on a finite-dimensional complex vector space V . Then there exists a basis for V for which the matrix of F is upper triangular. The elements on the diagonal in this triangular matrix are the eigenvalues for F , counting multiplicities.*

Proof. We proceed by induction on $\dim V$, the statement is trivial if $\dim V = 1$ since every 1×1 -matrix is upper triangular. Let $\dim V = n$ and assume the statement is true for all vector spaces of dimension $< n$. We know that F has an eigenvalue, so assume $F(u_1) = \lambda u_1$. Pick any $(n-1)$ -dimensional subspace U that doesn't contain u_1 ; then $V = \text{span}(u_1) \oplus U$. Pick a basis (u_2, \dots, u_n) of U . Then with respect to the basis (u_1, \dots, u_n) of V , the matrix of F looks like

$$[F] = \left(\begin{array}{c|ccc} \lambda & b_2 & \cdots & b_n \\ \hline 0 & & & \\ \vdots & & A & \\ 0 & & & \end{array} \right)$$

where A is the matrix of the projection onto U of the restriction $F|_U$. By the induction hypothesis, for this operator there exists a basis for U for which A becomes upper triangular. Together with the first basis vector u_1 , we get a basis for V for which the matrix for F is upper triangular. Let $T = (t_{ij})$ be this upper triangular matrix, then since the determinant of a triangular matrix is the product of the

¹⁹Technically, to get a map $U \rightarrow U$ we should compose $F|_U$ with the projection onto the subspace U , and form $P_U \circ F|_U$, but it is standard to just write $F|_U$ when the context is understood.

diagonal elements, we have $\det(T - \lambda I) = (t_{11} - \lambda) \cdots (t_{nn} - \lambda)$ is the characteristic polynomial of T , and therefore of F , and its roots t_{11}, \dots, t_{nn} are the diagonal elements of T , so these are the eigenvalues of F . \square

Corollary 3.13. *Let A be an operator (or a square matrix) with eigenvalues $\lambda_1, \dots, \lambda_n$ (repeating according to algebraic multiplicity). Then the trace of A is the sum of the eigenvalues, and the determinant of A is the product of the eigenvalues:*

$$\operatorname{tr}(A) = \lambda_1 + \dots + \lambda_n \quad \text{and} \quad \det(A) = \lambda_1 \cdots \lambda_n.$$

Proof. We know that trace and determinant is independent of the choice of basis, so use Theorem 3.12 to find a basis for which the matrix of A is upper triangular with the eigenvalues on the diagonal. Then the result follows. \square

This result is usually proved in a first course only for diagonalizable operators, now we see that it holds in general for operators on complex vector spaces.

3.4 Matrix polynomials

There is a natural way to "plug in" a square matrix into a polynomial.

Definition 3.14. Let $p(t) = a_n t^n + \dots + a_1 t + a_0$ be a polynomial with coefficients $a_i \in \mathbb{F}$, and let A be a square matrix with coefficients in the same field \mathbb{F} . Then we define

$$p(A) = a_n A^n + \dots + a_1 A + a_0 I.$$

Example 3.15. Let $p(t) = t^7 + 12t^4 - t^3 + 2t^2 + 5t + 3$ and $N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Then

$$p(N) = N^7 + 12N^4 - N^3 + 2N^2 + 5N + 3I = 2N^2 + 5N + 3I = \begin{pmatrix} 3 & 5 & 2 \\ 0 & 3 & 5 \\ 0 & 0 & 3 \end{pmatrix}$$

here the calculation was made easy by the fact that N is nilpotent, $N^k = 0$ for $k > 2$, so the first terms disappeared.

\triangle

Note that $p(A)$ is a square matrix of the same size as A , and that $P(A)$ commutes with A since each term does. Moreover, if $B = SAS^{-1}$, we get $p(B) = Sp(A)S^{-1}$. For this reason it also makes sense to define $p(F)$ where $F : V \rightarrow V$ is an operator of an \mathbb{F} -vector space.

For a given $n \times n$ -matrix A , can we always find a nonzero polynomial $p(t)$ such that $p(A) = 0$? Yes - consider the matrices $I, A, A^2, \dots, A^{n^2}$. These are $n^2 + 1$ vectors in an n^2 -dimensional vector space $\operatorname{Mat}_n(\mathbb{F})$, therefore they are linearly dependent and there are scalars λ_k such that

$$\sum_{k=0}^{n^2} \lambda_k A^k = 0$$

so with $p(t) = \sum_{k=0}^{n^2} \lambda_k t^k$ we have $p(A) = 0$.

Can we find find such a polynomial of smaller degree?

Example 3.16. Let $p(t) = t^2 - 5t - 2$ and $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then

$$p(A) = A^2 - 5A - 2I = \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix} - 5 \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} - 2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

So in this case $p(A)$ is the zero matrix. Note that the characteristic polynomial of A is $\det(A - \lambda I) =$

$$\mid \lambda^2 - 5\lambda - 2.$$

△

The example above shows an example where a *matrix satisfies its characteristic equation*: $p_A(A) = 0$. This is in fact always true, it's a famous result of linear algebra:

Theorem 3.17. (The Cayley-Hamilton theorem) *If A is a square matrix with characteristic polynomial $p_A(\lambda) = \det(A - \lambda I)$, then*

$$p_A(A) = 0.$$

The theorem holds over any field, but we prove it only for the complex numbers below:

Proof. Use Theorem 3.12 to find a matrix S and an upper triangular matrix T (with the eigenvalues $\lambda_1, \dots, \lambda_n$ of A on the diagonal) such that $S^{-1}AS = T$. We have previously shown that similar matrices have the same characteristic polynomial, so $p_A(\lambda) = p_T(\lambda)$, and

$$p_A(A) = p_T(A) = p_T(STS^{-1}) = Sp_T(T)S^{-1}$$

so it suffices to show that $p_T(T) = 0$ for an upper triangular matrix T .

Now factor $p_T(\lambda)$ completely over \mathbb{C} , we know that the eigenvalues of T are its diagonal entries, so

$$p_T(\lambda) = (t_{11} - \lambda) \cdots (t_{nn} - \lambda).$$

Then $p_T(T) = (t_{11}I - T) \cdots (t_{nn}I - T)$, and by taking successive products from the right we see that $p_T(T)v = 0$ for every vector v , as illustrated below when $n = 3$:

$$\begin{aligned} p_T(T)v &= (t_{11}I - T)(t_{22}I - T)(t_{33}I - T)v = \begin{pmatrix} 0 & * & * \\ 0 & * & * \\ 0 & 0 & * \end{pmatrix} \begin{pmatrix} * & * & * \\ 0 & 0 & * \\ 0 & 0 & * \end{pmatrix} \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \\ &= \begin{pmatrix} 0 & * & * \\ 0 & * & * \\ 0 & 0 & * \end{pmatrix} \begin{pmatrix} * & * & * \\ 0 & 0 & * \\ 0 & 0 & * \end{pmatrix} \begin{pmatrix} * \\ * \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & * & * \\ 0 & * & * \\ 0 & 0 & * \end{pmatrix} \begin{pmatrix} * \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

where in the calculation above, $*$ means some arbitrary number. Each multiplication gives one more zero in the resulting vector. This shows that $p_T(T)v = 0$ for all v , which means that T is the zero matrix. In light of the above remarks, this completes the proof. □

Example 3.18. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Let us find a polynomial p such that $p(A) = A^{-1}$. Since $p_A(\lambda) = \lambda^2 - 5\lambda - 2$ we know by Cayley-Hamilton that $p_A(A) = 0$, so solving for I we get

$$A^2 - 5A - 2I = 0 \Leftrightarrow I = \frac{1}{2}(A - 5I)A \Leftrightarrow A^{-1} = \frac{1}{2}(A - 5I).$$

This idea works in general whenever A is invertible, because then 0 is not an eigenvalue, and the constant term in p_A is nonzero.

△

The Cayley-Hamilton also provides a quicker way to evaluate a polynomial of high degree at A without computing many matrix products. For example, in our previous example we had $p_A(t) = t^2 - 5t - 2$, so to evaluate $p(A)$ for some complicated polynomial $p(t)$, first use standard polynomial division to find polynomials $q(t)$ and $r(t)$ with $\deg r(t) < 2$ such that

$$p(t) = p_A(t)q(t) + r(t).$$

Then by Cayley-Hamilton we have

$$p(A) = p_A(A)q(A) + r(A) = 0 \cdot q(A) + r(A) = r(A),$$

so we only have to evaluate $r(A)$ where r has degree ≤ 1 .

Minimal polynomial

Let A be an $n \times n$ matrix. Cayley-Hamilton says that there exists a polynomial of degree n (the characteristic polynomial) satisfying $p(A) = 0$, but a polynomial of even lower degree may exist.

Definition 3.19. Let A be a square matrix. The **minimal polynomial** of A is the monic polynomial $m_A(t)$ of lowest degree for which $m_A(A) = 0$.

One can also define the minimal polynomial of any operator on a finite-dimensional vector space. The adjective *monic* just means that the coefficient of the highest degree term is 1.

Proposition 3.20. *The minimal polynomial exists and is unique.*

Proof. Cayley-Hamilton shows that there exists some monic polynomial p annihilating A (meaning $p(A) = 0$), namely $\pm p_A$. Now if two monic polynomials p_1 and p_2 of the same minimal degree n both annihilate A , then $(p_1 - p_2)(A) = p_1(A) - p_2(A) = 0 - 0 = 0$, so $p_1 - p_2$ also annihilates A , and it has lower degree than n (and can be made monic by dividing by its leading coefficient). This contradicts minimality unless $p_1 = p_2$ which shows uniqueness. \square

Proposition 3.21. *The minimal polynomial divides any polynomial that annihilates A :*

$$p(A) = 0 \Rightarrow p(t) = m_A(t)q(t).$$

In particular, the minimal polynomial divides the characteristic polynomial.

Proof. Assume $p(A) = 0$. Divide $p(t)$ by $m_A(t)$ using polynomial division. We obtain a polynomial equation $p(t) = q(t)m_A(t) + r(t)$ where $\deg r(t) < \deg m_A$ or $r(t) = 0$. Replacing t by A we get $p(A) = q(A) \cdot m_A(A) + r(A)$ and $0 = q(A) \cdot 0 + r(A)$, so $r(A) = 0$. But then r is the zero-polynomial, otherwise it would be a polynomial of lower degree than m_A that annihilates A , which would contradict the minimality of m_A . We conclude that $p(t) = q(t)m_A(t)$ so $m_A(t)$ divides $p(t)$. \square

Proposition 3.22. *The characteristic and minimal polynomial have the same zeros:*

$$p_A(\lambda) = 0 \Leftrightarrow m_A(\lambda) = 0.$$

Proof. Any zero of m_A is a zero of p_A since $m_A | p_A$. On the other hand, let λ be a zero of p_A . Then λ is an eigenvalue and there is some eigenvector v with $Av = \lambda v$. Then $0 = m_A(A)v = m_A(\lambda)v$ so $m_A(\lambda)$ is zero too. \square

Example 3.23. Let us find the minimal polynomial of $A = \begin{pmatrix} 1 & 1 & 1 \\ -4 & 4 & 3 \\ -4 & 1 & 6 \end{pmatrix}$. We compute and factor the characteristic polynomial: $p_A(t) = -(t-3)^2(t-5)$. Now since $m_A(t)$ divides $p_A(t)$ and still has 3 and 5 as zeros, there are only two options: either $m_A(t) = -p_A(t) = (t-3)^2(t-5)$ or $m_A(t) = (t-3)(t-5)$. We test whether the second option annihilates A , but find that $(A-3I)(A-5I)$ is not the zero-matrix. Therefore $m_A(t) = (t-3)^2(t-5)$.

\triangle

If we know the eigenvalues of some matrix A , what can be said about the eigenvalues of $p(A)$ where p is some polynomial? The answer is given by the *spectral mapping theorem*.

Theorem 3.24. Spectral mapping theorem. *For any polynomial p , we have*

$$\sigma(p(A)) = p(\sigma(A)).$$

In other words, λ is an eigenvalue for A if and only if $p(\lambda)$ is an eigenvalue for $p(A)$.

Proof. Note that $p(\sigma(A))$ is defined as $\{p(\lambda) \mid \lambda \in \sigma(A)\}$. One direction is easy: if $Av = \lambda v$ and $p(t) = \sum a_k t^k$, then $p(A)v = \sum a_k A^k v = \sum a_k \lambda^k v = p(\lambda)v$. The other direction is left as an exercise. \square

Intuitively, the spectrum of A is some finite set of points in \mathbb{C} . The theorem says that if we apply the polynomial p to each of these points we get the spectrum of the operator $p(A)$.

4 Jordan normal form

Our goal of this section is to show that for *any* linear operator on a complex finite-dimensional vector space, there is a basis such that the matrix for the operator has a particular canonical format called the *Jordan form*²⁰. We shall soon prove this, and we shall discuss the algorithm for *Jordanizing* a matrix, but first, let's investigate the properties of matrices in this form.

4.1 Properties of matrices on Jordan form

Definition 4.1. The **Jordan block** of size n and with eigenvalue λ is defined as the $n \times n$ -matrix

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

In other words, $J_n(\lambda)$ has λ 's on the diagonal, and ones on the super-diagonal (one step over the diagonal), and zeros elsewhere.

Example 4.2. The matrix

$$J = J_3(5) = \begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{pmatrix}$$

is a Jordan block. We note that $p_J(t) = \det(J - tI) = (5 - t)^3$ so the only eigenvalue is 5, and since

$$J - 5I = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \text{ only multiples of } \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

are eigenvectors, so the geometric multiplicity of $t = 5$ is 1 and J is not diagonalizable.

△

The same goes in general, the characteristic polynomial of $J = J_n(\lambda)$ is $p_J(t) = \det(J - tI) = (\lambda - t)^n$, and the eigenspace $E_\lambda = \ker(J - \lambda I)$ is spanned by the first standard basis vector e_1 . In particular, the geometric multiplicity g_λ is 1, so it is not diagonalizable (except for the trivial case $n = 1$).

Definition 4.3. A matrix J is said to be in **Jordan form** if it is a block-diagonal matrix where each block is a Jordan block. In other words,

$$J = \text{diag}(J_{n_1}(\lambda_1), J_{n_2}(\lambda_2), \dots, J_{n_k}(\lambda_k)).$$

Note that the blocks may have different sizes, and that some of the λ_i can coincide.

Recall that direct sums of linear maps correspond to block-diagonal matrices. For this reason another common notation for the Jordan-matrix in the definition is

$$J = J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \dots \oplus J_{n_k}(\lambda_k).$$

Example 4.4. The following matrix is in Jordan form, for clarity it is common to omit off-diagonal

²⁰Named after Camille Jordan who first stated what is now known as the Jordan decomposition Theorem in 1870.

zeros and to draw boxes to indicate the Jordan blocks:

$$J = \begin{pmatrix} 5 & 1 & 0 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} \boxed{\begin{matrix} 5 & 1 \\ & 5 & 1 \\ & & 5 \end{matrix}} & & \\ & \boxed{\begin{matrix} 2 & 1 \\ & 2 \end{matrix}} & \\ & & \boxed{2} \end{pmatrix} = J_3(5) \oplus J_2(2) \oplus J_1(2)$$

Here we see that the characteristic polynomial is $p_J(t) = (t - 5)^3(t - 2)^3$ so the eigenvalues are 5 and 2. Each Jordan-block corresponds to an eigenvector, in this example we see that e_1 is an eigenvector with eigenvalue 5, and that e_4 and e_6 both are eigenvectors of eigenvalue 2.

△

Note in particular that any Jordan block is in Jordan form (with a single block), and that any diagonal matrix is in Jordan form (all the blocks have size 1×1).

How can we determine the minimal polynomial for a matrix in Jordan form? Let us first consider an example where all the eigenvalues coincide:

Example 4.5. For the matrix J below, we have $p_J(t) = (2 - t)^7$, so the minimal polynomial has form $(t - 2)^n$ for some $1 \leq n \leq 7$, we plug in J and compute:

$$J = \begin{pmatrix} \boxed{\begin{matrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{matrix}} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \boxed{\begin{matrix} 2 & 1 \\ 0 & 2 \end{matrix}} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \boxed{\begin{matrix} 2 & 1 \\ 0 & 2 \end{matrix}} \end{pmatrix} \Rightarrow (J - 2I)^n = \begin{pmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}^n & & \\ & \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^n & \\ & & \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^n \end{pmatrix}$$

and we see that this is zero if and only if $n \geq 3$, because then the *largest Jordan block* is annihilated by the polynomial, so $m_J(t) = (t - 2)^3$.

△

We see that the analogous argument works in general: For each zero λ of $p_J(t)$, the corresponding exponent of $(t - \lambda)$ in $m_J(t)$ is the *size of the largest Jordan block of eigenvalue λ* .

For example, in Example 4.4, the largest Jordan block for eigenvalue 5 had size 3, and for eigenvalue 2, the largest block had size 2. So the minimal polynomial is $m_J(t) = (t - 5)^3(t - 2)^2$. Here it is easy to verify that this polynomial annihilates J : we have

$$\begin{aligned} m_J(J) &= (J - 5I)^3(J - 2I)^2 = \begin{pmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}^3 & & \\ & \begin{bmatrix} -3 & 1 \\ 0 & -3 \end{bmatrix}^3 & \\ & & [-3]^3 \end{pmatrix} \begin{pmatrix} \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}^2 & & \\ & \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^2 & \\ & & [0]^2 \end{pmatrix} \\ &= \begin{pmatrix} 0 & | & | \\ \hline & * & | \\ \hline & & * \end{pmatrix} \begin{pmatrix} * & | & | \\ \hline & 0 & | \\ \hline & & 0 \end{pmatrix} = \begin{pmatrix} 0 & | & | \\ \hline & 0 & | \\ \hline & & 0 \end{pmatrix} = 0. \end{aligned}$$

To summarize the above discussion, we collect what we know about a matrix in Jordan form.

- The eigenvalues is on the diagonal of J , repeating according to algebraic multiplicity. In other words, the algebraic multiplicity of λ is the sum of the sizes of all λ -Jordan blocks in J .

- The geometric multiplicity of λ is the number of λ -Jordan blocks. The top-left position of each Jordan block corresponds to an eigenvector, if a Jordan-block $J_n(\lambda)$ sits in J with top-left corner in position (i, i) , then e_i is an eigenvector of eigenvalue λ .
- If $p_J(t) = \pm(t - \lambda_1)^{n_1} \cdots (t - \lambda_k)^{n_k}$, the minimal polynomial is $m_J(t) = (t - \lambda_1)^{m_1} \cdots (t - \lambda_k)^{m_k}$, where m_i is the size of the largest λ_i -Jordan block in J .

Algebraic multiplicities, geometric multiplicities, characteristic- and minimal polynomials are invariant under a change of basis, so if we know these invariants for an arbitrary matrix A , and we are trying to write $J = S^{-1}AS$, we get some information about the shape of J .

4.2 Structure theory for nilpotent operators

Our goal in this section is to from a given a linear map $F : V \rightarrow V$ find a basis for which the matrix of F is in Jordan form. This can be somewhat complicated, so let us first restrict ourselves to *nilpotent* maps F .

If F is nilpotent, then $F^d = 0$ for some minimal d which is the nilpotency degree of F . Then the minimal polynomial must be $m_F(t) = t^d$ which shows that zero is the only eigenvalue of F .

We are looking for a special type of basis for V which clearly reveals structure of F :

Definition 4.6. Let $F : V \rightarrow V$ be nilpotent. A **string** (for F) in V is a sequence of nonzero vectors (v_1, \dots, v_n) , such that $F(v_i) = v_{i-1}$ and $F(v_1) = 0$, or visually

$$v_n \mapsto \cdots \mapsto v_2 \mapsto v_1 \rightarrow 0.$$

Here we call v_n the **first vector**, and we call v_1 the **last vector** of the string, and the **length** of the string is n . We shall sometimes draw simpler arrows and sometimes omit the vectors when drawing strings.

A **string basis** for V is a basis for V which is a union of strings.

Strings are sometimes called *Jordan chains* or just *chains*, I will use the word string when referring to nilpotent operators and reserve the other words for the more general case that we will investigate later.

Example 4.7. Suppose that a linear map $F : \mathbb{C}^8 \rightarrow \mathbb{C}^8$ has a string basis (e_1, \dots, e_8) which looks like in the left diagram below. In this basis, the matrix of F will have the form to the right:

$$\begin{array}{l}
 e_3 \rightarrow e_2 \rightarrow e_1 \rightarrow 0 \\
 e_5 \rightarrow e_4 \rightarrow 0 \\
 e_7 \rightarrow e_6 \rightarrow 0 \\
 e_8 \rightarrow 0
 \end{array}
 \quad
 [F] =
 \begin{pmatrix}
 \boxed{\begin{matrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix}} & \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \\
 \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \boxed{\begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix}} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix} \\
 \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \boxed{0}
 \end{pmatrix}$$

Here we note that $[F] = J_3(0) \oplus J_2(0) \oplus J_2(0) \oplus J_1(0)$ is in Jordan form, and that the nilpotency degree of F is 3, which corresponds to the length of the longest string. This also shows that $m_F(t) = t^3$. We see that each string corresponds to a Jordan block, and that the rightmost vectors of the strings are the eigenvectors with eigenvalue 0.

The images and kernels $\text{Im}(F^k)$ and $\ker(F^k)$ can easily be visualized from the string-diagram. For example, $\ker(F)$ consists of vectors that are mapped to zero, so this subspace is spanned by the rightmost vectors of all the strings: $\ker(F) = \text{span}(e_1, e_4, e_6, e_8)$. Similarly, the subspace $\ker(F^2)$ is spanned by vectors which are mapped to zero when F is applied twice, so this is spanned by all the standard basis vectors except e_3 . Clearly $\ker(F^3) = \mathbb{C}^8$.

Dually, the image of F is spanned by all basis vectors except the first (leftmost) vectors of each

| string, $\text{Im}(F) = \text{span}(e_1, e_2, e_4, e_6)$, while $\text{Im}(F^2) = \text{span}(e_1)$. Clearly $\text{Im}(F^3) = 0$.

△

Lemma 4.8. *If the last vectors of a set of strings are linearly independent (the eigenvectors with eigenvalue 0), then so are the whole strings. More precisely, if $\mathcal{S} = \{v_k^{(i)}\}$ are strings for F with $F(v_1^{(i)}) = 0$ and $F(v_k^{(i)}) = v_{k-1}^{(i)}$, then if the last vectors $\{v_1^{(i)}\}$ is a linearly independent set, then so is \mathcal{S} .*

Proof. Following the same idea as in the proof of Proposition 3.8, assume a minimal nontrivial linear combination of string-vectors is zero. Then applying the operator F shifts all the vectors in the linear combination one step to the right, so by applying F until a vector in the linear combination is mapped to zero, we have obtained a linear combination with fewer elements which is zero, contradicting the minimality of our original linear combination. □

Theorem 4.9. *For any nilpotent map $F : V \rightarrow V$ there exists a string basis, and with respect to this basis, the matrix $[F]$ is in Jordan form with zeroes on the diagonal.*

Proof. We proceed by induction on $\dim V$, the base case $\dim V = 0$ is obvious.

Now assume $F : V \rightarrow V$ is nilpotent and $\dim V = n$, and that the statement is true for all nilpotent maps on vector spaces of dimension lower than n . Now if F is the zero map the statement is obvious (any basis is a string basis). Otherwise, $0 < \dim \ker F < n$, so by rank-nullity $\text{Im}(F)$ is a proper subspace. Consider the restriction $F|_{\text{Im}(F)} : \text{Im}(F) \rightarrow \text{Im}(F)$, by the induction hypothesis there exists a string basis for this map, denote this basis

$$\{e_j^{(i)} \mid 1 \leq j \leq k_i\}$$

where the top index indicate the different strings, and k_i is the length of string number i , and where $F(e_j^{(i)}) = e_{j-1}^{(i)}$ and $F(e_1^{(i)}) = 0$. Consider the leftmost vectors $e_{k_i}^{(i)}$ in these strings, they belong to $\text{Im}(F)$, so we can extend each string westwards by picking $e_{k_i+1}^{(i)} \in V$ such that $F(e_{k_i+1}^{(i)}) = e_{k_i}^{(i)}$. Now our extended string-vectors are linearly independent, and they span a subspace $U \subset V$ containing $\text{Im}(F)$. Now by construction, $F(U) = \text{Im}(F)$. The subspace U might not be all of V , so we extend U by vectors v_1, \dots, v_s to a basis of V . But $F(v_i)$ belongs to $\text{Im}(F) = F(U)$, so $F(v_i) = F(u_i)$ for some $u_i \in U$. Now take $w_i = v_i - u_i$. Then the w_i still span a complement to U , and we have $F(w_i) = F(v_i) - F(u_i) = 0$, so $w_i \in \ker(F)$. But then take our string basis for U and adjoin the w_i as strings of length 1, then we have a string basis for V . □

If you want to understand the proof, it might be instructive to study how it applies to the map:

$$\begin{array}{c} e_2 \mapsto e_1 \mapsto 0 \\ \nearrow \\ e_3 \end{array}$$

Here the image is spanned by e_1 , we extend the only string to a string of length 2: $e_2 \mapsto e_1 \mapsto 0$. Then $U = \text{span}(e_1, e_2)$ is not all of V , so we extend by $v = e_3$, and note that $F(e_3) = e_1 \in F(U)$, and we pick an element w of U that is also mapped to e_1 , let's take $u = e_2$. Then $w = v - u = e_3 - e_2$ is mapped to zero, and is a string of length 1. So in this case our derived string basis would be $(e_1, e_2, e_3 - e_2)$.

Corollary 4.10. *Let A be a nilpotent matrix. Then there exists an invertible matrix S and a matrix J in Jordan-form (with zeros on the diagonal) such that $S^{-1}AS = J$. The matrix J is unique up to permutation of the Jordan blocks.*

Proof. Use the Theorem 4.9 to find a string basis, and let S be the matrix with the string-basis vectors as columns in order string by string from the right side of each string, then clearly $S^{-1}AS$ is in Jordan form. For the uniqueness claim, we note that the sequence of numbers $\dim \ker(A^k) = \dim \ker(J^k)$ will determine uniquely the number of strings and the length of strings in the string basis, and therefore the shape of the Jordan form up to permutation. □

Since the Jordan form of nilpotent operators is unique we shall speak of *the* Jordan form of a nilpotent operator²¹.

²¹Usually we consider two Jordan forms to be the same if they contain the same Jordan blocks in different order, if we define some arbitrary order on \mathbb{C} and on the size of blocks, the Jordan form would be truly unique. Note however that while J is unique, there may be several choices for the actual vectors of the string basis (the columns of S).

Example 4.11. Let us classify nilpotent operators on a 5-dimensional vector space V . The corollary above guarantees that any nilpotent operator is similar to a matrix in Jordan form, so we need only find all Jordan-forms of 5×5 matrices where the diagonal elements are zero. The block-partition of the Jordan matrix corresponds to a partition of the integer 5 as a sum of positive integers, and there are seven such partitions:

$$(5), (4+1), (3+2), (3+1+1), (2+2+1), (2+1+1+1), (1+1+1+1+1).$$

Here are the corresponding Jordan forms:

The corollary guarantees that for any nilpotent map F on a 5-dimensional vector space, there exists a basis for which the matrix of $[F]$ is one (and only one) of the matrices above. Alternatively stated, any nilpotent matrix A is similar to exactly one of the matrices above: $A = SJS^{-1}$ where J is one of the seven matrices above.

△

Given a nilpotent operator, how do you actually find a string basis algorithmically? Before going hunting for basis vectors, a good starting point is determining the number of strings and their lengths. This determines the Jordan form of the operator.

Lemma 4.12. For a nilpotent operator (or matrix) $A : V \rightarrow V$, define $n_k = \dim \ker(A^k)$, and let d_k be the number of strings of length k in any string basis, or equivalently, the number of Jordan blocks of size k in the Jordan form of A . Then for $k \geq 1$,

$$d_k = 2n_k - n_{k-1} - n_{k+1}.$$

Proof. Note $n_0 = 0$ and that $n_k = \dim V$ for all $k \geq m$ where m is the nilpotency-degree of A , and that n_0, n_1, \dots, n_m is a strictly increasing integer sequence. Note also that the maximum length of a string is m , so $d_k = 0$ for $k > m$.

Visualize a generic string-diagram and consider what the different kernels $\ker(A^k)$ look like in terms of the strings. Now $n_1 = n_1 - n_0 = \dim \ker(A)$ equals the total number of strings, or in other words, the number of strings of length ≥ 1 . Similarly, $n_2 - n_1$ equals the number of strings of length ≥ 2 . By this line of reasoning, $n_i - n_{i-1}$ equals the number of strings of length $\geq i$. But then the number of strings of length *exactly* k is equal to the number of strings of length $\geq k$ minus the number of strings of length $k+1$, so

$$d_k = \underbrace{(n_k - n_{k-1})}_{\text{\#strings of length } \geq k} - \underbrace{(n_{k+1} - n_k)}_{\text{\#strings of length } \geq k+1} = 2n_k - n_{k-1} - n_{k+1}.$$

□

Example 4.13. Let's say that we have a matrix $A \in \text{Mat}_{11}(\mathbb{C})$ and we have found that $A^4 = 0$, and that $\text{rank}(A) = 7, \text{rank}(A^2) = 3, \text{rank}(A^3) = 1$. Let us determine the Jordan form of A from this information.

Let $n_k = \dim \ker(A^k)$ as in the lemma. By rank-nullity we get $n_0 = 0, n_1 = 11 - 7 = 4, n_2 = 11 - 3 = 8, n_3 = 11 - 1 = 10$, and $n_k = 11$ for $k > 4$. By the lemma, the numbers d_k of strings

can be computed as $d_k = 2n_k - n_{k-1} - n_{k+1}$, which gives

$$d_1 = 0, d_2 = 2, d_3 = 1, d_4 = 1,$$

and $d_k = 0$ for $k \geq 4$.

So we conclude that there are 2 strings of length 2, 1 strings of length 3, and 1 chain of length 4, so the Jordan form for A is:

$$S^{-1}AS = J_2(0) \oplus J_2(0) \oplus J_3(0) \oplus J_4(0).$$

△

Now, to find the actual vectors in a string basis, it is easiest to start with the *first vectors of the longest strings*.

Suppose that we know for some nilpotent map A that there are 2 strings of length 3, and no longer strings. Some starting vectors for these strings should lie in²² $\ker(A^3) \setminus \ker(A^2)$, so we pick two vectors v and w that span a complement²³ to $\ker(A^3)$ in $\ker(A^2)$. Then we apply A to get the rest of these strings:

$$v \mapsto Av \mapsto A^2v \mapsto 0 \text{ and } w \mapsto Aw \mapsto A^2w \mapsto 0.$$

Now we proceed to find the strings of length 2 by finding their first vectors which should lie in $\ker(A^2) \setminus \ker(A)$. However, we should also be careful not to pick any vector which is linearly dependent the six vectors v, Av, A^2v, w, Aw, A^2w from our previously chosen strings. Proceeding like this eventually produces a string basis.

In light of the discussion above, here is an algorithm²⁴ for finding a string basis for a nilpotent matrix (or operator).

Algorithm 4.14. To find a string basis for a nilpotent matrix A , do the following:

1. Write down A^k and find a basis in each subspace $\ker(A^k)$ until $A^m = 0$. Let $n_k = \dim \ker(A^k)$.
2. Find $d_k = 2n_k - n_{k+1} - n_{k-1}$, the number of strings of length k .
3. Sketch a string-diagram and write down the corresponding Jordan matrix J .
4. For each string-length k , from longest to shortest do:
 - (a) Let \mathcal{B} be the set of previously chosen vectors (\mathcal{B} is empty in the first step)
 - (b) Pick vectors v_1, \dots, v_{d_k} in $\ker(A^k)$ that are linearly independent both to $\ker(A^{k-1})$ and to the previously picked vectors in \mathcal{B} . These vectors will be first vectors in strings of length k .
 - (c) Compute the rest of the strings $v_i, Av_i, A^2v_i \dots$ until this is zero, adjoin all these nonzero vectors to \mathcal{B} .
 - (d) Decrease k by 1 and repeat until $k = 0$.
5. \mathcal{B} now contains a string basis. Order its vectors string by string from right to left, in the same order as your string diagram (and Jordan matrix J). Let S be the matrix with the string-basis vectors as columns in this order.
6. Verify that $SJS^{-1} = A$.

In step 1, things are easier if pick a basis in each successive $\ker(A^{k+1})$ by extending a basis from $\ker(A^k)$. If we know that the matrix S is invertible, it is easier to verify that $SJ = AS$ in the last step, even faster is to just verify that for each column s_i of S , we have either $As_i = 0$ or $As_i = s_{i-1}$.

The algorithm may seem complicated, especially step 4b. However, for small matrices some steps are quite obvious, let's illustrate:

²²Recall that $A \setminus B$ means the set-difference, it meaning the elements of A that are not in B . Don't confuse it with our notation V/U for quotient-spaces.

²³In other words, extend a basis of $\ker(A^2)$ to $\ker(A^3)$

²⁴Several variations of this can be found in the literature.

Example 4.15. Let $A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$. This gives $A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ and $A^3 = 0$. We see immediately that

$$\ker(A) = \text{span}(e_2, e_1 - e_4) \text{ and } \ker(A^2) = \text{span}(e_1, e_2, e_4).$$

So $(n_1, n_2, n_3) = (2, 3, 4)$ and $(d_1, d_2, d_3) = (1, 0, 1)$, so we are looking for two strings, one of length 1 and one of length 3. The Jordan form will therefore be $J_3(0) \oplus J_1(0)$.

Following the algorithm we are first looking for one vector v in $\ker(A^3) \setminus \ker(A^2) = \mathbb{C}^4 \setminus \text{span}(e_1, e_2, e_4)$. We pick $v = e_3$ as the first vector of this string. The rest of the string will be $Ae_3 = e_4$ and $Ae_4 = e_2$, and as we expect, this last vector is indeed mapped to 0 by A . So this is our string of length 3. There are no strings of length 2, so we look instead at the last string of length 1. Now we should pick a vector in $\ker(A^1) \setminus \ker(A^0) = \ker(A) \setminus \{0\}$ that is also linearly independent to our previously chosen vectors (e_3, e_4, e_2) . Since $\ker(A) = \text{span}(e_2, e_1 - e_4)$, a natural choice is to pick $e_1 - e_4$, this is indeed mapped to zero.

Now we have our string basis $\mathcal{B} = (e_2, e_4, e_3, e_1 - e_4)$, where the order matches our chosen Jordan form. We put these string basis vectors as columns of a matrix S , and we take J to be the corresponding Jordan matrix we found above:

$$S = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \quad J = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Now we can verify that we indeed have $S^{-1}AS = J$.

△

4.3 Jordan chains and Jordanization

Once we know how to deal with nilpotent matrices, the rest of the theory is not hard.

Generalized eigenspaces

Recall that the eigenspace E_λ for a linear operator (matrix) $A : V \rightarrow V$ consists of all vectors v for which $(A - \lambda I)v = 0$. *Generalized eigenvectors* for the eigenvalue λ are vectors that are *eventually* mapped to zero by $(A - \lambda I)$.

Definition 4.16. Let $A : V \rightarrow V$ be an operator (or matrix). A nonzero vector v is called a **generalized eigenvector** for the eigenvalue λ if there exists a positive integer n such that $(A - \lambda I)^n v = 0$. The set of all generalized eigenvectors for the eigenvalue λ (and the zero vector) is called the **generalized eigenspace** for λ , we denote it by \tilde{E}_λ :

$$\tilde{E}_\lambda = \{v \in V \mid (A - \lambda I)^n v = 0 \text{ for some } n\}.$$

We note that we have a sequence

$$E_\lambda = \ker(A - \lambda I) \subset \ker((A - \lambda I)^2) \subset \ker((A - \lambda I)^3) \subset \dots$$

so \tilde{E}_λ is the union of all these kernels. However, if V is finite-dimensional, this sequence of kernels eventually stabilizes, so there exists some minimal integer n such that $\ker((A - \lambda I)^n) = \ker((A - \lambda I)^{n+k})$ for all $k \geq 0$. Then $\tilde{E}_\lambda = \ker((A - \lambda I)^n)$, so \tilde{E}_λ is in fact a subspace. Note also that the operator $(A - \lambda I)$ acts nilpotently on the subspace \tilde{E}_λ , in other words:

$$((A - \lambda I)|_{\tilde{E}_\lambda})^n = 0.$$

But as a map $V \rightarrow V$ on the whole space, the operator $(A - \lambda I)$ typically is not nilpotent.

Example 4.17. Consider the matrix J on Jordan form below

$$A = \begin{pmatrix} \boxed{\begin{matrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{matrix}} & \begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \boxed{\begin{matrix} 8 & 1 \\ 0 & 8 \end{matrix}} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \boxed{\begin{matrix} 5 & 1 \\ 0 & 5 \end{matrix}} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \end{matrix} & \boxed{3} \end{pmatrix}$$

Here, the eigenvalues are 5, 3 and 8. In the matrix $A - 5I$, the 5's on the diagonal disappear, leaving two nilpotent Jordan blocks. We see that

$$\ker(A - 5I) = \text{span}(e_1, e_6), \ker((A - 5I)^2) = \text{span}(e_1, e_2, e_6, e_7), \ker((A - 5I)^3) = \text{span}(e_1, e_2, e_3, e_6, e_7).$$

But then the sequence clearly stabilizes and the $\ker((A - 5I)^n)$ doesn't change for $n \geq 3$. So the generalized eigenspace $\tilde{E}_5 = \ker((A - 5I)^3) = \text{span}(e_1, e_2, e_3, e_6, e_7)$ is 5-dimensional. Similarly, $\tilde{E}_8 = \text{span}(e_4, e_5)$ and $\tilde{E}_3 = \text{span}(e_8)$. Note that the direct sum of the generalized eigenspaces is the whole space

$$\mathbb{C}^8 = \tilde{E}_5 \oplus \tilde{E}_8 \oplus \tilde{E}_3,$$

and that the dimension of each \tilde{E}_λ is equal to the *algebraic multiplicity* of λ in p_A .

△

Jordan chains

The idea for finding the Jordan form of an arbitrary matrix is to treat each generalized eigenspace separately.

Definition 4.18. Let $A : V \rightarrow V$ be a linear operator (or matrix). A **Jordan chain** (for eigenvalue λ) in V is a sequence of nonzero vectors (v_1, \dots, v_n) such that $(A - \lambda I)v_i = v_{i-1}$ and $(A - \lambda I)v_1 = 0$. In other words, a Jordan chain is just a string in the sense of the previous section, but for the operator $(A - \lambda I)$.

A **Jordan basis** for V is a basis which is a union of Jordan chains (possibly with different λ).

So a Jordan basis looks like a string basis but where the chains may correspond to different eigenvalues. Note that the last vector in each string is an *eigenvector* since some $(A - \lambda I)$ maps it to zero. With respect to a Jordan basis, the matrix of the operator is clearly in Jordan form.

We can now generalize our algorithm for finding a string basis to an algorithm for finding a Jordan basis and the corresponding Jordan form, in other words, to **Jordanize** the operator. Given a square matrix A , the algorithm produces a matrix J in Jordan form, and an invertible matrix S (with columns forming a Jordan basis) such that $SJS^{-1} = A$.

Jordanizing a matrix

Algorithm 4.19. To Jordanize a square matrix A :

1. Find the characteristic polynomial, and solve $p_A(\lambda) = 0$ to find the eigenvalues.
2. For each eigenvalue λ do:
 - (a) Let \mathcal{B} be the set of previously chosen vectors (start with an empty set).
 - (b) Find the generalized eigenspace \tilde{E}_λ by computing $\ker((A - \lambda I)^k)$ for $k = 1, 2, \dots$ until the sequence stabilizes: $\ker((A - \lambda I)^n) = \ker((A - \lambda I)^{n+1})$, this happens when the dimension of $\dim \ker((A - \lambda I)^n)$ has reached the algebraic multiplicity of λ . Then $\tilde{E}_\lambda = \ker((A - \lambda I)^n)$.
 - (c) Let $N := (A - \lambda I)|_{\tilde{E}_\lambda} : \tilde{E}_\lambda \rightarrow \tilde{E}_\lambda$. This operator is nilpotent.
 - (d) Apply Algorithm 4.14 to find a string basis in \tilde{E}_λ for the operator N . This basis is a union of Jordan chains with eigenvalue λ , adjoin all these strings to \mathcal{B} .
3. Now \mathcal{B} should contain a Jordan basis. Let S be the matrix whose columns are the elements of the Jordan chains, order each chain from right to left (starting with the eigenvector). Take J to be the corresponding Jordan matrix with block sizes and eigenvalues corresponding to the ordering of the chains in S .
4. Verify that $S^{-1}AS = J$.

When we have found \tilde{E}_i , one possibility is to pick a basis and construct a smaller matrix for the operator $A|_{\tilde{E}_i} : \tilde{E}_i \rightarrow \tilde{E}_i$. This may simplify finding the string basis, but the basis should then be converted back to be written in the standard basis for the whole space. Alternatively, when following Algorithm 4.14, work with vectors sitting inside \tilde{E}_i , but expressed in the basis for the whole space.

Before looking at an example, let's consider why this algorithm actually works.

There are two important pieces missing: what if the Jordan chains we pick inside one generalized eigenspace are linearly dependent with chains from other generalized eigenspaces? And how can we be sure that all the Jordan chains we find in the algorithm actually span the whole space? These concerns are put to rest by the following lemma:

Lemma 4.20. *Let $A : V \rightarrow V$ be a linear operator (or a matrix) on a finite dimensional complex vector space. Then V is the direct sum of the generalized eigenspaces for the operator:*

$$V = \tilde{E}_{\lambda_1} \oplus \tilde{E}_{\lambda_2} \oplus \dots \oplus \tilde{E}_{\lambda_k}$$

where $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$.

Proof. First we claim that vectors from different generalized eigenspaces are linearly independent. To show this one can combine the proof-techniques used to prove Proposition 3.8 (vectors from different eigenspaces are linearly independent), and Lemma 4.8 (strings are linearly independent). We omit the details here.

It remains to prove that the eigenspaces span V . Let $\dim V = n$, factor the characteristic polynomial of A over \mathbb{C} :

$$p_A(t) = (-1)^n (t - \lambda_1)^{n_1} \dots (t - \lambda_k)^{n_k}.$$

Now perform a partial fraction decomposition of the rational function $\frac{1}{p_A(t)}$, and multiply it by $p_A(t)$, we get

$$1 = p_A(t) \cdot \frac{1}{p_A(t)} = p_A(t) \left(\frac{q_1(t)}{(t - \lambda_1)^{n_1}} + \dots + \frac{q_k(t)}{(t - \lambda_k)^{n_k}} \right) = h_1(t)q_1(t) + \dots + h_k(t)q_k(t),$$

where we introduced $h_i(t) = \frac{p_A(t)}{(t - \lambda_i)^{n_i}}$ in the last step. Note that the h_i are *polynomials* as the denominators cancel out. Now evaluate the polynomial identity above at the matrix A , define the operator $P_i := h_i(A)q_i(A)$, we get:

$$I = h_1(A)q_1(A) + \dots + h_k(A)q_k(A) = P_1 + \dots + P_k.$$

We now claim that P_i maps vectors into the generalized eigenspace \tilde{E}_{λ_i} , in other words, $\text{Im}(P_i) \subset \tilde{E}_{\lambda_i}$. Indeed for any $v \in V$, we have

$$(A - \lambda_i I)^{n_i} \cdot P_i v = (A - \lambda_i I)^{n_i} q_i(A) h_i(A) v = p_A(A) q_i(A) v = 0 \cdot q_i(A) v = 0,$$

which shows that vectors $P_i v$ in the image of P_i are killed by $(A - \lambda_i I)^{n_i}$, so $P_i v \in \tilde{E}_{\lambda_i}$ as claimed.

But then for $v \in V$ we have

$$v = Iv = (P_1 + \cdots + P_k)v = P_1 v + \cdots + P_k v,$$

so we can express any vector as a sum of vectors from each generalized eigenspace, so the \tilde{E}_{λ_i} span V . \square

Note that the proof of Lemma 4.20 actually provide an explicit way to construct projection maps onto each eigenspace: The map $P_i : V \rightarrow V$ where projects each vector v onto \tilde{E}_{λ_i} . It is not hard to prove that these projections satisfy

$$P_i^2 = P_i, \quad P_i P_j = 0, \quad P_1 + \cdots + P_k = I.$$

The proof also shows that the algebraic multiplicity of λ is equal to the dimension of the corresponding eigenspace:

$$m_\lambda = \dim \tilde{E}_\lambda.$$

This guarantees that the method of Algorithm 4.19 works and will produce a Jordan basis.

The results of this section can be summarized in a single theorem:

Theorem 4.21. Jordan Theorem. *Let $F : V \rightarrow V$ be an operator on a complex vector space. Then there exists a basis for V with respect to which the matrix of $[F]$ has Jordan form. Equivalently, any complex square matrix A has a factorization $A = SJS^{-1}$. The Jordan form is unique up to permutation of the blocks.*

Proof. Lemma 4.20 shows that Algorithm 4.19 will produce a Jordan-basis for the space V . For the uniqueness-claim, we note that the blocks of the Jordan form are uniquely determined by the numbers $\dim \ker((A - \lambda I)^k)$, and these are invariant under a change of basis. \square

Example 4.22. Let us Jordanize the matrix A below. In other words, we shall find an invertible matrix S and a matrix J in Jordan form such that $S^{-1}AS = J$.

$$A = \begin{pmatrix} 3 & -2 & -3 & 3 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 2 & -1 \\ 1 & -1 & 1 & 2 & -2 \\ 0 & 0 & -1 & 1 & 2 \end{pmatrix}$$

We find that the characteristic polynomial is $p_A(t) = \det(A - tI) = -(t - 1)^3(t - 3)^2$, so the eigenvalues are 1 and 3.

We start with the eigenvalue 1. We compute powers of $A - I$ and find that

$$A - I = \begin{pmatrix} 2 & -2 & -3 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 2 & -1 \\ 1 & -1 & 1 & 1 & -2 \\ 0 & 0 & -1 & 1 & 1 \end{pmatrix} \quad (A - I)^2 = \begin{pmatrix} 4 & -4 & -4 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 4 & -4 & 0 & 4 & -4 \\ 4 & -4 & 0 & 4 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

We solve $(A - I)v = 0$ and find that

$$\ker(A - I) = \text{span}((1, 0, 1, 0, 1), (1, 1, 0, 0, 0)).$$

Next we solve $(A - I)^2 v = 0$ and we find that $\ker((A - I)^2)$ is spanned by one more vector, so we extend our previous kernel and write

$$\ker((A - I)^2) = \text{span}((1, 0, 1, 0, 1), (1, 1, 0, 0, 0), (1, 0, 0, -1, 0)).$$

Now $\ker((A - I)^3)$, cannot be bigger (looking at the algebraic multiplicity of $\lambda = 1$).

We conclude that \tilde{E}_1 is three-dimensional, and the dimensions of the kernels tell us that we are looking for one Jordan chain of length 2, and one of length 1 (see Algorithm 4.14 for details). To find the longest chain, we start with a vector in $\ker((A - I)^2) \setminus \ker(A - I)$, let's take $v_2 := (1, 0, 0, -1, 0)$. Then take $v_1 := (A - I)v_2 = (1, 0, 1, 0, 1)$, so now we have a chain $v_2 \mapsto v_1 \mapsto 0$. For the second chain of length 1, we just pick a vector in $\ker(A - I)$ independent from v_1 and v_2 , let's take $v_3 = (1, 1, 0, 0, 0)$. Now we have our chains for the first generalized eigenspace $\tilde{E}_1 = \text{span}(v_1, v_2, v_3)$. With respect to this basis, the restriction of A onto \tilde{E}_1 will have Jordan form $J_2(1) \oplus J_1(1)$.

We move on to the second eigenvalue $\lambda = 3$ and find that

$$A - 3I = \begin{pmatrix} 0 & -2 & -3 & 3 & 1 \\ 0 & -2 & 0 & 0 & 0 \\ 1 & -1 & -2 & 2 & -1 \\ 1 & -1 & 1 & -1 & -2 \\ 0 & 0 & -1 & 1 & -1 \end{pmatrix} \quad (A - 3I)^2 = \begin{pmatrix} 0 & 4 & 8 & -8 & -4 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & -4 & 0 \\ 0 & 0 & -4 & 4 & 4 \\ 0 & 0 & 4 & -4 & 0 \end{pmatrix}$$

And there is no reason to look further, we are only missing two vectors. We find a basis for $\ker(A - 3I)$ and extend it to a basis for $\ker((A - 3I)^2)$:

$$\ker(A - 3I) = \text{span}(0, 0, 1, 1, 0) \quad \ker((A - 3I)^2) = \text{span}((0, 0, 1, 1, 0), (1, 0, 0, 0, 0)).$$

So we are looking for a single chain of length 2, we pick its first vector in $\ker((A - 3I)^2) \setminus \ker(A - 3I)$, let's take $v_5 := (1, 0, 0, 0, 0)$, and we compute $v_4 := (A - 3I)v_5 = (0, 0, 1, 1, 0)$.

Now we have our Jordan chains. We collect the data by creating $S = (v_1, v_2, v_3, v_4, v_5)$ with the chain vectors as columns, and we take J to be the corresponding Jordan matrix:

$$S = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \quad J = \begin{pmatrix} 1 & 1 & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 3 & 1 \\ & & & & 3 \end{pmatrix}$$

Then $S^{-1}AS = J$.

△

What if we consider a vector space over a field \mathbb{F} other than \mathbb{C} ? Well, we need the characteristic polynomial to factor completely over \mathbb{F} for the algorithm to work.

The general result is that if $A : V \rightarrow V$ is an operator on an \mathbb{F} -vector space, and if $p_A(t)$ factors completely over $\mathbb{F}[t]$ into linear factors:

$$p_A(t) = \det(A - tI) = \pm(t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_k)$$

for some $\lambda_i \in \mathbb{F}$, then we can Jordanize A : there exists a basis in V for which the matrix of A has Jordan form.

The previous computation was actually an example of this, we started with a *real* matrix in $\text{Mat}_5(\mathbb{R})$, and we managed to factor $p_A(t)$ completely as a product of *real* linear factors $(t - \lambda)$ where $\lambda \in \mathbb{R}$, therefore the algorithm worked and we obtained a factorization by real matrices $A = SJS^{-1}$.

4.4 Matrix functions and applications

Motivation

So what is the point of knowing how to Jordanize a matrix? Well, first of all it has many theoretical applications - if we want to prove a statement for any linear map $F : V \rightarrow V$ we may always pick a Jordan basis for V , so it suffices to prove the statement for matrices in Jordan form.

But aside from this there are several applications of a more practical nature. In a first course in linear algebra, we could approach a large class of problems via diagonalization:

- Dynamical systems in discrete time, such as predator prey models
- Explicit forms for recursively defined sequences, such as the Fibonacci numbers
- Linear systems of differential equations

Solutions to these problems were found by diagonalizing some constant coefficient-matrix. With the method of this chapter, these problems become tractable even when the coefficient matrix is non-diagonalizable.

Dynamical systems in discrete time

A dynamical system in discrete time is a system $X_{n+1} = AX_n$ where $X_n = (x_n^{(1)}, \dots, x_n^{(m)})^T$ is a set of m sequences of numbers, and A is a constant $m \times m$ -matrix. The system tells you how to go from X_n to X_{n+1} (we can think of this as a number of variables changing in the next "time step").

Now clearly $X_n = A^n X_0$, so to find the explicit form of the solutions we need to find a general formula for A^n . This could be done by the Jordan form. Write $A = SJS^{-1}$, where we can decompose $J = D + N$ into a diagonal matrix D and a nilpotent matrix N (with some ones on the superdiagonal). Then

$$A^n = (SJS^{-1})^n = SJ^nS^{-1} = S(D + N)^nS^{-1} = S(D + N)^nS^{-1}$$

So it suffices to find $(D + N)^n$. But the matrices D and N commute, so the binomial theorem applies:

$$(D + N)^n = \sum_{k=0}^n \binom{n}{k} N^k D^{n-k}.$$

But since N is nilpotent, only the first few terms of this sum are nonzero, and powers of a diagonal matrix is easy to compute.

Hopefully an example will clarify the general method:

Example 4.23. Rabbits and foxes are living in forest, the foxes are hunting the rabbits. In year number n there are r_n rabbits and f_n foxes in the forest. From the start there are 40 rabbits and 10 foxes, and we assume that the populations evolve according to the following model:

$$\begin{cases} r_{n+1} = 3r_n - f_n \\ f_{n+1} = r_n + f_n \end{cases}$$

Intuitively this should sort of make sense: If there are no foxes, the rabbit population increase 3-fold each year, but the more foxes there are the smaller the increase. Similarly, without rabbits the foxes can just sustain themselves, but more rabbits means an increase in the fox-population.

Let us find explicit formulas for r_n and f_n . First let $X_n = \begin{pmatrix} r_n \\ f_n \end{pmatrix}$ and $A = \begin{pmatrix} 3 & -1 \\ 1 & 1 \end{pmatrix}$. Then $X_0 = \begin{pmatrix} 40 \\ 10 \end{pmatrix}$ and $X_{n+1} = AX_n$ so $X_n = A^n X_0$.

To compute A^n we first Jordanize A and find that $A = SJS^{-1}$ for $J = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ and $S = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. We write $J = 2I + N$ where $N^2 = 0$. Then

$$J^n = (2I + N)^n = \sum_{k=0}^n \binom{n}{k} (2I)^{n-k} N^k = (2I)^n + n(2I)^{n-1}N = \begin{pmatrix} 2^n & 0 \\ 0 & 2^n \end{pmatrix} + \begin{pmatrix} 0 & n2^{n-1} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{pmatrix}.$$

So the solution is

$$X_n = A^n X_0 = (SJS^{-1})^n X_0 = SJ^nS^{-1} X_0 = \dots = 5 \cdot 2^n \begin{pmatrix} 8 + 3n \\ 2 + 3n \end{pmatrix},$$

in other words, in year n there are $2^n(40 + 15n)$ rabbits and $2^n(10 + 15n)$ foxes. Note that the solution gives qualitative information about what happens to the populations in the long run; in this example we get $\frac{r_n}{f_n} \rightarrow 1$ as $n \rightarrow \infty$, so in the long run there will be roughly as many foxes as rabbits.

△

Obviously a model like the one in the example will not match reality exactly, for example, if there are too many foxes, r_k can turn negative which doesn't make sense. Also, both populations will tend towards infinity here which is not possible in reality. However, such models can provide a good first approximation of a real dynamical system, and they may be useful for reasonably small populations.

Recursively defined sequences

Consider a simple recursively defined sequence $x_{n+1} = 5x_n$ with $x_0 = 3$. Since the next term is obtained by multiplication by 5, the explicit form of this sequence is clearly $x_n = 3 \cdot 5^n$.

The exact same formula actually works when we have several variables.

Example 4.24. Let us find an explicit formula for the sequence b_n defined by

$$b_n = 4b_{n-1} - 4b_{n-2}, \text{ where } b_0 = 0, b_1 = 1.$$

We introduce $X_n := \begin{pmatrix} b_{n+1} \\ b_n \end{pmatrix}$, $X_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and $A = \begin{pmatrix} 4 & -4 \\ 1 & 0 \end{pmatrix}$. Then $X_{n+1} = AX_n$, so $X_n = A^n X_0$.

We Jordanize A and find that $A = SJS^{-1}$ with $S = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$ and $J = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$.

But then we can calculate have

$$X_n = A^n X_0 = (SJS^{-1})^n X_0 = SJ^n S^{-1} X_0 = S(2I + N)^n S^{-1} X_0.$$

Here we can expand $(2I + N)^n = (2I)^n + n(2I)^{n-1}N = 2^n I + n2^{n-1}N$ by the binomial theorem, as higher powers of N are zero. The expression above simplifies to

$$\begin{aligned} X_n &= S(2^n I + n2^{n-1}N)S^{-1}X_0 = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} n2^{n-1} \\ 2^n \end{pmatrix} = \begin{pmatrix} (n+1)2^n \\ n2^{n-1} \end{pmatrix}, \end{aligned}$$

And since b_n is the bottom coordinate of X_n we obtain the explicit formula

$$b_n = n2^{n-1}.$$

△

Note that the same technique works if we just want to find a general solution to a recursively defined sequence without given starting values. In the above example, such a formula would look like $b_n = 2^n(C + Dn)$ for arbitrary scalars C, D .

Analytic matrix functions

Before looking at systems of differential equations we need to dig a bit deeper into the topic of matrix functions.

For a square matrix A , can we evaluate $f(A)$ when f is not a polynomial? For many functions the answer is yes, and this will turn out to be a useful thing to do.

Definition 4.25. Let $f : \mathbb{C} \rightarrow \mathbb{C}$ be a function whose Maclaurin-series converges for all x :

$$f(x) = f(0) + f'(0)x + \frac{f''(0)x^2}{2!} + \cdots = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)x^k}{k!}.$$

For a square matrix A we then define

$$f(A) := f(0)I + f'(0)A + \frac{f''(0)A^2}{2!} + \cdots = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)A^k}{k!}.$$

Note that $f(A)$ is an infinite sum of matrices, so here we should ask ourselves what it means for an infinite sum of matrices to converge. We will return for this question later when we talk about inner products and norms, but for now we shall think of it as element-wise convergence: If A_n is a sequence of matrices, we say that $A_n \rightarrow B$ as $n \rightarrow \infty$ if $(A_n)_{ij} \rightarrow B_{ij}$ for each position (i, j) . The condition that the Maclaurin-series converges for all x guarantees that we can evaluate $f(A)$ at any square matrix A .

However, we can use the same definition for $f(A)$ for a larger class of functions, for example $\log(A)$ - but then this expression will only make sense for a certain class of matrices²⁵.

Example 4.26. For the matrices $A = \begin{pmatrix} 4 & 0 \\ 0 & -5 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix}$, let us compute e^A , e^B , $\cos(A)$, $\cos(B)$. Recall that

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \dots \text{ and } \cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \dots$$

Therefore

$$e^A = 1 + A + \frac{A^2}{2} + \frac{A^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{A^k}{k!} = \begin{pmatrix} \sum_{k=0}^{\infty} \frac{4^k}{k!} & 0 \\ 0 & \sum_{k=0}^{\infty} \frac{(-5)^k}{k!} \end{pmatrix} = \begin{pmatrix} e^4 & 0 \\ 0 & e^{-5} \end{pmatrix}.$$

Similarly, we get $\cos(A) = \begin{pmatrix} \cos(4) & 0 \\ 0 & \cos(-5) \end{pmatrix}$.

Since B is nilpotent with $B^3 = 0$, the computation becomes easy as all but the first terms of the series disappear:

$$e^B = I + B + \frac{B^2}{2} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \cos(B) = I - \frac{B^2}{2} = \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

△

Note that equalities of functions in one variable are preserved when evaluating them at a matrix, for example $\sin(2x) = \frac{e^{ix} + e^{-ix}}{2}$ are two ways of writing the same function, and the Maclaurin-series are preserved when taking sums, so the corresponding equality will hold if we replace x by any square matrix A .

Now we can always Jordanize a matrix as $A = SJS^{-1}$, and by factoring out S and S^{-1} from each term of the Maclaurin-expansion we get $f(A) = f(SJS^{-1}) = Sf(J)S^{-1}$, so we only need to be able to evaluate f at a Jordan-matrix. Since the different Jordan-blocks do not interact when taking powers and sums of the matrix, it is in fact enough to consider how to evaluate f at a single Jordan block.

Proposition 4.27. *Let $f : \mathbb{C} \rightarrow \mathbb{C}$ be a function whose Maclaurin-series convergent at every x as in Definition 4.25. For*

$$J = J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \lambda & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix} \text{ we have } f(J) = \begin{pmatrix} \frac{f(\lambda)}{0!} & \frac{f'(\lambda)}{1!} & \frac{f''(\lambda)}{2!} & \dots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ 0 & \frac{f(\lambda)}{0!} & \frac{f'(\lambda)}{1!} & \vdots & \frac{f^{(n-2)}(\lambda)}{(n-2)!} \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{f(\lambda)}{0!} & \frac{f'(\lambda)}{1!} \\ 0 & \dots & \dots & 0 & \frac{f(\lambda)}{0!} \end{pmatrix},$$

in other words, in $f(J)$, on the k 'th superdiagonal, we have the element $\frac{f^{(k)}(\lambda)}{k!}$.

Proof. Write $J = \lambda I + N$ and in each term of $f(J)$, expand $J^k = (\lambda I + N)^k$ using the binomial theorem. Then collect terms with the same term factor N^j and use the fact that $N^n = 0$, we omit the details. □

Indeed, the proposition above is another common way to define $f(J)$.

Lemma 4.28. *If A and B commute, we have*

$$e^{A+B} = e^A e^B.$$

²⁵Technically, the spectral radius $\max\{|\lambda| : \lambda \in \sigma(A)\}$ of A should be smaller than the radius of convergence for f , more on this later.

Proof. By definition we have

$$e^A e^B = \left(\sum_{i=0}^{\infty} \frac{A^i}{i!} \right) \left(\sum_{j=0}^{\infty} \frac{B^j}{j!} \right) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{A^i B^j}{i! j!}$$

Now reorder the terms so that terms with same total degree are written together. With $k := i + j$, the expression above can be further simplified to

$$= \sum_{k=0}^{\infty} \sum_{i=0}^k \frac{A^i B^{k-i}}{i! (k-i)!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{i=0}^k \binom{k}{i} A^i B^{k-i} = \sum_{k=0}^{\infty} \frac{1}{k!} (A + B)^k = e^{A+B}.$$

In the second to last step we used the fact that the binomial theorem holds for commuting matrices. \square

The numbers in our vectors or matrices may depend on some unknown parameter t . Then we can differentiate such matrices and vectors element-wise, for example:

$$\text{For } A(t) = \begin{pmatrix} t & 3 \\ t^2 & \sin(t) \end{pmatrix} \text{ we write } \frac{d}{dt} A(t) = A'(t) = \begin{pmatrix} 1 & 0 \\ 2t & \cos(t) \end{pmatrix}.$$

Lemma 4.29. For any square matrix A we have $\frac{d}{dt} e^{tA} = A e^{tA}$.

Proof.

$$\frac{d}{dt} e^{At} = \frac{d}{dt} \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = \sum_{k=0}^{\infty} \frac{d}{dt} t^k \frac{A^k}{k!} = \sum_{k=0}^{\infty} k t^{k-1} \frac{A^k}{k!}$$

The step where we differentiated termwise may require some motivation, but since the Maclaurin-series converges anywhere, this step is fine. Now the first term in the sum is zero, so we make a change of variables and introduce $j := k - 1$, the sum above becomes:

$$= \sum_{j=0}^{\infty} (j+1) t^j \frac{A^{j+1}}{(j+1)!} = \sum_{j=0}^{\infty} t^j A \frac{A^j}{j!} = A \sum_{j=0}^{\infty} \frac{(tA)^j}{j!} = A e^{tA}.$$

\square

Note that we can easily calculate e^J where J is a Jordan block, since $J = J_n(\lambda) = \lambda I + N$ where N is the nilpotent matrix with ones on the superdiagonal. Since λI and N commute, by Lemma 4.28 we have

$$e^J = e^{\lambda I + N} = e^{\lambda I} e^N = (e^{\lambda} I) e^N = e^{\lambda} e^N,$$

and e^N is easy to compute, since $N^n = 0$, we have $e^N = I + N + \frac{N^2}{2} + \cdots + \frac{N^{n-1}}{(n-1)!}$.

Since any Jordan matrix is a direct sum of such blocks, we can do the same thing in general in each block: write $J = D + N$ where $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ is the diagonal of J , and N is nilpotent. Then

$$e^J = e^{D+N} = e^D e^N = \text{diag}(e^{\lambda_1}, \dots, e^{\lambda_n}) e^N.$$

Now let A be any square matrix. Jordanize A and write $A = S J S^{-1}$. Now

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = \sum_{k=0}^{\infty} \frac{(S J S^{-1})^k}{k!} = \sum_{k=0}^{\infty} \frac{S J^k S^{-1}}{k!} = S \left(\sum_{k=0}^{\infty} \frac{J^k}{k!} \right) S^{-1} = S e^J S^{-1},$$

which we can calculate with the previous method.

So with this method we can explicitly find e^A for any square matrix A .

Example 4.30. Let us compute e^A for $A = \begin{pmatrix} 5 & -1 \\ 1 & 3 \end{pmatrix}$. We first Jordanize A . We get $p_A(t) = \det(A - tI) = (t-4)^2$ so the only eigenvalue is 4. We have $A - 4I = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$, and $(A - 4I)^2 = 0$, so $\ker((A - 4I)^2) = \mathbb{C}^2$ and $\ker(A - 4I) = \text{span}(1, 1)$. So we will get a single chain of length 2, and as first vector we pick any vector not in $\ker(A - 4I)$, let's take $v_2 = (1, 0)$, and then we take $v_1 = (A - 4I)v_2 = (1, 1)$. We put v_1 and v_2 as columns in a matrix S and take J as the corresponding Jordan matrix, we then have $A = S J S^{-1}$ where

$$S = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \quad J = \begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = 4I + N \quad S^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}.$$

Now we can evaluate

$$e^A = e^{SJS^{-1}} = Se^J S^{-1} = Se^{4I+N} S^{-1} = Se^J S^{-1} = Se^{4I} e^N S^{-1} = Se^4 I(I+N) S^{-1} = Se^4 (I+N) S^{-1}$$

$$e^4 \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} S^{-1} = e^4 \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = e^4 \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2e^4 & -e^4 \\ e^4 & 0 \end{pmatrix}$$

△

We now expand on this example by solving a related system of differential equations:

Example 4.31. Let us find the general solutions to the linear system of differential equations

$$\begin{cases} x_1'(t) = 5x_1(t) - x_2(t) \\ x_2'(t) = x_1(t) + 3x_2(t), \end{cases}$$

and let us then in particular find the solution which also satisfies the boundary condition $x_1(0) = 3$, $x_2(0) = 5$.

Let $X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$ and let $A = \begin{pmatrix} 5 & -1 \\ 1 & 3 \end{pmatrix}$. Then the system can be written simply as $X'(t) = AX(t)$. Now we claim that $X(t) = e^{tA}C$ is a solution for every 2×1 matrix C . Indeed, for $X(t) = e^{tA}C$ we have

$$X'(t) = \frac{d}{dt} e^{tA}C = Ae^{tA}C = AX(t).$$

So to write down the general solution we only need to compute e^{tA} for the matrix A above. This exact matrix A was Jordanized in the previous problem, we had $A = SJS^{-1}$ where $S = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ $J = \begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix} = 4I + N$.

So

$$e^{tJ} = e^{4tI+tN} = e^{4tI} e^{tN} = e^{4t} I(I + tN) = e^{4t} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{4t} & te^{4t} \\ 0 & e^{4t} \end{pmatrix}$$

so

$$e^{tA} = Se^{tJ} S^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} e^{4t} & te^{4t} \\ 0 & e^{4t} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = e^{4t} \begin{pmatrix} t+1 & -t \\ t & 1-t \end{pmatrix}.$$

So an arbitrary solution can be written

$$X(t) = e^{tA}C = e^{4t} \begin{pmatrix} t+1 & -t \\ t & 1-t \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = e^{4t} \begin{pmatrix} c_1(t+1) - c_2t \\ c_1t + c_2(1-t) \end{pmatrix} = \begin{pmatrix} (c_1 - c_2)t + c_1 \\ (c_1 - c_2)t + c_2 \end{pmatrix}.$$

In particular, taking $C = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$, we get $X(0) = e^{0A}C = IC = C = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$, which is the solution satisfying our boundary condition, explicitly this solution is

$$\begin{cases} x_1(t) = (-2t + 3)e^{4t} \\ x_2(t) = (-2t + 5)e^{4t} \end{cases}.$$

△

The method in the above example works in general. We state the result as a proposition:

Proposition 4.32. An $n \times n$ system of linear differential equations

$$\begin{cases} x_1'(t) = a_{11}x_1(t) + a_{12}x_2(t) + \cdots + a_{1n}x_n(t) \\ x_2'(t) = a_{21}x_1(t) + a_{22}x_2(t) + \cdots + a_{2n}x_n(t) \\ \vdots \\ x_n'(t) = a_{n1}x_1(t) + a_{n2}x_2(t) + \cdots + a_{nn}x_n(t) \end{cases}$$

can be written as $X'(t) = AX(t)$ where $A = (a_{ij})$ and $X(t) = (x_1(t), \dots, x_n(t))^T$.

The general solution to this system is $X(t) = e^{tA}C$ where $C = (c_1, \dots, c_n)^T$ is an arbitrary vector. We note that $X(0) = C$.

5 Inner product spaces

5.1 Inner products

In a vector space we only have the operations of adding vectors and multiplying vectors by scalars. Although we usually visualize a 2-dimensional vector space as a plane, we haven't defined the concept of distances and angles between vectors yet. For more abstract vector spaces such as the vector space $\text{Mat}_n(\mathbb{C})$ of matrices, or $\mathcal{P}(\mathbb{R})$ of polynomials, it is not even clear how we should define lengths and angles. So in order to be able to do some geometry in the vector space setting we need to introduce these concept. It turns out that all you need is an *inner product*.

Definition 5.1. Let V be a complex vector space. A function $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$ is called an **inner product** or a **scalar product** on V if for all $u, v, w \in V$ and $\lambda, \mu \in \mathbb{C}$ we have:

1. $\langle \lambda u + \mu v, w \rangle = \lambda \langle u, w \rangle + \mu \langle v, w \rangle$
2. $\langle u, v \rangle = \overline{\langle v, u \rangle}$
3. $\langle v, v \rangle \geq 0$ with equality if and only if $v = 0$.

A vector space equipped with an inner product is called a (complex) **inner product space**. If we replace \mathbb{C} by \mathbb{R} we get the definition of a *real inner product space*, also called a *Euclidean space*.

A few remarks:

- By applying the second axiom we get $\langle v, v \rangle = \overline{\langle v, v \rangle}$ which shows that $\langle v, v \rangle$ is a real number, and the third condition makes sense²⁶.
- Sometimes we want to talk about different inner products on the same space, other common notations for the inner product of two vectors include (u, v) , $\langle u, v \rangle_1$, $(u|v)$, $u \bullet v$, etcetera.
- Axioms (1) and (2) imply that inner products are conjugate-linear in the second argument:

$$\langle u, \lambda v \rangle = \bar{\lambda} \langle u, v \rangle.$$

So inner products are linear in the first argument, and satisfies half of the axioms for being linear in the second argument, for this reason inner products are called *sesqui-linear forms*, "one and a half"-linear forms.

- In some contexts, such as in quantum mechanics, but *not in this course*, axiom (1) is replaced by linearity in the second argument:

$$\langle w, \lambda u + \mu v \rangle = \lambda \langle w, u \rangle + \mu \langle w, v \rangle,$$

this is just a convention and the theory will be analogous.

- In some branches of physics, like in Lorenzian geometry, Pseudo-Riemannian geometry and when studying Minkowski spacetime, axiom (3) is relaxed to

$$\langle u, v \rangle = 0 \quad \forall u \Rightarrow v = 0,$$

which allows $\langle v, v \rangle < 0$ for some vectors.

We have a number of *standard inner products* on some of our familiar vector spaces:

Example 5.2. The following functions are inner products on the respective vector space:

- On \mathbb{C}^n we have the inner product $\langle u, v \rangle = \langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle = x_1 \bar{y}_1 + \dots + x_n \bar{y}_n$. This is also called the *dot-product* and can also be written $u \bullet v$, it is defined the same way on \mathbb{R}^n .
- On $\text{Mat}_{m \times n}(\mathbb{C})$, we have the *Frobenius scalar product* $\langle A, B \rangle = \text{tr}(AB^*)$. Although this looks different than the dot-product, upon closer inspection it really is the same thing: element-wise products of entries of A with the conjugate of the corresponding elements of B : $\text{tr}(AB^*) = \sum_{i,j} a_{i,j} \bar{b}_{i,j}$.

²⁶In general, when for $z \in \mathbb{C}$ we write $z > 0$, we mean that z is *real and positive*.

- On $\mathcal{C}[a, b]$, the space of continuous functions from $[a, b]$ to \mathbb{C} (or to \mathbb{R}), we have $\langle f(x), g(x) \rangle = \int_a^b f(x) \overline{g(x)} dx$. The same works on subspaces, such as the polynomials.

It is straight forward to verify that the axioms of Definition 5.1 hold for these inner products.

△

We shall assume our vector spaces are equipped with these standard inner products unless otherwise stated.

Definition 5.3. Let V be an inner product space with inner product $\langle \cdot, \cdot \rangle$ and let $\|\cdot\|$ be the norm derived from the inner product: $\|v\| := \sqrt{\langle v, v \rangle}$. Then for $u, v \in V$:

- We define the **length** of v , also called the **norm** of v , to be $\|v\| = \sqrt{\langle v, v \rangle}$.
- We define the **distance** between u and v to be $\|u - v\|$.
- We define the **angle** between u and v to be $\arccos(\frac{\langle u, v \rangle}{\|u\| \cdot \|v\|})$ (when V is a real inner product space).
- We say that u and v are **orthogonal** if $\langle u, v \rangle = 0$.

Example 5.4. The space $\mathcal{P}(\mathbb{R})$ of polynomials with real coefficients becomes a real inner product space when equipped with the inner product

$$\langle p(x), q(x) \rangle := \int_0^1 p(x)q(x)dx.$$

Let us find the angle θ between the polynomials 1 and x with respect to this inner product.

We have

$$\langle 1, 1 \rangle = \int_0^1 1dx = [x]_0^1 = 1, \quad \langle 1, x \rangle = \int_0^1 xdx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}, \quad \langle x, x \rangle = \int_0^1 x^2dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3},$$

So

$$\theta = \arccos\left(\frac{\langle 1, x \rangle}{\|1\| \cdot \|x\|}\right) = \arccos\left(\frac{\langle 1, x \rangle}{\sqrt{\langle 1, 1 \rangle} \cdot \sqrt{\langle x, x \rangle}}\right) = \arccos\left(\frac{\frac{1}{2}}{1 \cdot \frac{1}{\sqrt{3}}}\right) = \arccos\left(\frac{\sqrt{3}}{2}\right) = \frac{\pi}{6}.$$

△

We summarize a number of direct consequences of the definitions:

Proposition 5.5. Let V be an inner product space with inner product $\langle \cdot, \cdot \rangle$ and let $\|v\| := \sqrt{\langle v, v \rangle}$ be the norm derived from the inner product. Then for $u, v \in V$ and $\lambda \in \mathbb{C}$ we have:

1. $\|\lambda v\| = |\lambda| \cdot \|v\|$.
2. $\|u + v\| \leq \|u\| + \|v\|$ (the triangle inequality)
3. $\|v\| \geq 0$ with equality if and only if $v = 0$.
4. $|\langle u, v \rangle| \leq \|u\| \cdot \|v\|$ with equality if and only if $u = \|u\|v$ (the Cauchy-Schwartz inequality)
5. $\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2$ (the parallelogram identity)
6. $4\langle u, v \rangle = \begin{cases} \|u + v\|^2 - \|u - v\|^2 & \text{if } V \text{ is real.} \\ \|u + v\|^2 - \|u - v\|^2 + i\|u + iv\|^2 - i\|u - iv\|^2 & \text{if } V \text{ is complex.} \end{cases}$

We leave the proofs as an exercise.

Note that the Cauchy-Schwarz inequality guarantees that the definition of angles in Definition 5.3 makes sense. The last formula is also interesting as it shows that the inner product can be calculated from the norm.

5.2 Norms

In some contexts, we want to define norms without necessarily having a corresponding inner product. It turns out that the first three properties of Proposition 5.5 are the key to make norms useful as a concept.

Definition 5.6. Let V be a complex vector space. A **norm** on V is a function $\|\cdot\| : V \rightarrow \mathbb{R}$ such that for $u, v \in V$ and $\lambda \in \mathbb{C}$ we have:

1. $\|\lambda v\| = |\lambda| \cdot \|v\|$
2. $\|u + v\| \leq \|u\| + \|v\|$
3. $\|v\| \geq 0$ with equality if and only if $v = 0$

The definition is the same if V is a real vector space, except that $\lambda \in \mathbb{R}$. A vector space equipped with a norm is called a **normed space**. We still define the length of v as $\|v\|$ and the distance between u and v as $\|u - v\|$ in any normed space.

Norms appear frequently in functional analysis and physics when investigating some important infinite-dimensional classes of vector spaces such as Banach-spaces²⁷ and Hilbert spaces²⁸.

Proposition 5.5 shows that every inner product gives rise to a norm, for example, the Frobenius-norm on $\text{Mat}_n(\mathbb{C})$ is $\|A\|_F = \sqrt{\sum |a_{ij}|^2}$. But the opposite is *not true*, there are norms that do not come from any inner product. Here are a few examples:

- On \mathbb{C}^n the **maximum norm** is defined as

$$\|(x_1, \dots, x_n)\|_{\max} = \|(x_1, \dots, x_n)\|_{\infty} = \max\{|x_1|, \dots, |x_n|\}.$$

- On \mathbb{C}^n the **Manhattan norm** or the **taxicab norm** is defined as

$$\|(x_1, \dots, x_n)\|_{\text{Mh}} = |x_1| + \dots + |x_n|.$$

- More generally, for $p \geq 1$, the **p-norm** on \mathbb{C}^n is defined as

$$\|(x_1, \dots, x_n)\|_p = (|x_1|^p + \dots + |x_n|^p)^{\frac{1}{p}}.$$

For $p \neq 2$ this does not correspond to an inner product.

- The **spectral norm** on $\text{Mat}_{m \times n}(\mathbb{C})$ is defined²⁹ as

$$\|A\|_{\sigma} = \max\{\sqrt{\lambda} \mid \lambda \in \sigma(A^*A)\}.$$

- The **operator norm** of a linear operator between inner product spaces $F : V \rightarrow W$ is defined as

$$\|F\|_{\text{op}} = \max\{\|F(v)\| \mid \|v\| = 1\}.$$

- The **supremum norm**³⁰ on $\mathcal{C}[a, b]$ is defined as

$$\|f(x)\|_{\text{sup}} = \max\{|f(x)| \mid x \in [a, b]\}.$$

These all satisfy the conditions in Definition 5.6. In fact many of these are the p -norm $\|\cdot\|_p$ in disguise: The standard norm is $\|\cdot\|_2$, the Manhattan norm is $\|\cdot\|_1$, and the maximum norm is $\|\cdot\|_{\infty}$ in the sense that $\|v\|_{\max} = \lim_{p \rightarrow \infty} \|v\|_p$.

²⁷A Banach-space is a complete normed vector space. "Complete" means that every Cauchy-sequence v_n converges to some v with respect to the norm: $\|v_n - v\| \rightarrow 0$.

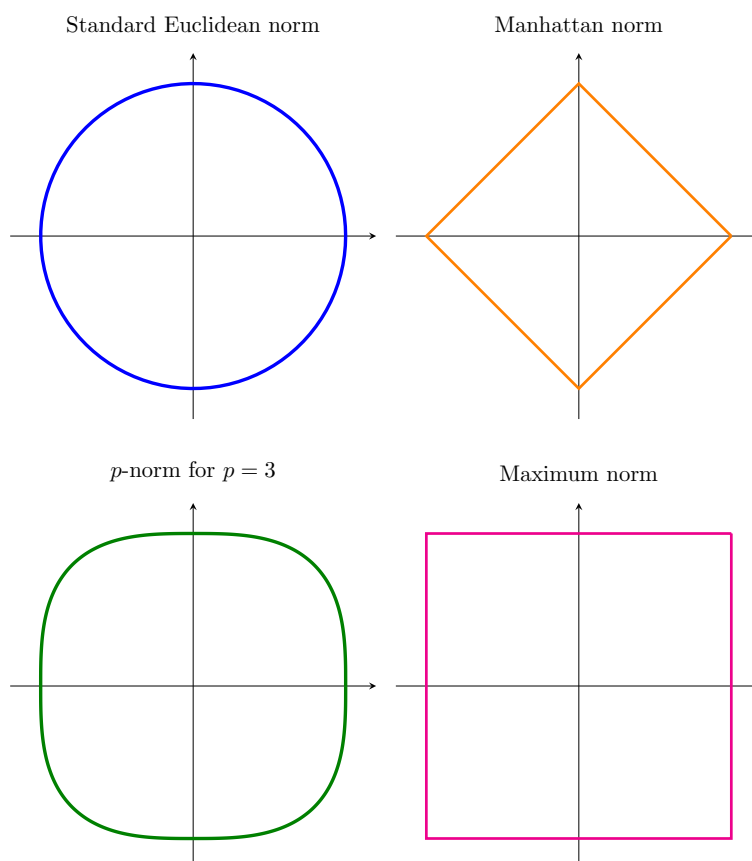
²⁸A Hilbert space is an inner product space, which is complete with respect to the norm induced from the inner product.

²⁹For square matrices A satisfying $A^*A = AA^*$, this is the same as the *spectral radius* of A , the largest absolute value of the eigenvalues of A . We will discuss this in more detail later.

³⁰The analogous definition works on any vector space of continuous functions from some compact subset $X \subset \mathbb{C}^n$ to \mathbb{C} .

Why should we consider different norms, isn't the distance between two points objective? Well, depending on the context it will make sense to measure the distance in different ways. For example, if you are on Manhattan, the streets form an orthogonal grid, and the distance you need to travel to go between the points is the sum of the vertical and the horizontal distances between the points (since you cannot drive diagonally through buildings). For this reason the Manhattan norm is appropriate in this context. In another setting we might be able to move vertically and horizontally independently with the same speed, such as a king on a chessboard - in this setting the maximum norm is the appropriate way to measure distance.

The point is, different norms give rise to different geometries on the vector space. For example, here is how the unit circle $\{v \in \mathbb{R}^2 : \|v\| = 1\}$ looks for different choices of norms $\|\cdot\|$ on \mathbb{R}^2 :



What norms can be derived from inner products? The answer is given in:

Theorem 5.7. *Let V be a vector space equipped with a norm $\|\cdot\|$. There exists an inner product $\langle \cdot, \cdot \rangle$ on V such that $\|v\|^2 = \langle v, v \rangle$ if and only if the norm satisfies the parallelogram identity:*

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2 \text{ for all } u, v \in V.$$

Proof. The idea of the proof is given by the last point of Proposition 5.5: The only option is to define $\langle u, v \rangle = \frac{1}{4}(\|u + v\|^2 - \|u - v\|^2)$ (when V is a real vector space), and then show that this function is an inner product if and only if the parallelogram law holds. It is quick to check that axioms (1) and (2) of Definition 5.1 follows from the norm axioms, but showing that $\langle \cdot, \cdot \rangle$ is linear in the first argument takes some work. See the literature for details. \square

As soon as we have a norm on a vector space we can talk about the concept of convergence in V with respect to the norm.

Definition 5.8. Let $\|\cdot\|$ be a norm on V . We say that a sequence of vectors v_1, v_2, \dots in V **converges** to some vector $v \in V$ with respect to the norm $\|\cdot\|$ if and only if

$$\|v_n - v\| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

This just means that for each $\varepsilon > 0$ there exists N such that $\|v_n - v\| < \varepsilon$ for $n > N$.

Here we might ask ourselves if it is possible that the sequence v_n converges with respect to one norm and diverges with respect to another. When V is finite-dimensional such problems do not arise:

Proposition 5.9. *If V is finite-dimensional and $v_n \rightarrow v$ with respect to some norm, then $v_n \rightarrow v$ with respect to every norm.*

This lets us choose norms freely when trying to show that a sequence of vectors converges.

5.3 Orthonormal bases and projections

In this section V is a real or complex inner product space of finite dimension unless otherwise stated, we write $\langle \cdot, \cdot \rangle$ for the inner product and $\|\cdot\|$ for the norm derived from the inner product.

Definition 5.10. A basis (e_1, \dots, e_n) for V is called an **orthonormal basis** or an **ON-basis**, if the basis vectors have length 1 and are pairwise orthogonal:

$$\langle e_i, e_j \rangle = \delta_{ij}.$$

If we drop the condition that the lengths have to be 1, we call the basis an **orthogonal basis**.

The same definition works in infinite-dimensional spaces.

Orthonormal bases make many calculations easier: if (e_1, \dots, e_n) is an ON-basis, we have

$$v = \langle v, e_1 \rangle e_1 + \dots + \langle v, e_n \rangle e_n.$$

We know that every subspace of a vector space has a complement; for $U \subset V$ we can always find another subspace U' such that $U \oplus U' = V$. In an inner product space there is a canonical choice of the complement U^\perp :

Definition 5.11. For a subspace $U \subset V$ we define its *orthogonal complement* to U as

$$U^\perp := \{v \in V \mid \langle v, u \rangle = 0 \text{ for all } u \in U\}.$$

So U^\perp is the set of vectors that are orthogonal to every vector in U . It is not too hard to show that $V = U \oplus U^\perp$ and that $(U^\perp)^\perp = U$ when V is finite-dimensional.

The canonical choice of complement also gives canonical choices for projection maps onto subspaces:

Definition 5.12. If U is a subspace of V we define the map $P_U : V \rightarrow V$ by $P_U(v) = u$ where $v = u + u'$ is the unique expression of v as a sum of $u \in U$ and $u' \in U^\perp$.

If $U = \text{span}(u)$ is one-dimensional, P_U can be calculated explicitly by the familiar projection formula^a

$$P_U(v) = P_u(v) = \frac{\langle v, u \rangle}{\langle u, u \rangle} u,$$

Otherwise, if U has higher dimension, we can pick an orthogonal basis (u_1, \dots, u_m) of U , and then we explicitly have

$$P_U(v) = P_{u_1}(v) + \dots + P_{u_m}(v) = \frac{\langle v, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 + \dots + \frac{\langle v, u_m \rangle}{\langle u_m, u_m \rangle} u_m.$$

We call $P_U(v)$ the **(orthogonal) projection** of v onto the subspace U (or onto the vector u if $U = \text{span}(u)$).

^aNote the order of the vectors in $\langle v, u \rangle$, for complex vector spaces these cannot be switched.

It is not hard to show that $P_U^2 = P_U$ and that $P_U(v)$ is the vector in U with minimal distance to v . Note that it is important that the basis u_1, \dots, u_m of the subspace we are projecting onto are indeed pairwise orthogonal, otherwise the result will be wrong.

We also remark that $P_v(u) = P_{\lambda v}(u)$ whenever λ is a nonzero complex number so the length of the vector we project onto is irrelevant, and when projecting onto a subspace it is enough to project onto the vectors of an orthogonal basis of the subspace.

5.4 Gram-Schmidt

The Gram-Schmidt process is an algorithm for converting a basis for a finite-dimensional inner product space (or a subspace) into an orthonormal basis for the same space.

Theorem 5.13. (Gram-Schmidt) *Let v_1, \dots, v_n be linearly independent vectors in an inner product space V , and let $U = \text{span}(v_1, \dots, v_n)$.*

$$\begin{aligned} e_1 &= v_1 & &= v_1 \\ e_2 &= v_2 - P_{e_1}(v_2) & &= v_2 - \frac{\langle v_2, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 \\ e_3 &= v_3 - P_{e_1}(v_3) - P_{e_2}(v_3) & &= v_3 - \frac{\langle v_3, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 - \frac{\langle v_3, e_2 \rangle}{\langle e_2, e_2 \rangle} e_2 \\ &\vdots & &\vdots \\ e_n &= v_n - P_{e_1}(v_n) - \dots - P_{e_{n-1}}(v_n) & &= v_n - \frac{\langle v_n, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 - \dots - \frac{\langle v_n, e_{n-1} \rangle}{\langle e_{n-1}, e_{n-1} \rangle} e_{n-1} \end{aligned}$$

Then (e_1, \dots, e_n) is an orthogonal basis for U . If we normalize it and define $f_i := \frac{1}{\|e_i\|} e_i$, the vectors (f_1, \dots, f_n) will be an orthonormal basis for U .

Proof. Since each e_k is defined as $v_k - u$ for some u in the span of the previous vectors e_i , we get $\text{span}(v_1, \dots, v_k) = \text{span}(e_1, \dots, e_k)$ for each $1 \leq k \leq n$. So we need only verify that the vectors e_i and e_k are orthogonal for $i \neq k$. Without loss of generality we can assume $i < k$. We do this by induction. Assume that the vectors e_i are pairwise orthogonal for $i < k$. We shall prove that e_k is orthogonal to each of these vectors e_i . By the definition of e_k in the Gram-Schmidt process we have

$$\begin{aligned} \langle e_k, e_i \rangle &= \left\langle v_k - \frac{\langle v_k, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 - \dots - \frac{\langle v_k, e_{k-1} \rangle}{\langle e_{k-1}, e_{k-1} \rangle} e_{k-1}, e_i \right\rangle = \left\langle v_k - \sum_{j=0}^{k-1} \frac{\langle v_k, e_j \rangle}{\langle e_j, e_j \rangle} e_j, e_i \right\rangle \\ &= \langle v_k, e_i \rangle - \sum_{j=0}^{k-1} \frac{\langle v_k, e_j \rangle}{\langle e_j, e_j \rangle} \langle e_j, e_i \rangle = \langle v_k, e_i \rangle - \frac{\langle v_k, e_i \rangle}{\langle e_i, e_i \rangle} \langle e_i, e_i \rangle = \langle v_k, e_i \rangle - \langle v_k, e_i \rangle = 0. \end{aligned}$$

□

It is also possible to normalize each e_i at each step in Gram-Schmidt. This produces the ON-basis f_i directly, but the calculations of the projections will typically involve square roots even if we start with integer-vectors.

We also remark that the Gram-Schmidt process still can be applied if the set of vectors is linearly dependent, then some e_i will be zero, but if we remove these we will end up with an ON-basis of $\text{span}(v_1, \dots, v_n)$.

Example 5.14. Consider the vector space $\mathcal{C}[0, 1]$ of real-valued continuous functions on the unit interval with the standard inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$.

In this example we shall use Gram-Schmidt to construct an ON-basis for \mathcal{P}_1 , the subspace of polynomials of degree ≤ 1 (with domains restricted to $[0, 1]$). We shall then use this basis to find the function in \mathcal{P}_1 closest to $f(x) = x^2$.

Since $\mathcal{P}_1 = \text{span}(1, x)$, we apply Gram-Schmidt to convert the basis $(v_1, v_2) = (1, x)$ of \mathcal{P}_1 to an orthogonal basis (e_1, e_2) of \mathcal{P}_∞ . By the algorithm we should take $e_1 = v_1 = 1$ and we note that $\|e_1\|^2 = \langle e_1, e_1 \rangle = 1$, so we get

$$e_2 = v_2 - P_{e_1}(v_2) = v_2 - \frac{\langle v_2, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 = x - \frac{\int_0^1 x \cdot 1 dx}{\int_0^1 1 \cdot 1 dx} 1 = x - \frac{1}{2}$$

and we note that $\|e_2\|^2 = \langle e_2, e_2 \rangle = \int_0^1 (x - \frac{1}{2})^2 dx = \frac{1}{12}$. So we conclude that

$$(e_1, e_2) = (1, x - \frac{1}{2})$$

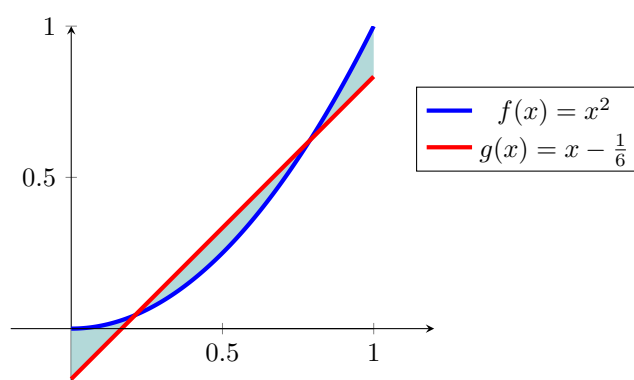
is an orthogonal basis for \mathcal{P}_1 . The second basis vector does not have length 1 though, so if we need an ON-basis we should also normalize. So with $f_1 = e_1$, and $f_2 = \frac{1}{\|e_2\|}e_2$ we get an ON-basis for \mathcal{P}_1 :

$$(f_1, f_2) = (1, \sqrt{3}(2x - 1)).$$

Now the function $g(x)$ in \mathcal{P}_1 closest to $f(x) = x^2$ is the projection of x^2 onto \mathcal{P}_1 , and this can be found by projecting onto each of our basis vectors f_1, f_2 in our ON-basis. To avoid square roots in our calculations we might as well project onto e_1 and e_2 instead:

$$g(x) = P_{\mathcal{P}_1}(x^2) = P_{f_1}(x^2) + P_{f_2}(x^2) = \frac{\langle x^2, 1 \rangle}{\langle 1, 1 \rangle} \cdot 1 + \frac{\langle x^2, (x - \frac{1}{2}) \rangle}{\langle (x - \frac{1}{2}), (x - \frac{1}{2}) \rangle} \cdot (x - \frac{1}{2}) = \frac{1}{3} + (x - \frac{1}{2}) = x - \frac{1}{6}.$$

So $g(x) = x - \frac{1}{6}$ is the function in \mathcal{P}_1 closest to x^2 . Concretely, by our definition of the inner product, this means that $\|x^2 - (x - \frac{1}{6})\|^2 = \int_0^1 (x^2 - (x - \frac{1}{6}))^2 dx$ is minimized, which we can think of as minimizing the area between the graphs of $x - \frac{1}{6}$ and x^2 on $[0, 1]^a$.



So in a sense $x - \frac{1}{6}$ is the "best approximation" of x^2 by a line if we only care about how the functions behave on the unit interval. If wanted to find the best approximation for another interval, we could perform the analogous computation for a different choice of inner product.

^aTechnically the area under the *squared difference* of the functions is being minimized.

△

The idea of the example still works if we replace polynomials by another class of functions. Projection onto the subspace produces an approximation of a given function as a linear combination of functions from this class. For example, in *Fourier-analysis*, we approximate periodic functions by *trigonometric functions*, as illustrated by the following example.

Example 5.15. Consider the space of all real 2π -periodic functions equipped with the inner product^a

$$\langle f(x), g(x) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)g(x)dx.$$

Let $\mathcal{F}_n = \text{span}(\sin(x), \sin(2x), \dots, \sin(nx))$. Let us find the function $g(x) \in \mathcal{F}_3$ which best approximates the *square wave function* $f(x)$, which is defined as $\text{sgn}(x)$ on $[-\pi, \pi)$ and is 2π -periodic.

One can check that $\{\sin(kx)\}_{k \in \mathbb{N}}$ is actually an orthonormal set of functions with respect to our inner product, so we obtain our approximation immediately as

$$g(x) = P_{\mathcal{F}_3}(f(x)) = \langle f(x), \sin(x) \rangle \sin(x) + \langle f(x), \sin(2x) \rangle \sin(2x) + \langle f(x), \sin(3x) \rangle \sin(3x)$$

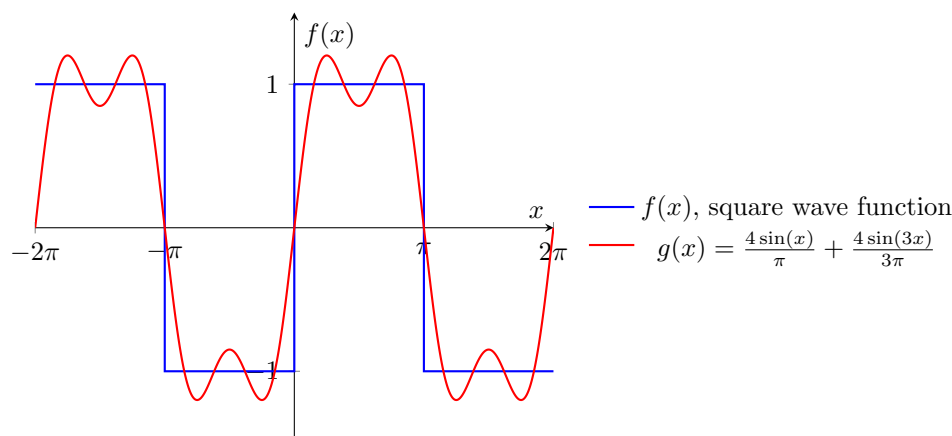
For general k we calculate $\langle f(x), \sin(kx) \rangle$. Since $f(x) \sin(kx)$ is an even function and $f(x) = 1$ for $x \in (0, \pi)$ we get

$$\langle f(x), \sin(kx) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx = \frac{2}{\pi} \int_0^{\pi} \sin(kx) dx = \frac{2}{\pi} \left[\frac{-\cos(kx)}{k} \right]_0^{\pi}$$

$$= \frac{2(1 + \cos(k\pi))}{k\pi} = \frac{2(1 - (-1)^k)}{k\pi} = \begin{cases} \frac{4}{k\pi} & \text{for odd } k \\ 0 & \text{for even } k. \end{cases}$$

This shows that

$$g(x) = \langle f(x), \sin(x) \rangle \sin(x) + \langle f(x), \sin(2x) \rangle \sin(2x) + \langle f(x), \sin(3x) \rangle \sin(3x) = \frac{4}{\pi} \sin(x) + \frac{4}{3\pi} \sin(3x).$$



Our calculation actually directly tells us the projection of the square wave function onto \mathcal{F}_{2m+1} for arbitrary m :

$$P_{\mathcal{F}_{2m+1}}(f(x)) = \sum_{k=0}^m \frac{4}{k\pi} \sin((2k+1)x).$$

Higher m gives better approximations of the function.

^aSome technical restrictions of the whole space of functions is needed in order for the integral in the inner product to exist, let us assume that all functions are continuous except for finitely many points on $[-\pi, \pi)$. We shall also consider two functions in the space to be equal if they differ for only finitely many points on that interval (otherwise the inner product would not be positive definite).

△

Approximating functions by trigonometric functions has plentiful applications in signal-processing, audio-compression, etcetera.

5.5 QR-decomposition

Definition 5.16. A **QR-factorization** or **QR-decomposition** of a matrix A is a factorization of form

$$A = QR$$

R is an upper-triangular matrix and Q is a square matrix satisfying $Q^*Q = I$.

The condition $Q^*Q = I$ is equivalent to saying that the columns of Q form an ON-basis with respect to the standard inner product on \mathbb{C}^n , such matrices are called *unitary*, we will return to these soon.

If A is an invertible matrix, A has a *unique* QR factorization if we require that the diagonal entries of R are real positive.

The QR -factorization can be obtained by applying Gram-Schmidt³¹ to the columns of the matrix as the following example demonstrates:

³¹However, the Gram-Schmidt algorithm is unstable numerically, so modern computer algorithms use different methods to find QR-factorizations.

Example 5.17. Let us find the QR -factorization of the matrix

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

We are seeking a matrix Q satisfying $QQ^* = I$, and an upper triangular matrix R such that $A = QR$.

We apply Gram-Schmidt to the columns of A with respect to the standard inner product on \mathbb{C}^3 , call the columns v_1, v_2, v_3 so that $A = (v_1 \ v_2 \ v_3)$.

We obtain $e_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, for which we have $\|e_1\|^2 = \langle e_1, e_1 \rangle = 2$, so we take $f_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.

Next we get

$$e_2 = v_2 - \frac{\langle v_2, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 = v_2 - 2e_1 = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} - 2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix},$$

for which $\|e_2\|^2 = \langle e_2, e_2 \rangle = 6$, so we take $f_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$. And finally

$$e_3 = v_3 - \frac{\langle v_3, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 - \frac{\langle v_3, e_2 \rangle}{\langle e_2, e_2 \rangle} e_2 = v_3 - \frac{3}{2} e_1 - \frac{1}{2} e_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} - \frac{3}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix},$$

and we have $\|e_3\|^2 = \langle e_3, e_3 \rangle = 3$ so we take $f_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$.

The steps in Gram-Schmidt give us a way to express our original vectors v_1, v_2, v_3 as linear combinations of our new basis vectors f_1, f_2, f_3 . We had:

$$\begin{cases} v_1 = e_1 & = \sqrt{2}f_1 \\ v_2 = 2e_1 + e_2 & = 2\sqrt{2}f_1 + \sqrt{6}f_2 \\ v_3 = \frac{3}{2}e_1 + \frac{1}{2}e_2 + e_3 & = \frac{3\sqrt{2}}{2}f_1 + \frac{\sqrt{6}}{2}f_2 + \sqrt{3}f_3 \end{cases}$$

which can be expressed in matrix form

$$\begin{aligned} A &= \begin{pmatrix} 1 & 3 & 1 \\ 1 & 1 & 2 \\ 0 & 2 & 2 \end{pmatrix} = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} | & | & | \\ f_1 & f_2 & f_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} \sqrt{2} & 2\sqrt{2} & \frac{3\sqrt{2}}{2} \\ 0 & \sqrt{6} & \frac{\sqrt{6}}{2} \\ 0 & 0 & \sqrt{3} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 2\sqrt{2} & \frac{3\sqrt{2}}{2} \\ 0 & \sqrt{6} & \frac{\sqrt{6}}{2} \\ 0 & 0 & \sqrt{3} \end{pmatrix} = QR. \end{aligned}$$

△

In the example we obtained our matrix R by keeping track of the coefficients relating our original vectors v_i to our new vectors f_i . However, note that if $A = QR$ is a QR-decomposition, we have

$$R = IR = Q^*QR = Q^*A,$$

so we can just perform Gram-Schmidt to obtain the matrix Q where the new ON-basis are columns, and then use it to compute $R = Q^*A$. If we followed Gram-Schmidt correctly, R will be upper triangular.

The method in the example works in general for finding a QR-factorization. Although the columns of an $m \times n$ -matrix are vectors of \mathbb{C}^m , they may not span \mathbb{C}^m , but we can always use Gram-Schmidt to obtain an ON-basis for their span³², and then we extend this to an ON-basis of the full space \mathbb{C}^m , then

³²In the G-S-algorithm, some resulting vectors may be zero, we drop these when constructing the basis.

we take Q as the square matrix with this ON-basis as columns, and $R = Q^*A$ will be upper triangular. We summarize:

Proposition 5.18. *Any matrix $A \in \text{Mat}_{m \times n}(\mathbb{C})$ has a QR-factorization $A = QR$ with Q unitary ($Q^*Q = I$) and R upper triangular. If $m \geq n$ we can express this in block form as*

$$A = QR = (Q_1 \mid Q_2) \begin{pmatrix} R_1 \\ 0 \end{pmatrix} = Q_1 R_1,$$

where $Q \in \text{Mat}_m(\mathbb{C})$ is a unitary and R is upper triangular. In the block-decomposition, Q_1 contains first n columns of Q , and R_1 is of shape $n \times n$.

The more compact version $A = Q_1 R_1$ is sometimes called a **thin** QR-factorization (but note that Q_1 is typically not unitary since it is not square).

Example 5.19. Let $A = \begin{pmatrix} 1 & 5 \\ 2 & 7 \\ 2 & 4 \end{pmatrix}$. We apply Gram-Schmidt to the columns and get a matrix

$Q_1 = \frac{1}{3} \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 2 & -2 \end{pmatrix}$ whose columns are orthonormal and span the same space as the columns of A .

We then take $R_1 = Q_1^* A = \begin{pmatrix} 3 & 9 \\ 0 & 3 \end{pmatrix}$. We now have the thin QR-factorization

$$A = Q_1 R_1 = \frac{1}{3} \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 3 & 9 \\ 0 & 3 \end{pmatrix}.$$

To get the full QR-factorization we extend the basis obtained via Gram-Schmidt to an ON-basis for \mathbb{C}^3 , and adjoin this last column to Q_1 . Then $R = Q^* A = \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$. We then get a full QR-factorization:

$$A = QR = (Q_1 \mid Q_2) \begin{pmatrix} R_1 \\ 0 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{pmatrix} \begin{pmatrix} 3 & 9 \\ 0 & 3 \\ 0 & 0 \end{pmatrix}.$$

△

Applications of the QR-decomposition

In the last section we saw how we could solve minimization-problems via Gram-Schmidt: The vector in v closest to a subspace U is $P_U(v)$, and this can be computed by constructing an ON-basis for U via Gram-Schmidt, and then projecting v onto each of these basis vectors. The QR-factorization gives a matrix version of this.

We illustrate by an example:

Example 5.20. Let us use the QR-factorization to find the vector in $U = \text{span}((1, 2, 2), (5, 7, 4))$ closest to $(1, 2, 3)$ (with respect to the standard norm). We put the vectors generating U as columns in a matrix A - this matrix is the same as in the last example and we already have its QR-factorization.

For the minimization we are trying to find x_1, x_2 such that $x_1(1, 2, 2) + x_2(5, 7, 4) - (1, 2, 3)$ has minimal length, this length can be expressed in matrix form as:

$$\begin{aligned} \left\| \begin{pmatrix} 1 & 5 \\ 2 & 7 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\| &= \|AX - Y\| = \|QRX - Y\| = \|Q^*QRX - Q^*Y\| = \|RX - Q^*Y\| \\ &= \left\| \begin{pmatrix} R_1 \\ 0 \end{pmatrix} X - \frac{1}{3} \begin{pmatrix} 11 \\ -2 \\ 1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} R_1 X - \frac{1}{3} \begin{pmatrix} 11 \\ -2 \end{pmatrix} \\ -(\frac{1}{3}) \end{pmatrix} \right\|, \end{aligned}$$

where in one step we used that $\|u - v\| = \|Q(u - v)\|$ which works since $QQ^* = I$. We conclude

from the calculation that the distance is minimized for $R_1 X = \frac{1}{3} \begin{pmatrix} 11 \\ -2 \end{pmatrix}$ which lets us find $X = \frac{1}{3} R_1^{-1} \begin{pmatrix} 11 \\ -2 \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 17 \\ -2 \end{pmatrix}$, which gives $P_U((1, 2, 3)) = \frac{1}{9}(7, 20, 26)$.

△

For a computer there are efficient methods for finding QR-factorizations, so for large systems the above technique is useful for solving minimization problems.

Another useful application for QR-factorization in computer algebra is the *QR-algorithm*, an algorithm that is used to find eigenvalues and eigenvectors of a matrix. For large matrices A , our usual method of computing the characteristic polynomials and finding their roots is not tenable, instead the following works: Find a QR-factorization of our matrix, flip Q and R , and repeat. More precisely, start with $A_0 = A$, and for each k , QR-factorize $A_k = Q_k R_k$ and take $A_{k+1} = R_k Q_k$. Then the sequence R_k converges to an upper triangular matrix with the eigenvalues of A on the diagonal, and Q_k converges to a matrix with the corresponding eigenvectors as columns. The proof of this lies beyond the scope of this course.

Another application of QR-factorizations is that it simplifies some matrix products, later we shall need to compute products of form A^*A for non-square matrices A . If $A = QR$ we get $A^*A = (QR)^*(QR) = R^*Q^*QR = R^*IR = R^*R$ which is faster to compute since R is upper triangular. The calculation also shows that R^*R is an LU-factorization of A^*A .

5.6 Self-adjoint, unitary, and normal operators

Note that if $\mathcal{B} = (e_1, \dots, e_n)$ is an ON-basis for an inner product space V , then we have

$$\begin{aligned} \left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}_{\mathcal{B}}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}_{\mathcal{B}} \right\rangle &= \langle x_1 e_1 + \dots + x_n e_n, y_1 e_1 + \dots + y_n e_n \rangle = \sum_{i,j} \langle x_i e_i, y_j e_j \rangle \\ &= \sum_{i,j} x_i \overline{y_j} \langle e_i, e_j \rangle = \sum_{i,j} x_i \overline{y_j} \delta_{ij} = \sum_i x_i \overline{y_i} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}. \end{aligned}$$

In other words, if we choose an ON-basis, the inner product of two vectors corresponds to the standard dot-product of their coordinate vectors in \mathbb{C}^n .

We also remark that if we write vectors $X, Y \in \mathbb{C}^n$ as column-matrices ($X, Y \in \text{Mat}_{n \times 1}(\mathbb{C})$), the dot-product can be expressed in matrix form as a matrix product

$$X \bullet Y = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^n x_i \overline{y_i} = (\overline{y_1} \quad \dots \quad \overline{y_n}) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = Y^* X,$$

note the flipped order: $X \bullet Y = Y^* X$.

Linear functionals and adjoints

Since an inner product is linear in the first argument, if we fix a vector $w \in V$, we get a linear function $\alpha : V \rightarrow \mathbb{C}$ (a linear functional) by defining $\alpha(v) = \langle v, w \rangle$, also written $\alpha = \langle \cdot, w \rangle$. The next theorem says that *every* linear functional has this form when V is finite-dimensional.

Proposition 5.21. Riesz representation theorem. *Let V be finite dimensional inner product space and let $\alpha : V \rightarrow \mathbb{C}$ be linear. Then there exists a unique vector $v_\alpha \in V$ such that*

$$\alpha(u) = \langle u, v_\alpha \rangle \quad \text{for all } u \in V.$$

Proof. Let (e_1, \dots, e_n) be an orthonormal basis for V , and define $v_\alpha := \overline{\alpha(e_1)}e_1 + \dots + \overline{\alpha(e_n)}e_n$. Then for any $u \in V$ we have

$$\langle u, v_\alpha \rangle = \left\langle u, \overline{\alpha(e_1)}e_1 + \dots + \overline{\alpha(e_n)}e_n \right\rangle = \alpha(e_1) \langle u, e_1 \rangle + \dots + \alpha(e_n) \langle u, e_n \rangle = \alpha(\langle u, e_1 \rangle e_1 + \dots + \langle u, e_n \rangle e_n) = \alpha(u),$$

which shows the existence of such a v_α . For uniqueness, we note that if v'_α also satisfies $\alpha(u) = \langle u, v'_\alpha \rangle$, then for all u we have

$$0 = \alpha(u) - \alpha(u) = \langle u, v_\alpha \rangle - \langle u, v'_\alpha \rangle = \langle u, v_\alpha - v'_\alpha \rangle.$$

But this shows that $v_\alpha - v'_\alpha = 0$ by positive-definiteness of the inner product. \square

A similar theorem also applies in a more general context³³.

Proposition 5.22. *Let $F : V \rightarrow W$ be a linear map between finite-dimensional inner product spaces (real or complex). Then there exists a unique linear map $F^* : W \rightarrow V$ called the **adjoint** of F that satisfies*

$$\langle F(v), w \rangle = \langle v, F^*(w) \rangle \quad \text{for all } v \in V \text{ and } w \in W.$$

If we pick orthonormal bases for V and W , we have $[F^] = [F]^*$, in other words, the matrix of the adjoint map is the Hermitian conjugate of the matrix for the original map.*

Proof. For each fixed $w \in W$, let $\alpha_w : V \rightarrow \mathbb{C}$ be defined by $\alpha_w(v) = \langle F(v), w \rangle$. This map is linear, so by Proposition 5.21 there exists a unique element, v_{α_w} such that

$$\langle F(v), w \rangle = \alpha_w(v) = \langle v, v_{\alpha_w} \rangle,$$

so define a map $F^* : W \rightarrow V$ by $F^*(w) = v_{\alpha_w}$, then it is not too hard to verify that this map is linear. \square

This just means that we can talk about adjoints of linear maps without specifying bases.

Definition 5.23. Let $F : V \rightarrow V$ be an operator on an inner product space V .

F is called **self-adjoint** if $F = F^*$.

F is called **unitary** if $F \circ F^* = \text{id}_V = F^* \circ F$.

A square matrix $A \in \text{Mat}_n(\mathbb{C})$ is called **unitary** if $AA^* = I = A^*A$.

F is called **normal** if $F \circ F^* = F^* \circ F$.

A square matrix in $A \in \text{Mat}_n(\mathbb{C})$ is called **normal** if $AA^* = A^*A$

Some remarks are in order.

First, we see that F is self-adjoint if and only if the matrix of F with respect to an ON-basis is Hermitian.

We see that a matrix A is unitary if and only if it is square and $A^{-1} = A^*$, this also shows that a square matrix is unitary if and only if $A^*A = I$. Moreover, columns of a unitary matrix forms an ON-basis with respect to the standard inner product. To see this we denote the columns of A by A_1, \dots, A_n , and note that the block matrix product shows that

$$A^*A = \begin{pmatrix} A_1^* \\ \vdots \\ A_n^* \end{pmatrix} (A_1 \quad \cdots \quad A_n) = (A_i^* A_j)_{ij} = (A_j \bullet A_i)_{ij}$$

so $A^*A = I$ if and only if the columns of A (and equivalently the rows of A) form an orthonormal basis in \mathbb{C}^n .

If A is a real matrix, it is unitary if and only if it is "orthogonal"³⁴: $A^T A = I = A A^T$.

If F is unitary we also note that

$$\|F(v)\|^2 = \langle F(v), F(v) \rangle = \langle v, F^*(F(v)) \rangle = \langle v, \text{id}(v) \rangle = \langle v, v \rangle = \|v\|^2,$$

which means that $\|F(v)\| = \|v\|$ for all v , and consequently $\|F(u) - F(v)\| = \|u - v\|$ so applying F doesn't change distances in the vector space, and F is called an **isometry**. Intuitively we should think of unitary operators as a sort of rotation or reflection on \mathbb{C}^n .

We also remark that both self-adjoint and unitary operators are normal.

³³In an infinite-dimensional Hilbert space, the theorem says that every *bounded* linear functional α can be represented as $\alpha = \langle \cdot, v \rangle$ for some fixed v .

³⁴"Orthonormal matrix", or "ON-matrix" in some texts.

Theorem 5.24. (Schur's Theorem) Let $F : V \rightarrow V$ be an operator on a finite dimensional complex inner product space. Then there exists an orthonormal basis for V with respect to which the matrix of F is upper triangular. Equivalently, any square complex matrix A is **unitarily equivalent** to an upper triangular matrix T : there exists a unitary matrix U such that $A = UTU^*$.

Proof. This follows directly from a modification of the proof of Theorem 3.12. We just make sure to normalize the eigenvectors we pick, and in the induction step, instead of picking an arbitrary complement to the line spanned by an eigenvector v , we pick its orthogonal complement $\text{span}(v)^\perp$. \square

5.7 Spectral theorem for normal operators

We recall the real version of the spectral theorem from a first linear algebra course, it says that *symmetric* matrices are *orthogonally diagonalizable*:

Theorem 5.25. (Real spectral theorem) Let $A \in \text{Mat}_n(\mathbb{R})$. There exists an orthonormal basis for \mathbb{R}^n consisting of eigenvectors for A if and only if A is symmetric. This means that there exists a factorization $A = SDS^T$ where D is diagonal and S is orthogonal ($SS^T = I$).

We may now prove the more general complex spectral theorem which says that the *normal* operators are precisely the ones that are orthogonally diagonalizable:

Theorem 5.26. (Complex spectral theorem) Let $F : V \rightarrow V$ be an operator on a finite dimensional complex inner product space. Then there exists an orthonormal basis for V consisting of eigenvectors for F if and only if F is normal.

Equivalently, if A is a square complex matrix there exists unitary U and diagonal D such that $A = UDU^*$ if and only if A is normal.

Proof. We prove the statement in matrix form. Suppose first that $A = UDU^*$ as in the theorem. Then

$$AA^* = (UDU^*)(UD^*U^*) = UDD^*U^* = UD^*DU^* = (UD^*U^*)(UDU^*) = A^*A,$$

so A is normal.

In the other direction, suppose A is normal. We use Schur's Theorem 5.24 to write $A = UTU^*$, and we will show that the upper triangular matrix T is in fact diagonal. Since A is normal, so is T :

$$TT^* = (U^*AU)(UA^*U^*) = UAA^*U^* = UA^*AU^* = (U^*A^*U)(UAU^*) = T^*T.$$

We claim that normal upper triangular matrices are always diagonal, we prove this by induction. The statement is trivially true for 1×1 -matrices. Let $T = (t_{ij})_{ij}$. Then explicitly the normality $TT^* = T^*T$ looks like

$$\begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & t_{nn} \end{pmatrix} \begin{pmatrix} \overline{t_{11}} & 0 & \cdots & 0 \\ \overline{t_{12}} & \overline{t_{22}} & \cdots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \overline{t_{1n}} & \overline{t_{2n}} & \cdots & \overline{t_{nn}} \end{pmatrix} = \begin{pmatrix} \overline{t_{11}} & 0 & \cdots & 0 \\ \overline{t_{12}} & \overline{t_{22}} & \cdots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \overline{t_{1n}} & \overline{t_{2n}} & \cdots & \overline{t_{nn}} \end{pmatrix} \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & t_{nn} \end{pmatrix}$$

We compare position $(1, 1)$ in both sides of the product and get

$$|t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2 = |t_{11}|^2$$

which shows that $t_{12} = t_{13} = \cdots = t_{1n} = 0$. So T has block form $\begin{pmatrix} t_{11} & 0 \\ 0 & T' \end{pmatrix}$, where T' is normal since T is, and since T' is smaller than T , the inductive hypothesis gives that T' is diagonal which shows that T is diagonal. \square

To actually find the matrix U we follow the standard diagonalization algorithm, eigenvectors corresponding to different eigenvalues will then be orthogonal, but when the geometric multiplicity is greater than one we need to make sure to pick an orthogonal basis for each eigenspace, and also normalize all eigenvectors before putting them as columns in the matrix U .

Corollary 5.27. *Self-adjoint operators, Hermitian, skew-Hermitian, and unitary matrices, as well as all complex multiples of such matrices and operators are orthogonally diagonalizable.*

Proof. It is easy to verify that all the operators listed are normal, they commute with their adjoints. \square

We remark that if A is a normal matrix whose eigenvalues are all real, we get $D = D^*$ above, which implies

$$A^* = (UDU^*)^* = UD^*U^* = UDU^* = A,$$

and A is not only normal but *Hermitian*.

If $A = UDU^*$, and if we write u_i for the columns of U such that $U = (u_1 \cdots u_n)$, then we have:

$$A = UDU^* = (u_1 \cdots u_n) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} u_1^* \\ \vdots \\ u_n^* \end{pmatrix} = (u_1 \cdots u_n) \begin{pmatrix} \lambda_1 u_1^* \\ \vdots \\ \lambda_n u_n^* \end{pmatrix} = \lambda_1 u_1 u_1^* + \cdots + \lambda_n u_n u_n^*.$$

So here we have written A as a linear combination of $n \times n$ -matrices of form $u_i u_i^*$. Such matrices clearly have rank 1 and we can see that they are in fact all orthonormal with respect to the Frobenius-inner product:

$$\begin{aligned} \langle u_i u_i^*, u_j u_j^* \rangle_F &= \text{tr}(u_i u_i^* (u_j u_j^*)^*) = \text{tr}(u_i u_i^* u_j u_j^*) = \text{tr}(\underbrace{(u_i u_i^* u_j) u_j^*}_{n \times n}) \\ &= \text{tr}(\underbrace{u_j^* (u_i u_i^* u_j)}_{1 \times 1}) = \text{tr}((u_j^* u_i)(u_i^* u_j)) = \text{tr}((u_i \bullet u_j)(u_j \bullet u_i)) = \text{tr}(\delta_{ij} \delta_{ji}) = \delta_{ij}. \end{aligned}$$

We will return to such decomposition of matrices when we talk about singular values and the Schmidt-decomposition, but then in the more general context of rectangular matrices.

Example 5.28. Let $A = \begin{pmatrix} 3 & 2 \\ -2 & 3 \end{pmatrix}$. One easily verifies that $A^*A = AA^*$, so A is normal. The standard-method shows that $\begin{pmatrix} 1 \\ i \end{pmatrix}$ is an eigenvector with eigenvalue $3 + 2i$ and that $\begin{pmatrix} 1 \\ -i \end{pmatrix}$ is an eigenvector with eigenvalue $3 - 2i$, these two eigenvectors are indeed orthogonal with respect to the standard inner product on \mathbb{C}^2 : $\begin{pmatrix} 1 \\ i \end{pmatrix} \bullet \begin{pmatrix} 1 \\ -i \end{pmatrix} = 1 \cdot \bar{1} + i \cdot \overline{-i} = 0$. We normalize them and put them as columns in a matrix U , which will then be unitary, and we take D as the diagonal matrix with the eigenvalues in the same order: with

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 3+2i & 0 \\ 0 & 3-2i \end{pmatrix}$$

we have $A = UDU^*$.

If we want to express A as a linear combination of Frobenius-orthonormal matrices as discussed above, this looks like

$$\begin{aligned} A = UDU^* &= (u_1 \quad u_2) \begin{pmatrix} 3+2i & 0 \\ 0 & 3-2i \end{pmatrix} \begin{pmatrix} u_1^* \\ u_2^* \end{pmatrix} = (3+2i)u_1 u_1^* + (3-2i)u_2 u_2^* \\ &= (3+2i)\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \frac{1}{\sqrt{2}} (1 \quad -i) + (3-2i)\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} \frac{1}{\sqrt{2}} (1 \quad i) = (3+2i)\frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} + (3-2i)\frac{1}{2} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}. \end{aligned}$$

\triangle

5.8 Positive definite operators

Definition 5.29. Let $F : V \rightarrow V$ be a self-adjoint^a operator on a complex inner product space V . F is called...

- **Positive definite** if $\langle F(v), v \rangle > 0$ for all nonzero $v \in V$.
- **Positive semi-definite** if $\langle F(v), v \rangle \geq 0$ for all nonzero $v \in V$.
- **Negative definite** if $\langle F(v), v \rangle < 0$ for all nonzero $v \in V$.
- **Negative semi-definite** if $\langle F(v), v \rangle \leq 0$ for all nonzero $v \in V$.

These concepts are defined analogously for matrices, for example:
A Hermitian matrix A is called **positive definite** if $AX \bullet X > 0$, or equivalently

$$X^*AX > 0 \text{ for all nonzero columns } X.$$

^aThis condition is not necessary for complex matrices, every positive (semi-)definite matrix will be Hermitian, the proof of this is left as an exercise.

So we note that F is positive definite if and only if its matrix $[F]$ with respect to an *ON-basis* is positive definite.

Example 5.30. Let $A = \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 6 & 0 \\ 0 & 0 \end{pmatrix}$ and $C = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$, and let $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ be an arbitrary nonzero vector.

Then $X^*AX = (\overline{x_1} \ \overline{x_2}) \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3|x_1|^2 + 5|x_2|^2 > 0$ so A is positive definite.

For B we have $X^*BX = (\overline{x_1} \ \overline{x_2}) \begin{pmatrix} 6 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 6|x_1|^2 \geq 0$, but it can be zero for $X \neq 0$, take for example $X = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. So B is positive semi-definite.

For C we have $X^*CX = (\overline{x_1} \ \overline{x_2}) \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2|x_1|^2 + \overline{x_1}x_2 + |x_2|^2$. This may be non-real number (take for example $X = \begin{pmatrix} 1 \\ i \end{pmatrix}$), so it is neither ≥ 0 nor ≤ 0 for all X , and C is neither positive or negative (semi)-definite.

△

Suppose that A is positive definite and that A is unitarily equivalent to a matrix B via $B = U^*AU$. Then for $X \neq 0$ we have

$$X^*BX = X^*U^*AUX = (UX)^*A(UX) > 0,$$

and since X is arbitrary nonzero and U is bijective, $Y = UX$ is also arbitrary nonzero. This shows that B is positive definite too.

In particular, since A is necessarily Hermitian, by the spectral theorem it is unitarily equivalent to a diagonal matrix (with the eigenvalues of A on the diagonal), and for a diagonal matrix $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ we have

$$X^*DX = \lambda_1|x_1|^2 + \dots + \lambda_n|x_n|^2$$

which is clearly positive for all nonzero X if and only if all $\lambda_i > 0$. We conclude that a Hermitian matrix A is positive definite if and only if all its eigenvalues are positive (and positive semi-definite if all its eigenvalues are ≥ 0).

We note that if we write $X, Y \in \mathbb{C}^n$ as columns-matrices, $\langle X, Y \rangle := X^*AY$ defines an inner product on \mathbb{C}^n if and only if A is positive definite. The sesquilinearity and conjugate-symmetry follows directly, and positive-definiteness of the inner product $\langle X, X \rangle > 0$ is equivalent to A being positive definite.

The following result provides a useful condition for testing whether a Hermitian matrix is positive definite:

Theorem 5.31. (Sylvester's criterion) Let $A \in \text{Mat}_n(\mathbb{C})$ be a Hermitian matrix ($A = A^*$). The principal minor of size $m \times m$ in A is the determinant of the matrix obtained from removing all but

the first m rows and columns from A .

The Hermitian matrix A is positive definite if and only if all its principal minors are positive, and A is positive semi-definite if all its principal minors are ≥ 0 .

We don't prove this here, but we illustrate with an example:

Example 5.32. The matrix $A = \begin{pmatrix} 1 & i & 1 \\ -i & 2 & 1 \\ 1 & 1 & 5 \end{pmatrix}$ is Hermitian. Its three principal minors of size 1, 2, 3 are:

$$|1| = 1, \quad \begin{vmatrix} 1 & i \\ -i & 2 \end{vmatrix} = 1, \quad \begin{vmatrix} 1 & i & 1 \\ -i & 2 & 1 \\ 1 & 1 & 5 \end{vmatrix} = 2.$$

Since all of these are positive, by Sylvester's criterion A is positive definite.

△

Square roots

In calculus, for $x \in \mathbb{R}$ we normally define \sqrt{x} as the unique *non-negative* real number y satisfying $y^2 = x$. So for $x < 0$, the square root \sqrt{x} is normally *undefined*. For $x > 0$ there are two real numbers y satisfying $y^2 = x$, but only one of them is positive.

For matrices the situation is analogous if we replace "positive" by "positive definite":

Definition 5.33. Let A be a positive semi-definite matrix. Then A is necessarily Hermitian, and by the spectral theorem we may write $A = UDU^*$ for $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ and since A is positive semi-definite, so is the matrix D , so $\lambda_i \geq 0$. We define

$$\sqrt{D} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}) \quad \text{and} \quad \sqrt{A} = U\sqrt{D}U^*.$$

So we leave \sqrt{A} *undefined* when A is not positive semi-definite, even if there may exist matrices which squares to A .

Proposition 5.34. \sqrt{A} is well defined for positive semi-definite A .

\sqrt{A} is the unique positive semi-definite matrix whose square is A .

Proof. We have $\sqrt{A}^2 = U\sqrt{D}U^*U\sqrt{D}U^* = U\sqrt{D}\sqrt{D}U^* = UDU^* = A$, and \sqrt{A} is positive semi-definite since all its eigenvalues are ≥ 0 .

For the uniqueness claim, suppose B is any positive semi-definite matrix for which $B^2 = A$. Since B is positive semi-definite it is Hermitian, and we can use the spectral theorem to write $B = \tilde{U}\tilde{D}\tilde{U}^*$ where $\tilde{D} = \text{diag}(\mu_1, \dots, \mu_n)$ has the eigenvalues of B on the diagonal. Then $A = B^2 = \tilde{U}\tilde{D}^2\tilde{U}^*$ so $A\tilde{U} = \tilde{U}\tilde{D}^2$ which shows that the columns of \tilde{U} are orthonormal eigenvectors for A with eigenvalues μ_i^2 , this shows that $\mu_i^2 = \lambda_i$. Since A is positive semi-definite and Hermitian, all its eigenvalues λ_i are real and ≥ 0 , which shows that $\mu_i = \pm\sqrt{\lambda_i}$, but since B is required to be positive semi-definite, only the plus-signs can occur. But then \sqrt{A} and B acts the same way on a basis for the vector space (the columns of \tilde{U}), so $B = \sqrt{A}$. □

Example 5.35. Let $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. By Sylvesters criterion, A is positive definite and it has a square root \sqrt{A} . Let us compute it:

We find the eigenvalues of A and an orthonormal basis of eigenvectors, this gives us

$$A = UDU^* \quad \text{where} \quad U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}.$$

So we have

$$\sqrt{A} = U\sqrt{D}U^* = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2 + \sqrt{2} & 2 - \sqrt{2} \\ 2 - \sqrt{2} & 2 + \sqrt{2} \end{pmatrix}.$$

Note that each of the four matrices $B = U \begin{pmatrix} \pm 2 & 0 \\ 0 & \pm \sqrt{2} \end{pmatrix} U^*$ satisfies $B^2 = A$, but only one of them is positive definite.

△

6 Problems

Vector spaces, subspaces, direct sum, quotients

- 1.1. Use the vector space axioms to prove that $v + v = 2 \cdot v$ holds in any complex vector space.
- 1.2. Consider $\mathbb{R}_+ =]0, \infty[$, the set of positive real numbers. We define an addition $+$ on \mathbb{R}_+ by

$$x + y := xy.$$

We define a multiplication of real numbers λ on \mathbb{R}_+ by

$$\lambda \bullet x := x^\lambda.$$

Under this addition and scalar multiplication, \mathbb{R}_+ is in fact a vector space.

- What is the zero element (the additive identity) of the vector space V ?
 - What is -5 ? (the additive inverse of the vector 5)?
 - Verify that the vector space axiom $(\lambda + \mu) \bullet v = \lambda \bullet v + \mu \bullet v$ holds in V .
 - Verify that $\lambda \bullet (\mu \bullet v) = (\lambda\mu) \bullet v$.
- 1.3. Which of the following subsets of \mathbb{R}^2 are subspaces? Which satisfy the additive property, and which satisfy the homogeneity property?
- $S_1 = \{(x, y) \in \mathbb{R}^2 \mid 2x + y = 3\}$
 - $S_2 = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$
 - $S_3 = \{(x, y) \in \mathbb{R}^2 \mid xy = 0\}$
 - $S_4 = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0 \text{ and } y \geq 0\}$
 - $S_5 = \{(x, y) \in \mathbb{R}^2 \mid x \in \mathbb{Z} \text{ and } y \in \mathbb{Z}\}$
 - $S_6 = \mathbb{R}^2$
 - $S_7 = \emptyset$
- 1.4. In \mathbb{R}^3 , let U be the plane $x + 2y + 3z = 0$ and let U' be the line spanned by $(1, 1, 1)$. Then $\mathbb{R}^3 = U \oplus U'$. Find the projection of $v = (1, 4, 1)$ on U with respect to this direct sum.
- 1.5. Show that $S = \{p(x) \in \mathcal{P}_3 \mid p(2) = 0\}$ is a subspace of \mathcal{P}_3 . Also find a basis for it.
- 1.6. Let S and S' be any two subspaces of a vector space V . Which of the following statements are true? For false statements, give a counterexample. For true statements, give a short proof.
- The intersection $S \cap S' := \{v \in V \mid v \in S \text{ and } v \in S'\}$ is a subspace.
 - The union $S \cup S' := \{v \in V \mid v \in S \text{ or } v \in S'\}$ is a subspace.
 - The sum $S + S' := \{u + v \mid u \in S, v \in S'\}$ is a subspace.
- 1.7. Let U_1, U_2, U_3 be three subspaces of V such that $U_1 + U_2 + U_3 = V$ and $U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{0\}$. Show that we do not necessarily have $V = U_1 \oplus U_2 \oplus U_3$.
- 1.8. Let \mathcal{F} be the vector space of all functions $\mathbb{R} \rightarrow \mathbb{R}$. Let \mathfrak{e} be the set of all even functions, and let \mathfrak{o} be the set of all odd functions in \mathcal{F} .
- Show that \mathfrak{e} and \mathfrak{o} are subspaces of \mathcal{F} .
 - Show that $\mathcal{F} = \mathfrak{e} \oplus \mathfrak{o}$.
 - Find the projection of e^x onto the subspace \mathfrak{e} .

d) Find the projection of $f(x) = \sin(x)x^{10} \arctan(x)$ onto \mathfrak{o} .

1.9. Which of the following maps are linear?

- a) $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ where $F(x, y) = (y + x, 2x - 1)$.
- b) $I : \mathcal{P}_n \rightarrow \mathcal{P}_{n+1}$ where $I(p(x)) = \int_0^1 p(x) dx$
- c) $G : \text{Mat}_{n \times n}(\mathbb{C}) \rightarrow \mathbb{C}$ where $G(A) = \text{tr}(A)$
- d) $H : \text{Mat}_{n \times n}(\mathbb{C}) \rightarrow \text{Mat}_{n \times n}(\mathbb{C})$ where $H(A) = A^T + 3A$
- e) $T : \mathcal{C}(\mathbb{R}) \rightarrow \mathcal{C}(\mathbb{R})$ where $T(f(x)) = f(x + 1)$
- f) $C : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ where $C(v) = v \times (1, 2, 3)$.

1.10. Let $T : \text{Mat}_n(\mathbb{C}) \rightarrow \text{Mat}_n(\mathbb{C})$ where $T(A) = A - A^T$. Show that T is linear and determine $\ker(T)$ and $\text{Im}(T)$.

1.11. Show that a linear map F is injective if and only if $\ker(F) = \{0\}$.

1.12. Let V be the vector spaces of all infinite sequences (a_1, a_2, \dots) where $a_i \in \mathbb{R}$; the sum and scalar action is defined coordinate-wise. Let $F : V \rightarrow V$ be the right-shifting operator $F(a_1, a_2, \dots) = (0, a_1, a_2, \dots)$, this is a linear map. Is F injective? Surjective? Does F have an inverse?

1.13. Let \mathcal{P}_4 be the space of polynomials with real coefficients and of degree ≤ 4 . Define a linear map $F : \mathcal{P}_4 \rightarrow \mathcal{P}_4$ by $F(p(x)) = p(x + 1)$. Find the matrix of F with respect to the standard basis of \mathcal{P}_4 . Also find the inverse of F .

1.14. In Lie theory an important object of study is \mathfrak{sl}_2 , the set of complex matrices with trace zero:

$$\mathfrak{sl}_2 = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{Mat}_{2 \times 2}(\mathbb{C}) \mid a + d = 0 \right\},$$

this is a subspace of $\text{Mat}_{2 \times 2}(\mathbb{C})$ with basis $(X, H, Y) = \left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right)$. Define two maps $F, G : \mathfrak{sl}_2 \rightarrow \mathfrak{sl}_2$ by

$$F(A) = HA - AH \text{ and } G(A) = XA - AX.$$

Find the matrices of F and of G with respect to the basis (X, H, Y) .

1.15. Let U be a subspace of a vector space V . Let $v + U$ and $w + U$ be two affine subsets. Prove first that $v + U = (v + u') + U$ whenever $u' \in U$. Use this to prove that $v + U = w + U$ whenever $v - w \in U$.

1.16. Let $V = \mathbb{R}^2$ and let U be the subspace spanned by $(1, 1)$. Let $A = (2, 3) + U$, $B = (1, 0) + U$, and $C = (4, 5) + U$ be affine subsets.

- a) Are any of A, B, C equal?
- b) Sketch the affine subsets A, B , and $A + B$ in a picture of \mathbb{R}^2 .
- c) Show that A is a basis for V/U and express B in this basis.

1.17. Let $U \subset \mathbb{R}^3$ be the subspace spanned by $(1, 2, 3)$. Then the pair $(A, B) = ((1, 1, 0) + U, (0, 1, 1) + U)$ is a basis for \mathbb{R}^3/U . Express the vector $C = (1, 1, 1) + U$ of V/U in this basis.

1.18. Let $V = \mathbb{R}^4/\ell$ where $\ell = \text{span}(3, 2, 1, 2)$. Is

$$(e_1, e_2, e_3) = ((1, 1, 0, 0) + \ell, (0, 1, 1, 0) + \ell, (0, 0, 1, 1) + \ell)$$

a basis for V ?

1.19. Let $F : V \rightarrow V$ be a linear map, and let $U \subset V$ be a subspace. Show that if $F(u) = 0$ for all $u \in U$, then the map defined by $\tilde{F} : V/U \rightarrow V/U$ with $\tilde{F}(v + U) = F(v) + U$ is well-defined and linear. How does the matrix of F and \tilde{F} look?

1.20. Write $\text{Aspan}(p_1, \dots, p_n)$ for the **affine span** of vectors (or think of them as points) p_1, \dots, p_n , defined as the smallest affine subset containing all vectors p_1, \dots, p_n . Find a geometric description of each affine span below, and express it as $v + U$ for suitable vector v and subspace U .

- a) $\text{Aspan}((1, 2), (3, 4))$ in \mathbb{R}^2 .
- b) $\text{Aspan}((1, 0, 0), (0, 1, 0), (0, 0, 1))$ in \mathbb{R}^3 .

1.21. A map $F : V \rightarrow V$ is called an **affine map** if $F(v) = G(v) + w$ where G is a linear map and w is a fixed translation-vector.

- a) Show that the composition of affine maps is affine.
- b) Show that the map that reflects points of \mathbb{R}^2 in the line $(1 + t, t)$ is an affine map.

1.22. In this problem we consider vector spaces and matrices over the field \mathbb{Z}_3 , it consists of only three elements $\{0, 1, 2\}$ which can be added and multiplied modulo 3 as indicated in these tables:

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

·	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

- a) Compute $2(1, 2, 1, 1) + (2, 2, 0, 1)$
- b) Solve the linear system $\begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$
- c) Compute $\det \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$
- d) Find $\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}^{-1}$
- e) How many elements does $\text{Mat}_{2 \times 2}(\mathbb{Z}_3)$ have?

1.23. Let $X = \{1, 2, 3, 4, 5\}$ be a finite set, and let $\mathcal{P}(X)$ be the set of all subsets of X . Define an addition on $\mathcal{P}(X)$ by the symmetric difference operator

$$S_1 + S_2 := S_1 \Delta S_2 = (S_1 \cup S_2) \setminus (S_1 \cap S_2).$$

- a) How should scalar multiplication be defined in order for $\mathcal{P}(X)$ to become a vector space over \mathbb{Z}_2 ? What is the additive identity in $\mathcal{P}(X)$?
- b) Compute $\{1, 3, 5\} + \{1, 2, 3\}$
- c) Compute $-\{1, 3, 5\}$
- d) Find a basis for $\mathcal{P}(X)$
- e) Are the vectors $\{1, 3, 4\}, \{1, 2\}, \{1, 4, 5\}, \{2, 3, 4\} \in \mathcal{P}(X)$ linearly dependent?

1.24. Let V and W be vector spaces over \mathbb{Q} , and let $F : V \rightarrow W$ be a map satisfying $F(u + v) = F(u) + F(v)$ for all $u, v \in V$. Show that F is linear.

Matrices, echelon forms, LU-factorization

- 2.1. Let A and B be Hermitian matrices of the same size. Which of the matrices below are guaranteed to be Hermitian? Give a proof or a counterexample to each.

$$A + B \qquad AB \qquad \lambda A \qquad A^T \qquad AB^* + BA^*$$

- 2.2. Any complex matrix can be written uniquely as $A + Bi$ where A and B are real matrices. Prove that $A + Bi$ is Hermitian if and only if A is symmetric and B is skew-symmetric.

- 2.3. Prove that $\text{tr}(AB) = \text{tr}(BA)$ whenever both matrix products are defined. Recall that the trace of a matrix is the sum of its diagonal entries.

2.4.

- a) Show that if an $n \times n$ -matrix A has eigenvalues $\lambda_1, \dots, \lambda_n$ (all different), then

$$\text{tr}(A) = \lambda_1 + \dots + \lambda_n.$$

- b) Suppose that A is a 3×3 -matrix with 3 different eigenvalues $\lambda_1, \lambda_2, \lambda_3$. Find a formula for $\text{tr}(A^n)$.

- c) Suppose that A is 3×3 and $\text{tr}(A) = 1$, $\text{tr}(A^2) = 6$, and $\text{tr}(A^3) = 10$. Find the eigenvalues of A .

- 2.5. The matrix P below is an example of what is called a *permutation matrix*. Compute P^n for each integer n .

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

- 2.6. A **permutation matrix** is a matrix which has a single 1 in each column and in each row.

- a) Prove that a permutation matrix must be square.
 b) Prove that the product of two permutation matrices is a permutation matrix.
 c) Prove that if P is a permutation matrix, then $P^n = I$ for some $n > 0$.
 d) We define the *order* of a permutation matrix P as the minimal positive n for which $P^n = I$. Find an example of an $n \times n$ -permutation matrix whose order is greater than its size n .

- 2.7. Let $A \in \text{Mat}_n(\mathbb{R})$ be a square matrix, and let $\mathcal{C}_A := \{B \in \text{Mat}_n(\mathbb{R}) \mid AB = BA\}$ be the **commutant** of A , the set of matrices that commute with A . Show that \mathcal{C}_A is a subspace of $\text{Mat}_n(\mathbb{R})$. Then find the commutant of $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$.

- 2.8. Let $\mathfrak{C} = \{A \in \text{Mat}_n(\mathbb{C}) \mid AB = BA \text{ for all } B \in \text{Mat}_n(\mathbb{C})\}$ be the set of matrices that commute with every other matrix. Describe the set \mathfrak{C} explicitly.

- 2.9. In algebra, an object x satisfying $x \cdot x = x$ is called an **idempotent**. For example, there are exactly two idempotents in \mathbb{R} : 1 and 0. Show that there are infinitely many idempotents in $\text{Mat}_2(\mathbb{R})$.

- 2.10. Let A be a real diagonalizable 5×5 -matrix. What values are possible for the dimension of the commutant $\dim \mathcal{C}_A$?

- 2.11. Recall that matrix N is called nilpotent if $N^d = 0$ for some d . Let N and M be nilpotent matrices that commute. Show that NM and $N + M$ are both nilpotent. Is the statement still true if the matrices do not commute?

- 2.12. Let $F : V \rightarrow V$ be nonzero operator satisfying $\text{Im}(F) \subset \ker(F)$. Show that F is nilpotent, and find its nilpotency degree.

2.13. Show that if $N \in \text{Mat}_n(\mathbb{C})$ is nilpotent, then its nilpotency degree is $\leq n$.

2.14. Show that $I + N$ is invertible whenever N is nilpotent.

2.15. Let $M = \left(\begin{array}{c|c} 2I & 0 \\ \hline A & 3I \end{array} \right)$ be the blockmatrix of size $2n \times 2n$, where each of the blocks has size $n \times n$ and A is some given matrix. Show that $M^{-1} = \frac{1}{6} \left(\begin{array}{c|c} 3I & 0 \\ \hline -A & 2I \end{array} \right)$.

2.16. Find a general formula for the inverse of the block matrix $M = \left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array} \right)$, where each block has size $n \times n$, and where A and C are invertible.

2.17. For the matrix below, find a row echelon form and find $\text{rank}(A)$. Also find the reduced row echelon form of A , and use your result to find the nullspace $\ker(A)$.

$$A = \begin{pmatrix} 1 & -1 & 1 & 2 & 1 \\ 2 & -2 & 4 & 3 & 1 \\ 3 & -3 & 5 & 5 & 2 \end{pmatrix}$$

2.18. For the matrix C below, find the reduced row echelon form, and use it to solve the linear system $CX = 0$.

$$C = \begin{pmatrix} 1 & 2i & 1+i & i \\ 2i & -4 & 1 & 3i \end{pmatrix}$$

2.19. A is a 3×5 matrix with $\ker(A) = \{(-2s + 3t - r, s, t, -r, 2r) \mid s, t, r \in \mathbb{R}\}$. Find the reduced row echelon form of A .

2.20. Later, when determining Jordan canonical form of a matrix, we will have to find bases for various subspaces related to the matrix.

$$\text{Let } A = \begin{pmatrix} 0 & 1 & 0 & 2 & -1 \\ -1 & 1 & 1 & -1 & 0 \\ 1 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 1 & -2 & 0 \end{pmatrix}$$

- Find a basis for $\text{Im}(A)$
- Find a basis for $\ker(A)$
- Find a basis for $\ker(A) \cap \text{Im}(A)$

2.21. Give an example of a matrix A such that $\ker(A)$ and $\text{Im}(A)$ have a nontrivial intersection, and neither is a subset of the other

2.22. Find the inverse of each elementary matrix below:

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix} \quad E_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then describe a general formula for the inverse of an elementary matrix.

2.23. Row and column operations can be achieved by matrix multiplication. Let A be a 3×3 matrix. What matrix should we multiply A by, and from what side, to have the effect of

- Adding two times the first row of A to the third row of A
- Multiplying the middle column of A by 3

c) Switching the first two rows in A

2.24. Write $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ as a product of elementary matrices. Then do the same for A^{-1} .

2.25. Recall the Gauss-Jordan method for finding the inverse of a square matrix A : Write down the block-matrix $[A|I]$ and do row operations on this matrix until it has form $[I|B]$, then $A^{-1} = B$. Prove that the Gauss-Jordan method works.

2.26. Let

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

Find the LU-decomposition and the LDU-decomposition of A .

2.27. Let

$$A = \begin{pmatrix} 1 & -1 & 1 & 1 \\ -2 & 5 & 0 & -1 \\ 3 & 3 & 1 & 2 \end{pmatrix}.$$

Find the LU-decomposition and the LDU-decomposition of A .

2.28. The matrix A below does not admit an LU -decomposition. Find instead a decomposition $PA = LU$ where L is lower triangular, U is upper triangular, and P is a permutation matrix.

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

2.29. Let U be a row echelon form of A . Prove that if the columns of A satisfy a linear dependence relation

$$\lambda_1 A_1 + \lambda_2 A_2 + \cdots + \lambda_n A_n = 0$$

then the columns of U satisfy the same relation:

$$\lambda_1 U_1 + \lambda_2 U_2 + \cdots + \lambda_n U_n = 0.$$

Conclude that the dimension of the span of the columns of A is equal to the number of pivots in U .

2.30. We say that a matrix A is in (reduced) **column echelon form** if and only if A^T is in (reduced) row echelon form. Describe the set of matrices that are both in reduced row echelon form and reduced column echelon form.

2.31. Solving a linear system by row operations involves multiplying and adding numbers (and making a few divisions too to figure out what row operations to make, and to solve the final diagonal system, but let's ignore all these). For a computer, addition is a lot faster than multiplication, so the number of multiplications is the limiting factor when solving a linear system.

- How many multiplications are required to solve a generic system $AX = b$ where A is a 4×4 -matrix with the standard Gaussian elimination algorithm? (here generic means that no unexpected zeros appear when performing the row operations)
- Assume now that $A = LU$ is an LU-factorization of the matrix above. The system $AX = b$ can then be written $L(UX) = b$, and we can solve it by solving the two triangular systems $LY = b$ and then $UX = Y$. How many multiplications are required in total?
- Assume that in (b) the matrix A is an $n \times n$ -matrix. Determine the number of multiplications needed.

2.32. Find the Cholesky-factorization for each of the matrices below:

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 6 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 2-i \\ 2+i & 9 \end{pmatrix}.$$

2.33. Find the Cholesky-factorization of the matrix A below.

$$A = \begin{pmatrix} 9 & 3 & -3 \\ 3 & 2 & 1 \\ -3 & 1 & 10 \end{pmatrix}$$

2.34. Prove that if A admits a Cholesky-factorization, then A is Hermitian. Also prove that the reverse implication does not hold.

2.35. Find the reduced row echelon form of the matrix $A \in \text{Mat}_{3 \times 4}(\mathbb{Z}_3)$ below, and use your result to solve $AX = 0$.

$$A = \begin{pmatrix} 2 & 1 & 2 & 1 \\ 2 & 1 & 1 & 0 \\ 1 & 2 & 2 & 0 \end{pmatrix}.$$

Introductory spectral theory

3.1. Find the spectrum $\sigma(F)$, and the dimension of the corresponding eigenspaces for each of the linear maps $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ described below:

- a) Projection onto a line
- b) Reflection in a plane
- c) Rotation around an axis
- d) The identity map
- e) A nilpotent operator of nilpotency-degree 3

3.2. Find all eigenvalues and eigenvectors of the linear map on \mathbb{C}^2 given by the matrix $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Also diagonalize A , in other words, find matrices $S, D \in \text{Mat}_2(\mathbb{C})$ such that D is diagonal and $A = SDS^{-1}$.

3.3. Let \mathcal{P} be the space of all polynomials with real coefficients. Find all eigenvectors and eigenvalues of the operator $F : \mathcal{P} \rightarrow \mathcal{P}$ defined by $F(p(x)) = xp'(x)$. What is the spectrum $\sigma(F)$?

3.4. Show that if A is a real matrix and λ is an eigenvalue, then so is $\bar{\lambda}$.

3.5. We know that a linear operator on \mathbb{C}^2 with matrix $A \in \text{Mat}_2(\mathbb{R})$ has $2 + 3i$ as an eigenvalue with corresponding eigenvector $(1, 1 + i)$. Find the matrix A .

3.6. Let $R : \text{Mat}_2(\mathbb{C}) \rightarrow \text{Mat}_2(\mathbb{C})$ be the linear map that rotates matrices a quarter of a turn clockwise like so:

$$R \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & a \\ d & b \end{pmatrix}.$$

- a) Find all eigenvalues and eigenvectors of R . Is R diagonalizable?
- b) Do the same for the corresponding map $\text{Mat}_3(\mathbb{C}) \rightarrow \text{Mat}_3(\mathbb{C})$.

3.7. Let V be a three-dimensional complex vector space with basis (e_1, e_2, e_3) , and let $P : V \rightarrow V$ be the linear map that permutes the basis vectors cyclically: $P(e_1) = e_2$, $P(e_2) = e_3$, $P(e_3) = e_1$. Show that P is diagonalizable and find a new basis of V consisting of eigenvectors of P .

3.8. Let $A = \begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}$. Find all eigenvalues and eigenvectors of the 4×4 -matrix $A^T A$ without writing down the matrix first.

3.9. An integer sequence a_n is defined recursively by $a_0 = 4, a_1 = 6$ and

$$a_n = 2a_{n-1} - 2a_{n-2} \quad \text{for } n \geq 2.$$

Find an explicit expression for a_n .

3.10. Compute $p(A)$ and $p(B)$ where $p(t) = t^4 + 2t^2 - 5t + 3$ and $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$.

3.11. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$. Find a polynomial $p(t)$ for which $p(A) = 0$ by computing I, A, A^2, \dots until these matrices become linearly dependent in $\text{Mat}_n(\mathbb{R})$.

3.12. The Cayley-Hamilton says that $p_A(A) = 0$ where $p_A(t) = \det(A - tI)$. A famous "fake proof" of the theorem goes like this:

$$p_A(A) = \det(A - A \cdot I) = \det(0) = 0.$$

This proof is incorrect, why?

3.13. Verify that the Cayley-Hamilton theorem holds for $A = \begin{pmatrix} -2 & 5 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.

3.14. Let $p(t) = t^2 - 4t + 3$ and $q(t) = (t-1)^2(t-2)^2(t-3)^2$. Compute $p(A)$ and $q(A)$ where $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

3.15. We know that $p_A(\lambda) = \lambda^2 + (-3-i)\lambda + 2 + 2i$. Find all eigenvalues of $B = A^2 + 3A - 5I$.

3.16. We know that a matrix A has characteristic polynomial $p_A(t) = -t^5 - 2t^4 - t^3$. What are the possible expressions for the minimal polynomial $m_A(t)$?

3.17. We know that a matrix A has minimal polynomial $t^2 - 1$. Simplify $A^3 + 2A^2 + 2A$.

3.18. Find the minimal polynomial of each matrix below.

$$\text{a) } A = \begin{pmatrix} 4 & -1 \\ 1 & 2 \end{pmatrix} \quad \text{b) } B = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 4 & 5 & 6 \\ 0 & 0 & 7 & 8 \\ 0 & 0 & 0 & 9 \end{pmatrix} \quad \text{c) } C = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

3.19. Let $R : \text{Mat}_n(\mathbb{C}) \rightarrow \text{Mat}_n(\mathbb{C})$ be the linear operator that rotates a matrix M a quarter of a turn counter-clockwise: $R\left(\begin{bmatrix} A \end{bmatrix}\right) = \begin{bmatrix} \swarrow A \end{bmatrix}$. Find the minimal polynomial of R .

Jordan canonical form

4.1. Which of the following matrices are in Jordan normal form?

$$\text{a) } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix} \quad \text{b) } \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad \text{c) } \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

4.2. For each Jordan-matrix below, determine the algebraic and the geometric multiplicity of each eigenvalue. Also find the characteristic polynomial and the minimal polynomial.

$$\text{a) } \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{pmatrix} \quad \text{b) } \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{c) } \begin{pmatrix} 5 & 1 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

4.3. For a matrix A we know that its characteristic polynomial is $\det(A - \lambda I) = (\lambda - 3)^5(\lambda + 1)^3$. Find the number the possible Jordan forms of A (up to permutation of the blocks).

4.4. How many *different* Jordan forms of nilpotent 6×6 -matrices are there?

4.5. Find two square matrices A and B with *different* Jordan forms such that

- a) A and B has the same characteristic and minimal polynomials.
- b) A and B has the same characteristic and minimal polynomials, and the same dimension of all the eigenspaces.

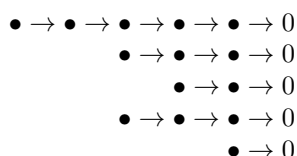
4.6. Show that for any square matrix A , the trace $\text{tr}(A)$ is the sum of the eigenvalues and $\det(A)$ is the product of the eigenvalues (counting multiplicities)

4.7. Prove that for any nilpotent map $N : \mathbb{C}^n \rightarrow \mathbb{C}^n$ we have $N^n = 0$.

4.8. Find all matrices that commute with the Jordan block $J = \begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{pmatrix}$. Generalize your result to matrices that commute with an arbitrary Jordan block.

4.9. For a certain nilpotent operator A we have $p_A(t) = t^7$, $m_A(t) = t^4$, and we know that the geometric multiplicity of the eigenvalue 0 is 3. Determine the Jordan form of A .

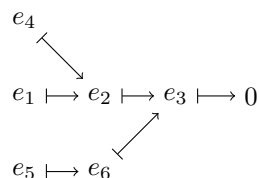
4.10. We know that a nilpotent linear map F on a 14-dimensional vector space has a string basis that looks like the diagram below, where each dot represents a vector in the string basis.



Find the dimension of:

- a) $\ker(F)$ b) $\text{Im}(F)$ c) $\ker(F^3)$ d) $\text{Im}(F^2)$ e) $\ker(F) \cap \text{Im}(F^2)$

4.11. Suppose $F : \mathbb{R}^6 \rightarrow \mathbb{R}^6$ is defined by the diagram below. Without using any matrices, find the characteristic and minimal polynomials for F . Also find a string basis for F .



4.12. For the nilpotent matrix N below, find a matrix S and a matrix J in Jordan form such that $SJS^{-1} = N$.

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

4.13. For the nilpotent matrix M below, find a matrix S and a matrix J in Jordan form such that $SJS^{-1} = M$.

$$M = \begin{pmatrix} 0 & 0 & -1 \\ 2 & -1 & -3 \\ -2 & 1 & 1 \end{pmatrix}$$

4.14. Jordanize the matrix A below. In other words, find a matrix J in Jordan form and an invertible matrix S such that $S^{-1}AS = J$.

$$A = \begin{pmatrix} -1 & 0 & 0 \\ -3 & 2 & 0 \\ 3 & 0 & 2 \end{pmatrix}$$

- 4.15.** Jordanize the matrix A below. In other words, find a matrix J in Jordan form and an invertible matrix S such that $S^{-1}AS = J$.

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 3 & 1 & -2 & 1 \\ -2 & 0 & 4 & 0 \\ 3 & -1 & -2 & 3 \end{pmatrix}$$

- 4.16.** Jordanize the matrix A below. Its only eigenvalue is 2.

$$A = \begin{pmatrix} 1 & 1 & 5 & -1 & 7 \\ -1 & 3 & 9 & 4 & 10 \\ 0 & 0 & 0 & -2 & -2 \\ 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 1 & 1 & 3 \end{pmatrix}.$$

- 4.17.** Jordanize the matrix

$$A = \begin{pmatrix} 9+i & 9 \\ -4 & -3+i \end{pmatrix}.$$

- 4.18.** Find a non-diagonalizable matrix 2×2 -matrix A for which $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector of eigenvalue 3.

- 4.19.** Let $F : \mathcal{P}_3 \rightarrow \mathcal{P}_3$ be the shifting operator $F(p(x)) = p(x+1)$. Find the Jordan form of F .

- 4.20.** Prove that a square matrix A is invertible if and only if 0 is not an eigenvalue of A .

- 4.21.** Prove that if two matrices A and B have the same Jordan form J , then A and B are similar.

- 4.22.** Let U be the subspace of $\mathcal{C}(\mathbb{R})$ generated by $(e^x, xe^x, x^2e^x, \sin(x), \cos(x))$, and let D be complexification of the differentiation operator acting on U . Find the Jordan form of D .

- 4.23.** Prove that if A is an operator whose minimal polynomial has simple zeros (the multiplicity of each zero is 1), then A is diagonalizable.

- 4.24.** Prove that for $n \geq 2$, the Jordan block $J_n(0)$ does not have a square root: no matrix X satisfies $X^2 = J_n(0)$.

- 4.25.** Suppose a matrix A has the Jordan form J below. Find the Jordan form of A^n for each $n > 0$.

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

- 4.26.** Find the Jordan form of $\begin{pmatrix} 2 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix} \in \text{Mat}_3(\mathbb{Z}_3)$.

- 4.27.** Find a matrix in $\text{Mat}_2(\mathbb{Z}_2)$ which does not admit a Jordan-decomposition, meaning that $A \neq SJS^{-1}$ for any $S, J \in \text{Mat}_2(\mathbb{Z}_2)$.

- 4.28.** Compute A^n for the Jordan matrix

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & -6 & 1 \\ 0 & 0 & 0 & 0 & -6 \end{pmatrix}.$$

4.29. Compute e^A and e^B for each of the matrices below.

$$A = \begin{pmatrix} -3 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

4.30. Compute e^J where J is the matrix below.

$$J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

4.31. Suppose that v is an eigenvector of A with eigenvalue λ . What can be said about eigenvalues and vectors of the matrix e^A ?

4.32. Compute $\sin(A_k)$ and $\cos(A_k)$ for $k = 1, 2, 3$ where

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad A_2 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad A_3 = \begin{pmatrix} \frac{\pi}{3} & 1 & 0 & 0 \\ 0 & \frac{\pi}{3} & 0 & 0 \\ 0 & 0 & \frac{\pi}{2} & 0 \\ 0 & 0 & 0 & \pi \end{pmatrix}.$$

4.33. Prove that $\frac{d}{dt} \sin(At) = A \cos(At)$ and that $\frac{d}{dt} \cos(At) = -A \sin(At)$. Find a differential equation that can be solved using this fact.

4.34. Show that if $\text{tr}(A) = 0$, then $\det(e^A) = 1$.

4.35. A discrete dynamical system evolves according to the model

$$\begin{cases} a_{n+1} = -2a_n + b_n \\ b_{n+1} = -a_n - 4b_n \end{cases} \quad \text{where } a_0 = 2 \text{ and } b_0 = 0.$$

Find explicit expressions for a_n and b_n and determine the limit of $\frac{a_n}{b_n}$ as $n \rightarrow \infty$.

4.36. Jordanize the matrix below by finding S and J such that $SJS^{-1} = A$.

$$A = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 3 & 0 \\ -4 & 7 & -1 \end{pmatrix}$$

4.37. Solve the initial value problem

$$\begin{cases} x'_1(t) = x_1(t) + x_2(t) \\ x'_2(t) = -x_1(t) + 3x_2(t) \\ x'_3(t) = -4x_1(t) + 7x_2(t) - x_3(t) \end{cases} \quad x_1(0) = 0, \quad x_2(0) = 1, \quad x_3(0) = 0.$$

Inner products and norms

5.1. Let $\langle \cdot, \cdot \rangle$ be the standard inner product on \mathbb{C}^2 : $\langle (x_1, x_2), (y_1, y_2) \rangle = x_1 \overline{y_1} + x_2 \overline{y_2}$. Find:

a) $\langle (i, 1+i), (3, 1+2i) \rangle$ b) $\|(1+i, 3)\|$ c) All vectors orthogonal to $(1, 1+i)$

5.2. Find the length of $p(x) = x^2 + x + 1$ in the real vector space of polynomials equipped with the inner product $\langle p(x), q(x) \rangle = \int_0^1 p(x)q(x)dx$.

- 5.3.** Consider a complex inner product space V . Show that the inner product function is conjugate-linear in the second argument:

$$\langle u, \lambda v + \mu w \rangle = \bar{\lambda} \langle u, v \rangle + \bar{\mu} \langle u, w \rangle.$$

- 5.4.** Which of the following rules define inner products on the given vector space? In case it is not an inner product, give an example of an axiom that is being violated.

- a) $\langle (x_1, y_1), (x_2, y_2) \rangle = 2x_1x_2 + y_1y_2$ on \mathbb{R}^2 .
- b) $\langle (x_1, y_1), (x_2, y_2) \rangle = x_1x_2 + x_1y_2 + y_1y_2$ on \mathbb{R}^2 .
- c) $\langle (x_1, y_1), (x_2, y_2) \rangle = x_1y_1 + y_1y_2$ on \mathbb{C}^2 .
- d) $\langle (x_1, y_1), (x_2, y_2) \rangle = x_1\bar{x}_2$ on \mathbb{C}^2 .
- e) $\langle x, y \rangle = |xy|$ on \mathbb{R}^2 .
- f) $\langle f(x), g(x) \rangle = \int_0^1 f(x)g(x)dx$ on \mathbb{R} (the real vector space of continuous functions $\mathbb{R} \rightarrow \mathbb{R}$).
- g) $\langle A, B \rangle = \text{tr}(A + B)$ on $\text{Mat}_{n \times n}(\mathbb{R})$.
- h) $\langle A, B \rangle = \text{tr}(AB)$ on $\text{Mat}_{n \times n}(\mathbb{R})$.

- 5.5.** Let $\langle \cdot, \cdot \rangle$ be an inner product on a real vector space V . Show that $(u|v) := 2\langle u, v \rangle$ defines a new inner product on V . Inner products are used to define lengths and angles in V - how are lengths and angles changed by using $(\cdot|\cdot)$ instead of $\langle \cdot, \cdot \rangle$ as the inner product on V ?

- 5.6.** Does there exist an inner product $\langle \cdot, \cdot \rangle$ on \mathbb{C}^2 for which $u = (1, 1)$ and $v = (1, i)$ are orthogonal? Answer the same question for $u' = (1, 1 + i)$ and $v' = (1 - i, 2)$?

- 5.7.** Let V be an inner product space. Show that the norm arising from the inner product satisfies the parallelogram law:

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

- 5.8.** Let V be an inner product space. Show that the norm arising from the inner product is indeed a norm in the sense of Definition 5.6: It satisfies absolute homogeneity, the triangle inequality, and positivity:

$$\|\lambda v\| = |\lambda| \cdot \|v\|, \quad \|u\| + \|v\| \geq \|u + v\|, \quad \|v\| \geq 0 \text{ with equality only for } v = 0.$$

- 5.9.** Show that the maximum norm on \mathbb{R}^2 can not be derived from an inner product.

- 5.10.** Let V be a complex vector space. Then the set of all functions $V \times V \rightarrow \mathbb{C}$ is also a vector space that we call \mathcal{F} . Let \mathcal{I} be the set of all inner products on V . Is \mathcal{I} a subspace of \mathcal{F} ?

- 5.11.** Find the distance between $(2, 1, 3)$ and $(4, -2, 6)$ with respect to:

- a) The standard norm
- b) The maximum norm
- c) The Manhattan norm
- d) The p -norm for $p = 3$

- 5.12.** For the matrix $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, find the Frobenius norm $\|A\|_F$ and the spectral norm $\|A\|_\sigma$.

- 5.13.** Prove that an equivalent definition of the operator norm of $F : V \rightarrow W$ is

$$\|F\|_{op} = \max_{v \neq 0} \left\{ \frac{\|F(v)\|}{\|v\|} \right\}.$$

5.14. Let $C = \{(\cos^{\frac{3}{2}}(t), \sin^{\frac{3}{2}}(t)) \mid t \in \mathbb{R}\}$. Find p such that C is the unit circle with respect to the p -norm.

5.15. Show that the definition of the p -norm does in fact not give a norm for $0 < p < 1$.

5.16. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are called *equivalent* if there exists positive constants C and D such that

$$C\|v\|_2 \leq \|v\|_1 \leq D\|v\|_2$$

holds for all $v \in V$. Show that on \mathbb{R}^2 , the maximum norm is equivalent to the standard norm, find the maximal possible C and the minimal possible D .

5.17. Show that if any two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent (as defined in the previous problem), then $v_n \rightarrow v$ with respect to $\|\cdot\|_1$ if and only if $v_n \rightarrow v$ with respect to $\|\cdot\|_2$.

5.18. The familiar vector product on \mathbb{R}^3 is given by

$$(x_1, y_1, z_1) \times (x_2, y_2, z_2) = (y_1 z_2 - y_2 z_1, -x_1 z_2 + x_2 z_1, x_1 y_2 - x_2 y_1).$$

With this rule we know that $u \times v$ is orthogonal to both u and v . Does the same property hold if we use the same definition of a vector product on \mathbb{C}^3 ? Give a proof or a counter-example.

5.19. On \mathbb{R}^3 , define $((x_1, x_2, x_3) | (y_1, y_2, y_3)) := x_1 y_1 + x_2 y_2 - x_3 y_3$. Show that $(\cdot | \cdot)$ is not an inner product, and describe the set of all $v \in \mathbb{R}^3$ which has length zero (meaning $(v | v) = 0$)

5.20. Let U be a subspace of a finite-dimensional inner product space V . Recall that

$$U^\perp := \{v \in V \mid \langle v, u \rangle = 0 \text{ for all } u \in U\}.$$

a) Show that U^\perp is a subspace of V .

b) Show that $V^\perp = \{0\}$, in other words, the only vector orthogonal to all of V is the zero vector.

c) Show that $U \cap U^\perp = \{0\}$.

5.21. In an inner product space V , for $u, v \in V$ with $u \neq 0$, define

$$P_u(v) = \frac{\langle v, u \rangle}{\langle u, u \rangle} u.$$

a) Show that $P_u(v) = P_{\lambda u}(v)$ for complex $\lambda \neq 0$

b) Show that $P_u : V \rightarrow V$ is a linear map

c) Show that $\text{Im}(P_u) = \text{span}(u)$ and that $\ker(P_u) = \text{Im}(P_u)^\perp$

d) Show that $v - P_u(v)$ is orthogonal to u .

e) Show that $P_u^2 = P_u$

5.22. Prove that if (e_1, \dots, e_n) is an orthonormal basis in an inner product space, then

$$v = \langle v, e_1 \rangle e_1 + \dots + \langle v, e_n \rangle e_n.$$

5.23. Prove the *Pythagorean theorem* in the inner product space setting: If u is orthogonal to v we have

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

5.24. Let $U \subset V$ be a subspace, and fix $v \in V$. Show $u = P_U(v)$ is the vector in U closest to v .

5.25. Consider \mathbb{C}^3 with the standard inner product, and let

$$U = \text{span}((1, i, 0), (0, 2i, 1))$$

be a subspace. Use the Gram-Schmidt process to find an ON-basis for U , and extend this basis to an ON-basis for \mathbb{C}^3 .

5.26. Find the function $g(x) \in \mathcal{P}_1$ that best approximates e^x with respect to the standard inner product on $\mathcal{C}[0, 1]$:

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx.$$

5.27. Consider the space \mathcal{F} of 2π -periodic real valued continuous functions equipped with the inner product

$$\langle f, g \rangle := \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)g(x)dx.$$

Show that

$$\left\{ \frac{1}{\sqrt{2}}, \sin(x), \cos(x), \sin(2x), \cos(2x), \sin(3x), \cos(3x), \dots \right\}$$

is an orthonormal set of functions in \mathcal{F} , in other words, show that these functions are pairwise orthogonal, and that each function has length 1 with respect to our given inner product.

5.28. Let $f(x)$ be a *triangular wave function* with period 2π which equals $|x|$ on $[-\pi, \pi]$. Find the function of form $g(x) = a + b \cos(x) + c \sin(x)$ that best approximates f (with respect to the inner product from the previous problem).

5.29. Find a QR-factorization of the matrix $A = \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix}$. In other words, find a matrix Q which is unitary ($Q^*Q = I$), and a matrix R which is upper triangular, such that $QR = A$.

5.30. Find a QR-factorization of the matrix $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 3 & 2 \\ 1 & 1 & 3 \end{pmatrix}$.

5.31. Find QR-factorizations of $A = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 4 & 1 \end{pmatrix}$ and of $B = \begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix}$.

5.32. Let $F : V \rightarrow W$ be a linear operator between inner product spaces. Recall that the adjoint of F is the map $F^* : W \rightarrow V$ where $F^*(w)$ is defined as the unique vector in V such that $\langle F(v), w \rangle = \langle v, F^*(w) \rangle$ for all $v \in V$. Show that F^* is a linear map (without using the fact that $[F^*] = [F]^*$).

5.33. Show $(G \circ F)^* = F^* \circ G^*$ and that $(F^*)^* = F$.

5.34. Let $F : V \rightarrow W$, and pick ON-bases in V and in W . Prove that $[F^*] = [F]^*$, in other words, the matrix of the adjoint of F is the Hermitian conjugate of the matrix for F with respect to the same ON-bases for V and for W .

5.35. Let operators F, G, H be given by the following matrices with respect to an ON-basis:

$$[F] = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad [G] = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad [H] = \frac{1}{5} \begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}$$

Which of the operators F, G, H are...

a) Self-adjoint? b) Unitary? c) Normal?

5.36. Find an operator that is normal, but that is neither unitary nor self-adjoint.

5.37. Determine for what $a, b \in \mathbb{C}$ the matrix $A = \frac{1}{5} \begin{pmatrix} 3 & a \\ b & 3 \end{pmatrix}$ is...

- a) Self adjoint b) Unitary c) Normal

Then answer the same question in the special case when $a, b \in \mathbb{R}$.

5.38. Show that the spectral radius of a unitary operator is always 1.

5.39. Let $F : V \rightarrow V$ be unitary. Show that F preserves inner products: $\langle F(u), F(v) \rangle = \langle u, v \rangle$ for all $u, v \in V$.

5.40. Prove that self-adjoint maps are normal and that unitary maps are normal.

5.41. Let $F : V \rightarrow W$ be a linear map between inner product spaces. Show that:

a) $\ker(F^*) = \text{Im}(F)^\perp$

b) $\text{Im}(F^*) = \ker(F)^\perp$

5.42. Prove that compositions of unitary operators are unitary.

5.43. Let A be a normal matrix: $AA^* = A^*A$.

a) Show that $A + \lambda I$ is normal for all λ .

b) Show that $\ker(A) = \ker(A^*)$.

c) Show that if $Av = \lambda v$ then $A^*v = \bar{\lambda}v$.

d) Show that if u and v are eigenvectors of A corresponding to different eigenvalues, then $u \perp v$.

5.44. Recall that a matrix A is called positive definite if $X^*AX > 0$ for all nonzero column-matrices $X \in \mathbb{C}^n$.

a) Show that a Hermitian matrix need not be positive definite.

b) Prove that if a matrix is positive definite then it must be Hermitian.

c) Find a non-Hermitian real matrix A such that $X^*AX > 0$ for all nonzero $X \in \mathbb{R}^n$.

d) Why does the result in (c) not contradict the result in (b)?

5.45. For what $a, b, c \in \mathbb{C}$ is the matrix $A = \begin{pmatrix} 2 & 1+i & 1 \\ a & b & 1 \\ 1 & 1 & c \end{pmatrix}$ positive definite?

5.46. Prove Sylvester's criterion for diagonal matrices: If D is diagonal it is positive definite if and only if all principal minors are positive.

5.47. Prove one of the implications in Sylvester's criterion: If one principal minor of a matrix is negative, then the matrix is not positive definite.

5.48. Write the matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ as a linear combination of two matrices which are orthonormal with respect to the Frobenius inner product.

5.49. For square matrices A and B we define

$$A \prec B \quad \Leftrightarrow \quad B - A \text{ is positive semi-definite.}$$

Prove that \prec is a partial order relation; in other words, show that

- $A \prec A$ for all A

- If $A \prec B$ and $B \prec C$, then $A \prec C$
- If $A \prec B$ and $B \prec A$, then $A = B$

5.50. Find the square root of the matrix $A = \frac{1}{5} \begin{pmatrix} 17 & 6 \\ 6 & 8 \end{pmatrix}$.

7 Hints

1.1. The equality looks obvious, but note that $+$ and \cdot can be defined in some non-standard way. Start from the left hand side and simplify it using the axioms. Start with the axiom that says that $1 \cdot v = v$.

1.2. The zero element $\mathbf{0}$ is the unique element of \mathbb{R}_+ satisfying $\mathbf{0} + v = v$ for all $v \in \mathbb{R}_+$.

1.3. Recall that a (real) subspace of V is a non-empty subset $S \subset V$ such that:

- $u, v \in S \Rightarrow u + v \in S$ (additivity)
- $\lambda \in \mathbb{R}, v \in V \Rightarrow \lambda \cdot v \in S$ (homogeneity).

To show that S is a subspace, these conditions have to be tested for all vectors and scalars. To show that S is not a subspace, it suffices to find a single counter-example to the conditions.

1.4. The projection is not orthogonal. Express $v = u + u'$ with $u \in U$ and $u' \in U'$.

1.5. For the basis, note that if a polynomial $p(x)$ lies in S , so does all its multiples $q(x)p(x)$.

1.6.

1.7. Look for a counterexample in $V = \mathbb{R}^2$

1.8. A function f is even if $f(-x) = f(x)$ for all x and odd if $f(-x) = -f(x)$. Show these properties are preserved when taking sums of functions or products of functions by scalars. To show that each f can be expressed as $e(x) + o(x)$, note that $f(x) + f(-x)$ is always even and $f(x) - f(-x)$ is odd.

1.9. For each map F , check whether $F(u + v) = F(u) + F(v)$ and $F(\lambda v) = \lambda F(v)$ holds for all vectors u, v in the domain of F and all scalars λ .

1.10.

1.11. Recall that the definition of F being injective is that $F(u) = F(v)$ only when $u = v$.

1.12. Recall the definitions. A map $F : V \rightarrow V$ is injective if $\ker(F) = 0$, surjective if $\text{Im}(F) = V$, and has an inverse G if both $G \circ F$ and $F \circ G$ is the identity map on V .

1.13. Consider where the standard basis $(1, x, x^2, x^3, x^4)$ is mapped, the images form the columns of $[F]$. For the inverse, figure it out without matrices.

1.14. The matrices of F and G will be 3×3 where the columns are the images of the basis vectors expressed in the given basis.

1.15. Recall that $v + U = \{v + u \mid u \in U\}$.

1.16.

- a) Two affine subsets are equal $v + U = w + U$ if and only if $v - w \in U$.
- b) Each subset is a line in \mathbb{R}^2 through the given point and in the direction $(1, 1)$.
- c) The coordinate for B in the basis (A) is the single number λ such that $\lambda A = B$, or in other words $\lambda(2, 3) + U = (0, 1) + U$.

1.17. Visually, elements of V/U are all the lines in direction $(1, 1, 1)$. The coordinates (λ_1, λ_2) you seek must satisfy $\lambda_1 A + \lambda_2 B = C$, these can be found by solving $\lambda_1(1, 1, 0) + \lambda_2(0, 1, 1) + \lambda_3(1, 2, 3) = (1, 1, 1)$.

1.18. e_1, e_2, e_3 are linearly dependent if $\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3 = 0 + \ell$ (the right side is the zero vector in V). This means that

$$\lambda_1(1, 1, 0, 0) + \lambda_2(0, 1, 1, 0) + \lambda_3(0, 0, 1, 1) = \lambda_4(3, 2, 1, 2),$$

solve this linear system.

1.19. If $v - w \in U$, then $v + U = w + U$, so we must have $\tilde{F}(v) = \tilde{F}(w)$ in order for \tilde{F} to be well-defined. The second question is vague, since it depends on what bases we pick, but consider a bases for (u_1, \dots, u_m)

of U and extend it to a basis $(u_1, \dots, u_m, v_1, \dots, v_n)$ of V , then a natural choice of basis for V/U is $(v_1 + U, \dots, v_n + U)$, consider how the matrices look with respect to these bases.

1.20. Note that if p_1, p_2 lies in an affine subset $v + U$, then $p_1 - p_2 \in U$, so U is spanned by all differences between the points.

1.21. For (b), to find the translation vector w , consider where the origin is mapped.

1.22. Basically everything you can do over a real vector space you can do the same way over \mathbb{Z}_3 . Recall that the only scalars are 0, 1, 2 though.

1.23. Note that $v = -v$ when the field is \mathbf{Z}_2 .

1.24. It remains to show that the property $F(\lambda v) = \lambda F(v)$ holds for all $\lambda \in \mathbb{Q}$. Start by showing that it holds for $\lambda \in \mathbb{Z}$.

2.1. Recall that $A^* = \overline{A}^T$, the conjugate transpose of A . A matrix is Hermitian if $A^* = A$.

2.2. That A is symmetric means that $A^T = A$ and that B is skew-symmetric means that $B^T = -B$.

2.3. Let A be $m \times n$ and let B be $n \times m$, use the sum-version of matrix multiplication: $(AB)_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$ - what are the diagonal elements of AB ?

2.4.

a) Since A is diagonalizable, $A = SDS^{-1}$, where $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. Apply the previous problem to prove that $\text{tr}(A) = \text{tr}(D)$.

b) $(S^{-1}AS)^n = D^n$

c) (the eigenvalues are integers)

2.5. Compute P^2 and P^3

2.6.

a) Count the 1's of the matrix first row by row, then column by column.

b) Think of the two matrices as performing permutations of the basis vectors.

c) Same hint as above.

d) You need at least a 5×5 -matrix.

2.7. For example, let $X = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}$ and solve the linear system $AX = XA$

2.8. Suppose that $A \in \mathfrak{C}$, and compute $e_{ij}A$ and Ae_{ij} for different choices of (i, j) (where e_{ij} is a matrix with a single 1 in position (i, j) and zeroes elsewhere), what does this say about the entries of A ?

2.9. $E = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is an idempotent.

2.10. Diagonalize A so that $S^{-1}AS = D$. Then A commutes with B , if and only if D commutes with $S^{-1}BS$, so $\dim \mathcal{C}_A = \dim \mathcal{C}_D$, so it suffices to investigate what matrices commute with a given diagonal matrix. Consider first the case where all eigenvalues are distinct, then the case where all eigenvalues coincide, and finally the case where some of them coincide.

2.11. Assume that $M^m = 0$ and $N^n = 0$, what can be said about $(AB)^{m+n}$ and $(M+N)^{m+n}$? For $M+N$, think about the binomial theorem. For the case when A and B do not commute, look for a 2×2 counterexample.

2.12. What can be said about $F(F(v))$?

- 2.13.** Consider the sequence of subspaces $\ker(N) \subset \ker(N^2) \subset \dots$. Show that the dimensions of these subspaces must be *strictly* increasing. (What would it mean if $\ker(N^k) = \ker(N^{k+1})$ for some k ?)
- 2.14.** You can write down $(I + N)^{-1}$ explicitly, think of the formula for geometric sums and its proof.
- 2.15.** Multiply M by M^{-1} in block form and show that the product is the identity matrix.
- 2.16.** Make an ansatz $M^{-1} = \left(\begin{array}{c|c} X_1 & X_2 \\ \hline X_3 & X_4 \end{array} \right)$ and multiply by M . Both MM^{-1} and $M^{-1}M$ should equal the identity matrix $\left(\begin{array}{c|c} I & 0 \\ \hline 0 & I \end{array} \right)$.
- 2.17.** The rank is the number of pivots in the REF (or RREF). $\ker(A)$ is the set of solutions to $AX = 0$, which is the same as the set of solutions to $\tilde{A}X = 0$ where \tilde{A} is the reduced row echelon form of A .
- 2.18.** Row operations works exactly the same over \mathbb{C} . Recall that a complex fraction can be simplified by multiplying numerator and denominator by the complex conjugate of the denominator.
- 2.19.** Remember that the parameters r, s, t correspond to non-pivot columns, find the rows of the RREF from bottom to top.
- 2.20.** $\text{Im}(A)$ is spanned by the columns of A , to remove linearly dependent columns, find the echelon form of A , and remove columns of A corresponding to non-pivot columns in the echelon form. $\ker(A)$ consists of all solutions X to $AX = 0$, such X can easily be found from the echelon form. The parametrization of the solutions yields a basis.
There are sometimes quick ways to see what the intersection is, but recall the standard algorithm: If u_1, \dots, u_m is a basis for U and v_1, \dots, v_n is a basis for V , then a vector w lies in the intersection if $\lambda_1 u_1 + \dots + \lambda_m u_m = w = \lambda_{m+1} v_1 + \dots + \lambda_{m+n} v_n$ has a solution, in other words there exists a $(m+n)$ -tuple $(\lambda_1, \dots, \lambda_{m+n})$ satisfying this equation. One can find all such λ_i by row-operating on the matrix which has all the $m+n$ vectors as columns.
- 2.21.** For example, consider a linear map on \mathbb{R}^4 such that $e_1 \mapsto e_2 \mapsto e_3 \mapsto 0$ and $e^4 \mapsto 0$. What are the three subspaces?
- 2.22.** Think of the matrices as performing row operations - what is the opposite of a given row operation?
- 2.23.** Multiplying A by elementary matrices from the **left** corresponds to doing row operations on A . Multiplication on the right corresponds to column operations.
- 2.24.** Reduce A to the identity matrix by multiplying by elementary matrices on the left (each multiplication corresponding to a row operation). If $E_3 E_2 E_1 A = I$, then $A = (E_3 E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} E_3^{-1}$. The factorization of the inverse is trivial once you have the factorization of A .
- 2.25.** Row operations on $[A|I]$ can be realized as multiplying the block matrix by elementary matrices on the left. Consider what happens to the left and to the right part.
- 2.26.** By multiplying A on the left by $E = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, we reduce A to echelon form: $EA = U$. Then $A = E^{-1}U$ is the LU decomposition.
- 2.27.** If $E_3 E_2 E_1 A = U$ where U is in row echelon form, then $A = (E_1^{-1} E_2^{-1} E_3^{-1})U = LU$ gives the LU -decomposition.
- 2.28.** After the first two standard row operations, note that rows 2 and 3 will be in the wrong order. So start by multiplying A by a permutation matrix P that switches rows 2 and 3, then PA admits an LU -factorization, proceed as usual.
- 2.29.** Such a dependence relation can be written $Av = 0$ where v is a column vector of the coefficients λ_i .
- 2.30.** Think of how a matrix in column echelon form looks, and think what happens when reducing it to row echelon forms
- 2.31.** First use the element in position $(1,1)$ to get zeros in positions $(2,1)$ and $(3,1)$ and $(4,1)$. This takes

4 multiplications each. Then use the element in position $(2, 2)$ to get zeros in positions $(3, 2)$ and $(4, 2)$, this takes 3 more multiplications each...

2.32. Recall that a Cholesky-factorization of A is $A = CC^*$ where C is lower triangular. First find the LDU-factorization $A = LDU$, then D is diagonal with positive entries, so find a diagonal matrix \tilde{D} such that $\tilde{D}^2 = D$ and let $C = L\tilde{D}$.

2.33. Same hint as the previous problem

2.34. For the first part it is enough to show that CC^* is Hermitian. For the second part it suffices to find a Hermitian matrix that can not be factored as CC^* - look for the smallest possible counter-example.

2.35. Remember that over \mathbb{Z}_3 there are only three scalars: 0, 1, 2, so any row operations can only involve these. Start by multiplying the first row by 2 to get a one top left. Remember that the solutions X should also be vectors in $(\mathbb{Z}_3)^4$.

3.1. Think about where vectors on the lines/planes/axes are mapped to.

3.2.

3.3. Compute $F(x^n)$.

3.4. We know that $Av = \lambda v$, take the complex conjugate. What can be said about the complex conjugate of a matrix product?

3.5. Use the previous problem to obtain S, D such that $A = SDS^{-1}$.

3.6. Assume that v is an eigenvector - what can be said about $R^4(v)$?

3.7. Find the eigenvalues and eigenvectors of P , it may help to introduce some new notation for the eigenvalues.

3.8. Consider $(A^T A)X$ when $X \in \mathbb{R}^4$ parallel with $(1, 2, 3, 4)$. Then consider the case when $AX = 0$.

3.9. Write $X_n := \begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix}$. Then $X_0 = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$ and $X_{n+1} = AX_n$ where $A = \begin{pmatrix} 2 & -2 \\ 1 & 0 \end{pmatrix}$. It follows that $X_n = A^n X_0$. This becomes easy to evaluate if X_0 is written as a linear combination of eigenvectors for A .

3.10. Find A^2 and A^4 , the definitions says that $p(A) = A^4 + 2A^2 - 5A + 3I$. Do the same for B .

3.11. The matrices (vectors) I, A, A^2 are linearly dependent. To find the dependence relation, express these three matrices in the standard basis $\mathbf{e} = (e_{11}, e_{12}, e_{21}, e_{22})$, and use the standard method.

3.12. Take a small (say 2×2) matrix A and write down the matrix $A - tI$, what happens when we try to replace t by A ?

3.13. $p_A(t) = -(t+2)^2(t-3)$, inserting A directly into this factorized form makes the computation trivial.

3.14. Factor $p(t)$ and note that $p(t)$ is a factor in $q(t)$.

3.15. Find the eigenvalues of A by factoring p_A , then use the spectral mapping theorem.

3.16. $m(t)$ is monic and divides $p(t)$, and each root of $p(t)$ is a root of $m(t)$.

3.17. Divide $t^3 + 2t + 2$ by $t^2 - 1$

3.18. For (a) the characteristic polynomial is $(t-3)^2$, so the only possible minimal polynomials are $m(t) = t-3$ or $m(t) = (t-3)^2$, remember that $m(A)$ should be the zero matrix. In (b) and (c) the matrix is triangular so the characteristic polynomial is easy to find.

3.19. What does the map R^4 do?

4.1. Try visualizing the Jordan-blocks on the diagonal. Each block should have the same value one the diagonal and ones on the super-diagonal

- 4.2.** The matrices are all on Jordan-form, so the algebraic multiplicity of λ is the number of occurrences of λ on the diagonal. Each Jordan block with λ 's on the diagonal corresponds to a single eigenvector for that eigenvalue, so the number of Jordan blocks corresponding to the eigenvalue λ is the geometric multiplicity of λ .
- 4.3.** The eigenvalues 3 and -1 can be treated independently. For the eigenvalue -1 there are three possibilities of the sizes of Jordan blocks: $(3), (2, 1), (1, 1, 1)$ corresponding to integer partitions of 3.
- 4.4.** Different Jordan forms correspond to decreasing sequences that add up to 6
- 4.5.** You need at least size 4×4 for (a), and 7×7 for (b). Take all eigenvalues as the same. Look for A and B with different partitions of Jordan blocks. Remember that for a given eigenvalue, the minimal polynomial determines the size of the largest Jordan block, and the geometric multiplicity is the number of Jordan blocks.
- 4.6.** We know that any matrix A can be Jordanized: $A = SJS^{-1}$ where J is in Jordan form.
- 4.7.** What eigenvalues can N have? Jordanize N , with $SNS^{-1} = J$.
- 4.8.** Try the 2×2 and the 3×3 case first.
- 4.9.** The given data decoded says that the matrix is of size 7×7 with zeros on the diagonal, the largest Jordan block should be of size 4, and there should be three Jordan blocks.
- 4.10.** Remember, a vector represented by a dot is in $\ker(F)$ if it is mapped to zero, it is in $\ker(F^2)$ if it lands in zero after moving along the arrows in one or two steps, and so on. The image consists of dots that are pointed to by some arrow, etc.
- 4.11.** F is clearly nilpotent, so zero is the only eigenvalue. For what n is F^n first zero? To get started on the string basis, consider where $e_1 - e_4$ is mapped.
- 4.12.** Find $\ker(N)$ and $\ker(N^2)$, think about what the chains should look like (or follow the standard algorithm).
- 4.13.** Same hint as the last problem.
- 4.14.** Follow the standard algorithm, it is easy in this case to find the characteristic polynomial.
- 4.15.** In this case, finding the characteristic and minimal polynomials will give enough information to determine the Jordan form. Find a basis for each eigenspace and extend to string bases.
- 4.16.** Since 2 is the only eigenvalue, the matrix $A - 2I$ will be nilpotent. A string basis for $A - 2I$ will be a Jordan basis for A , so just follow Algorithm 4.14.
- 4.17.** Proceed as usual, the only difference in this case is that the characteristic polynomial have non-real roots.
- 4.18.** You know the Jordan form and the first column of the matrix S , complete the matrix S in any way, and use that $SJS^{-1} = A$.
- 4.19.** The only eigenvalue is clearly 1. So the operator $F - \text{id}$ is nilpotent.
- 4.20.** Without loss of generality you can assume that A is in Jordan form.
- 4.21.** Remember that A and B are called similar if there exists T such that $TAT^{-1} = B$. Combine that $A = S_1JS_1^{-1}$ and that $B = S_2JS_2^{-1}$.
- 4.22.** Either find the matrix form of D relative to the given basis for U and go for there. Or think first, perhaps you can guess the eigenvalues and (generalized) eigenvectors.
- 4.23.** Pick a new basis for which the matrix of the operator is in Jordan form, then the same statement is simple to prove.
- 4.24.** Suppose $X^2 = J_n(0)$ where $X = SJS^{-1}$ is the Jordan form of X , consider what the matrix J must look

like.

- 4.25.** Consider the two Jordan blocks separately. Note that while A^n is similar to J^n , this latter matrix may not be in Jordan form, in particular A^2 is not in Jordan form.
- 4.26.** Follow the standard algorithm, but remember that all numbers, polynomial coefficients, and matrix elements are in \mathbb{Z}_3 . In particular, the matrices S and J should also belong to $\text{Mat}_3(\mathbb{Z}_3)$.
- 4.27.** The only way such a matrix can fail to have a Jordan-decomposition is if the characteristic polynomial doesn't factor completely over \mathbb{Z}_2 . The characteristic polynomial has degree 2 and there are only four polynomials of degree 2 in $\mathbb{Z}_2[x]$, namely x^2 , $x^2 + 1$, $x^2 + x$, $x^2 + x + 1$. Only one of them can not be factored, find a matrix which has this characteristic polynomial.
- 4.28.** You can consider each of the two Jordan blocks separately, write each block as $J = \lambda I + N$ and use the binomial theorem.
- 4.29.** Recall the definition: $e^A := \sum_{k=0}^{\infty} \frac{A^k}{k!}$. In A , consider the two diagonal positions of e^A separately. B is nilpotent so all but the first terms of the sum disappear.
- 4.30.** Write $J = D + N$ where D is diagonal and N is nilpotent. Verify that N and D commute, so that Show that $e^{D+N} = e^D e^N$. Compute the right side by considering the blocks separately.
- 4.31.** $Av = \lambda v$, so $e^A v = (I + A + \frac{A^2}{2} + \cdots)v = \dots$
- 4.32.** Recall that $\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$ and $\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$. A_1 is nilpotent, use Proposition 4.27 for A_2 and A_3 and consider the Jordan blocks separately.
- 4.33.** Follow the proof of the fact that $\frac{d}{dt} e^{At} = A e^{At}$. For a differential equation, if $X = \sin(At)C$, what is $X''(t)$?
- 4.34.** Trace and determinant is basis-independent, so you can assume that A is a Jordan matrix without loss of generality. Consider the diagonal elements of e^A .
- 4.35.** Follow the example with rabbits/foxes. Start by introducing $X_n = \begin{pmatrix} a_n \\ b_n \end{pmatrix}$ and expressing the system in matrix form, then Jordanize the coefficient matrix.
- 4.36.** The eigenvalues are 2 and -1 , so study the kernels of $(A - 2I)^k$ and of $(A + I)$ to find (generalized) eigenvectors forming Jordan chains
- 4.37.** The system can be written $X'(t) = AX(t)$ where A is the same as in the previous problem!
- 5.1.** Recall that the norm is defined via the inner product $\|v\| := \sqrt{\langle v, v \rangle}$, and that vectors are said to be orthogonal if their inner product is zero.
- 5.2.** $\|x^2 + x + 1\|^2 = \langle x^2 + x + 1, x^2 + x + 1 \rangle$
- 5.3.** Use symmetry, linearity in the first argument, and symmetry again.
- 5.4.**
- 5.5.** The axioms for an inner product are easy to verify. In the first case IP-space, $\|v\|_1 = \sqrt{\langle v, v \rangle}$, in the second it is $\|v\|_2 = \sqrt{\langle v, v \rangle}$. Recall that the angle between two vectors u and v of a real IP-space is defined as the number θ between 0 and π satisfying $\|u\| \cdot \|v\| \cos(\theta) = \langle u, v \rangle$.
- 5.6.** Is it possible to define an inner product such that (u, v) is an ON-basis?
- 5.7.** Expand the left side using the fact that $\|u \pm v\|^2 = \langle u \pm v, u \pm v \rangle$.
- 5.8.** For the triangle inequality, expand $\|u + v\|^2 = \langle u + v, u + v \rangle$ using sesqui-linearity, you will have to use Cauchy-Schwarz in one step.
- 5.9.** Recall that $\|(x, y)\|_{\max} = \max\{|x|, |y|\}$. It is enough to show that the norm does not satisfy the parallelogram law.

- 5.10. \mathcal{I} is clearly a subset of \mathcal{F} . Is the sum of two inner products an inner product? Is a scalar times an inner product still an inner product?
- 5.11. The distance is defined as $\|(2, 1, 3) - (4, -2, 6)\| = \|(2, 1, -4)\|$, so compute this for the various definitions of the norm.
- 5.12. $\|A\|_F = \sqrt{\langle A, A \rangle_F}$ where $\langle A, B \rangle_F = \text{tr}(AB^*)$. $\|A\|_\sigma = \max\{|\lambda| \mid \lambda \in \sigma(A)\}$.
- 5.13. Normalize v and write $v = \|v\| \left(\frac{1}{\|v\|} v \right) \|v\|$.
- 5.14. Choose p such that the identity $\cos^2(t) + \sin^2(t) = 1$ appear in the norm calculation.
- 5.15. Find an example of vectors violating the triangle inequality.
- 5.16. Look at the pictures of the unit circles in the different norms. How should the circles be scaled to contain one another?
- 5.17. Recall that a sequence of vectors v_n converges to a vector v with respect to a norm $\|\cdot\|$ if and only if $\|v_n - v\| \rightarrow 0$.
- 5.18. Look for a counterexample.
- 5.19. The form is symmetric and bilinear, but not positive definite.
- 5.20. For the first part, verify that for $u_1, u_2 \in U^\perp$ and $\lambda \in \mathbb{C}$ we have $u_1 + u_2 \in U^\perp$ and $\lambda u_1 \in U^\perp$, use linearity in the first argument of the inner product. For the second part, use the positive-definiteness.
- 5.21. The statements follow directly from the definition. They show that we can define *projections* in arbitrary inner product spaces.
- 5.22. Show that the difference $w := v - (\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n)$ is zero by showing that its inner product with each of the basis vectors e_i is zero.
- 5.23. The left hand side is $\langle u + v, u + v \rangle$.
- 5.24. $\|v - u\|^2 = \|(v - P_U(v)) + (P_U(v) - u)\|^2$, show that $v - P_U(v) \in U^\perp$ and $P_U(v) - u \in U$, and apply Pythagorean theorem.
- 5.25. Call the spanning vectors for u_1 and u_2 . Replace the second vector u_2 by itself minus its projection on u_1 , $u'_2 := u_2 - \frac{\langle u_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1$. Then u_1 and u'_2 are orthogonal and span U . Normalize to get an ON-basis. Solve a linear system to get the last basis vector.
- 5.26. Project e^x onto \mathcal{P}_1 , use the basis from Example 5.14.
- 5.27. If f is odd, $\int_{-a}^a f(x)dx = 0$. If g is even, $\int_{-a}^a g(x)dx = 2 \int_0^a g(x)dx$. Eulers formulas may also help.
- 5.28. The set $\{1, \cos(x), \sin(x)\}$ is an orthogonal set of vectors according to the last problem, so project f onto each of these vectors.
- 5.29. Applying Gram-Schmidt to the columns of A to get an ON-basis of \mathbb{C}^2 , put these vectors as columns in Q and find $R = Q^*AQ$.
- 5.30. Apply Gram-Schmidt to the columns.
- 5.31. Use Gram-Schmidt to find orthogonal vectors spanning the same set as the columns. To obtain a unitary square matrix Q , extend this basis to an ON-basis.
- 5.32. To show homogeneity $F(\lambda w) = \lambda F(w)$ we can calculate $\langle v, F^*(\lambda w) \rangle = \langle F(v), \lambda w \rangle = \bar{\lambda} \langle F(v), w \rangle = \bar{\lambda} \langle v, F^*(w) \rangle = \langle v, \lambda F^*(w) \rangle$, so $F^*(\lambda w) = \lambda F^*(w)$. Additivity can be proved similarly.
- 5.33. For the first part, $\langle u, (F^* \circ G^*)(v) \rangle = \langle u, F^*(G^*(v)) \rangle = \langle F(u), G^*(v) \rangle = \cdots$
- 5.34. Let $\{v_i\}$ and $\{w_j\}$ be the ON-bases, then $\langle v_i, F^*(w_j) \rangle = \langle F(v_i), w_j \rangle$

5.35. Since the matrices are given with respect to an ON-basis: Self-adjoint maps correspond to symmetric matrices, unitary maps corresponds to matrices whose columns form an ON-basis, and normal maps have matrices satisfying $NN^* = N^*N$.

5.36. Try a scaled rotation-matrix.

5.37. Self adjoint: $A^* = A$, Unitary: $AA^* = I$, Normal: $AA^* = A^*A$.

5.38. Show that if λ is an eigenvalue, then $|\lambda| = 1$.

5.39. $\langle F(u), F(v) \rangle = \langle u, F^*(F(v)) \rangle \dots$

5.40. This follows easily from the definitions

5.41. For (a), the key step is $F^*(w) = 0 \leftrightarrow \langle v, F^*(w) \rangle = 0 \forall v \in V$.

5.42.

5.43.

a) Show that $(A + \lambda I)(A + \lambda I)^* = (A + \lambda I)^*(A + \lambda I)$ by expanding both sides.

b) It suffices to show that $\|A^*v\|^2 = \langle A^*v, A^*v \rangle = \langle Av, Av \rangle = \|Av\|^2$.

c) Combine parts (a) and (b): $Av = \lambda v \Leftrightarrow v \in \ker(A - \lambda I) \Leftrightarrow v \in \ker((A - \lambda I)^*) = \ker(A^* - \bar{\lambda}I) \Leftrightarrow A^*v = \bar{\lambda}v$.

d) Assume $Au = \lambda u$ and $Av = \mu v$. Expand both sides of the equality $\langle Au, v \rangle = \langle u, A^*v \rangle$ using part (c) and subtract.

5.44.

a) Finding one counterexample suffices.

b) Write $A = \frac{A+A^*}{2} + i\frac{i(A^*-A)}{2} = B + Ci$, and show that $X^*AX > 0$ for all nonzero complex $X \in \mathbb{C}^n$ implies that $C = 0$.

c) Try an upper triangular 2×2 -matrix.

d) In (b) we allow X to have complex entries...

5.45. A must be Hermitian, then use Sylvesters criterion.

5.46. Show that all principal minors are positive if and only if all the diagonal elements of D are positive.

5.47. If the $k \times k$ principal minor is negative, consider columns X on block form $\begin{pmatrix} X' \\ 0 \end{pmatrix}$ where X' is of size $k \times 1$.

8 Answers

1.1. $v + v = 1 \cdot v + 1 \cdot v = (1 + 1) \cdot v = 2 \cdot v$. The condition that the vector space is complex is not really necessary, the only problematic step is the last one: $1 + 1 = 2$, as 2 may not be an element of the field.

1.2. The zero element is 1, the additive inverse of 5 is $\frac{1}{5}$

1.3.

- a) S_1 is not additive and not homogeneous, it is not a subspace. Geometrically it is a line not passing the origin.
- b) S_2 is both additive and homogeneous, it is a subspace. Geometrically it is a line through the origin.
- c) S_3 is not additive but it is homogeneous, it is not a subspace. Geometrically S_3 is the union of the coordinate axes.
- d) S_4 is additive but not homogeneous, it is not a subspace. Geometrically S_4 is the first quadrant.
- e) S_5 is additive but not homogeneous, it is not a subspace. Geometrically S_5 is a lattice.
- f) S_6 is additive and homogeneous, it is a subspace. By definition every vector space is a subspace of itself, although it is called a *non-proper* subspace.
- g) S_7 is technically both additive and homogeneous since a statement of form $\forall x \in M : P(x)$ is true whenever M is empty. However, it is not a subspace - by definition subspaces are required to be nonempty. Note however that the point set $\{0\}$ is a subspace.

1.4. The projection is $u = (-1, 2, -1)$.

1.5. One choice of basis is $((x - 2), x(x - 2), x^2(x - 2))$.

1.6. $S \cap S'$ and $S + S'$ are both subspaces. $S \cup S'$ is in general not a subspace, as illustrated by S_3 of the previous exercise.

1.7. For example, consider the three lines spanned by $(1, 0)$, $(0, 1)$, and $(1, 1)$.

1.8.

- a)
- b)
- c) $\cosh(x) = \frac{e^x + e^{-x}}{2}$
- d) 0 (the function is clearly even)

1.9. I, G, H, T, C are linear.

1.10. The kernel consists of all symmetric matrices, the image of all skew-symmetric matrices.

1.11. If $\ker(F)$ contains some nonzero v , then $F(0) = 0 = F(v)$, so F is not injective. On the other hand, if F is not injective with $F(u) = F(v)$ for two different vectors u and v , then by linearity $F(u - v) = F(u) - F(v) = 0$, so $u - v$ is a nonzero vector in the kernel.

1.12. F is injective but not surjective since the first coordinate is always zero in the image. F doesn't have an inverse (the function G that left-shifts sequences is a right inverse but not a left-inverse to G)

$$\mathbf{1.13.} [F] = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. F^{-1}(p(x)) = p(x - 1).$$

1.14. $[F] = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}$ and $[G] = \begin{pmatrix} 0 & -2 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

1.15.

1.16. $A = C$ since $(4, 5) - (2, 3) \in U$. Graphically, A , B , and $A + B$ are lines in \mathbb{R}^2 intersecting the y -axis in 1, -1 , and 0 respectively. In the basis (A) , B has coordinates (-1) , since $(-1)A = -1((2, 3) + U) = (-2, -3) + U = (0, 1) + U = B$.

1.17. $(\frac{1}{2}, -\frac{1}{2})$.

1.18. Since $3e_1 - 1e_2 + 2(0, 0, 1, 1) = 0 + \ell$ the vectors are linearly dependent and is not a basis.

1.19. Consider the choice of bases from the hint. If the matrix for \tilde{F} is A , then the matrix for F has block form $\left(\begin{array}{c|c} 0 & B \\ \hline 0 & A \end{array} \right)$ where B is some matrix.

1.20.

- a) The line $(1, 2) + t(1, 1)$ (so $v = (1, 2)$, $U = \text{span}(1, 1)$.)
- b) The plane $x + y + z = 1$ (so $v = (1, 0, 0)$, $U : x + y + z = 0$.)

1.21. For (b): The map is given by $F(v) = G(v) + w$ with $w = (1, -1)$ fixed and $G(x, y) = (y, x)$ linear.

1.22.

- a) $(1, 0, 2, 0)$
- b) $X = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$
- c) $2 \cdot 2 - 0 \cdot 1 = 1$
- d) $\begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix}$
- e) $3^4 = 81$

1.23.

- a) The zero vector is the empty set \emptyset since $S + \emptyset = S \Delta \emptyset = S$ for all subsets S . It follows from the axioms that $0 \cdot S = \emptyset$, and $1 \cdot S = S$ must hold, this defines scalar multiplication over \mathbb{Z}_2 completely.
- b) $\{1, 3, 5\} + \{1, 2, 3\} = \{2, 5\}$
- c) Compute $-\{1, 3, 5\} = \{1, 3, 5\}$
- d) The natural choice of basis are the singleton sets $(\{1\}, \{2\}, \{3\}, \{4\}, \{5\})$, any set is a (the natural linear combination of these).
- e) Yes, since $S_1 + S_2 = S_4$ we have $1 \cdot S_1 + 1 \cdot S_2 + 1 \cdot S_4 = \emptyset$

1.24.

2.1. $(A + B)^* = A^* + B^* = A + B$ so $A + B$ is always Hermitian when A and B are.

$(AB)^* = B^* A^* = BA$ so AB is only Hermitian when A and B commute. For a counterexample, take $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

$(\lambda A)^* = \bar{\lambda} A^* = \bar{\lambda} A$ so λA is only Hermitian when $\lambda \in \mathbb{R}$. For a counterexample, take $\lambda = i$ and $A = I$.

$(A^T)^* = (\overline{A^T})^T = (A^*)^T = A^T$ so A^T is always Hermitian when A is.

$(AB^* + BA^*)^* = B^{**}A^* + A^{**}B^* = BA^* + AB^* = AB^* + BA^*$, so $AB^* + BA^*$ is Hermitian when A and B are.

2.2. $(A + Bi)^* = A^* + \bar{i}B^* = \overline{A}^T - i\overline{B}^T = A^T - iB^T = A + Bi \Leftrightarrow A = A^T$ and $B = -B^T$.

2.3. After using the formula and taking the sum of the diagonals in both AB and BA , the trace of both is equal to $\sum_{i,k} a_{ik}b_{ki}$.

2.4.

a)

b) $\lambda_1^n + \lambda_2^n + \lambda_3^n$

c) $-1, 1, 2$

2.5. $P^3 = I$, which shows that

$$P^n = \begin{cases} I & \text{for } n = 3k \\ P & \text{for } n = 3k + 1 \\ P^2 & \text{for } n = 3k + 2 \end{cases}$$

where $k \in \mathbb{Z}$. Note that since $P^{-1} = P^2$ the formula also holds for negative n .

2.6. For the last part: take for example the matrix

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

and consider how it acts on basis vectors: $e_1 \rightarrow e_2 \rightarrow e_3 \rightarrow e_1$ and $e_4 \leftrightarrow e_5$. The first cycle becomes the identity in P^n when n is divisible by 3, and the second cycle when n is divisible by 2. The order of P is therefore 6, the least common multiple of 3 and 2.

2.7. $\mathcal{C}_A = \text{span}(A, I)$.

2.8. $\mathfrak{C} = \text{span}(I)$, only multiples of the identity-matrix commute with everything.

2.9. With E as in the hint, all the matrices of form $S^{-1}ES$ are also idempotents.

2.10. $\dim \mathcal{C}_A \in \{5, 7, 9, 11, 13, 17, 25\}$

2.11. For the last statement, neither MN nor $M + N$ need be nilpotent if $AB \neq BA$, take for example

$M = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $N = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. Then M and N are nilpotent, but $M + N$ and MN are not.

2.12. $F(F(v)) = 0$, since by definition, the inner argument $F(v)$ lies in the image, and therefore in the kernel, and thus is mapped to zero. So $F^2 = 0$ and F is nilpotent with nilpotency degree 2.

2.13. (If $0 \neq \ker(N^k) = \ker(N^{k+1})$ for some n , then N would act bijectively on $\ker(N^k)$, so $\ker(N^{k'}) = \ker(N^k)$ for all $k' > k$, which contradicts nilpotency).

2.14.

2.15.

2.16. $M^{-1} = \left(\begin{array}{c|c} A^{-1} & -A^{-1}BC^{-1} \\ \hline 0 & C^{-1} \end{array} \right)$

2.17.

$$\text{REF} = \begin{pmatrix} 1 & -1 & 1 & 2 & 1 \\ 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{RREF} = \begin{pmatrix} 1 & -1 & 0 & \frac{5}{2} & \frac{3}{2} \\ 0 & 0 & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Other answers are possible for the REF, but not for the RREF.

$$\ker(A) = \{(s - 5t - 3r, s, t + r, 2t, 2r) \mid s, t, r \in \mathbb{R}\} = \text{span}((1, 1, 0, 0, 0), (-5, 0, 1, 2, 0), (-3, 0, 1, 0, 2))$$

2.18. The RREF of C is $\begin{pmatrix} 1 & 2i & 0 & 1 \\ 0 & 0 & 1 & i \end{pmatrix}$. The solutions to $CX = 0$ are

$$(x_1, x_2, x_3, x_4) = (-2is, s, -it, t) \text{ where } s, t \in \mathbb{C}.$$

2.19. The RREF of A is $\begin{pmatrix} 1 & 2 & -3 & 0 & 1 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

2.20. $((0, -1, 1, 0, -1), (1, 1, 0, 0, 1), (2, -1, 4, 0, -2))$ is one choice of basis for $\text{Im}(A)$

$((1, 0, 1, 0, 0), (1, 1, 0, 0, 1))$ is one choice of basis for $\ker(A)$

$((1, 0, 1, 0, 0), (1, 1, 0, 0, 1))$ is one choice of basis for $\text{Im}(A) \cap \ker(A)$ (so in this case $\ker(A) \subset \text{Im}(A)$)

2.21. The suggestion from the hint gives: $\text{Im}(A) = \text{span}(e_2, e_3)$, while $\ker(A) = \text{span}(e_3, e_4)$, and their intersection is $\text{span}(e_3)$.

2.22.

$$E_1^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad E_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{5} \end{pmatrix} \quad E_3^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In general: $(I + \lambda e_{ij})^{-1} = I - \lambda e_{ij}$, $\text{diag}(d_1, d_2, \dots, d_n) = \text{diag}(\frac{1}{d_1}, \frac{1}{d_2}, \dots, \frac{1}{d_n})$, and $P^{-1} = P$ for matrices corresponding to switching two rows of the identity matrix ($P = I - e_{ii} - e_{jj} + e_{ij} + e_{ji} \Rightarrow P^{-1} = P$)

2.23.

$$(a) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} A \quad (b) A \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (c) \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} A$$

2.24.

$$A = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} \quad A^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}$$

(other answers are possible, multiply to verify)

2.25. If $EA = E_n \cdots E_2 E_1 A = I$ is a product of matrices that reduces A to the identity, this shows that $A = E^{-1}$ and $A = E$. When multiplying the block matrix on the left by E we get $E[A|I] = [EA|EI] = [I|E] = [E|A^{-1}]$.

2.26.

$$LU = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 3 & 4 \end{pmatrix} \quad LDU = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & \frac{4}{3} \end{pmatrix}$$

2.27.

$$LU = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & -6 & -3 \end{pmatrix} \quad LDU = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -6 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 & 1 \\ 0 & 1 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{2} \end{pmatrix}$$

2.28. $PA = LU$ where $L = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$, $P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, $U = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$.

2.29. Let E be a product of elementary matrices that reduces A to row echelon form U . Then $Av = 0 \Rightarrow EAv = 0 \Leftrightarrow Uv = 0$.

2.30. Such a matrix has block form $\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$, a diagonal matrix where the diagonal has a number of ones followed by a number of zeros.

2.31.

a) $(3 \cdot 4 + 2 \cdot 3 + 1 \cdot 2) + (3 + 2 + 1) = 26$. The first parenthesis corresponds to the operations to put A in echelon form, the rest corresponds to the back-substitution.

b) $2(3 + 2 + 1) = 12$

c) $2(1 + 2 + \dots + (n-1)) = n(n-1)$ (for the standard Gaussian elimination when A is $n \times n$, the answer is a polynomial of degree 3).

2.32. With $C = \begin{pmatrix} 1 & 0 \\ 2 & \sqrt{2} \end{pmatrix}$ we have $A = CC^*$.

With $G = \begin{pmatrix} 1 & 0 \\ 2+i & 2 \end{pmatrix}$ we have $B = GG^*$.

2.33. With $C = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 2 & \sqrt{5} \end{pmatrix}$ we have $A = CC^*$.

2.34. $(CC^*)^* = (C^{**})(C^*) = CC^*$ so CC^* is Hermitian. The 1×1 -matrix (-1) is Hermitian but does clearly not admit a factorization $(-1) = (\lambda)(\lambda)^* = (|\lambda|^2)$.

2.35.

$$\text{RREF: } \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad X = s(1, 1, 0, 0) + t(2, 0, 2, 1) \quad s, t \in \mathbb{Z}_3.$$

Note that since $s, t \in \{0, 1, 2\}$ there are exactly nine solutions to the linear system.

3.1.

a) $\sigma(F) = \{0, 1\}$, $\dim E_1 = 1$, $\dim E_0 = 2$

b) $\sigma(F) = \{1, -1\}$, $\dim E_1 = 2$, $\dim E_{-1} = 1$

c) $\sigma(F) = \{1\}$, $\dim E_1 = 1$ (unless the rotation angle is a multiple of π)

d) $\sigma(F) = \{1\}$, $E_1 = \mathbb{R}^3$ so $\dim E_1 = 3$.

e) $\sigma(F) = \{0\}$, $\dim E_0 = 1$.

3.2. $D = \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}$ and $S = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$ works. Other choices may also work.

3.3. $\sigma(F) = \mathbb{N}$ since $F(x^n) = nx^n$ for all integers $n \geq 0$. The eigenvectors for eigenvalue k are nonzero multiples of x^k . The operator F is sometimes called the *degree-operator*.

3.4. Since A is real we have $\bar{A} = A$. And we can compute

$$A\bar{v} = \bar{A}\bar{v} = \overline{Av} = \overline{\lambda v} = \bar{\lambda}\bar{v}$$

which shows that \bar{v} is an eigenvector of A with eigenvalue $\bar{\lambda}$.

3.5. The previous problem shows that $2 - 3i$ is an eigenvalue with eigenvector $(1, 1 - i)$, so

$$A = SDS^{-1} = \begin{pmatrix} 1 & 1 \\ 1+i & 1-i \end{pmatrix} \begin{pmatrix} 2+3i & 0 \\ 0 & 2-3i \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1+i & -i \\ 1-i & i \end{pmatrix} = \begin{pmatrix} -1 & 3 \\ -6 & 5 \end{pmatrix}.$$

3.6. In the 2×2 case the eigenvalues are $1, -1, i, -i$ with corresponding eigenvectors

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & i \\ -i & -1 \end{pmatrix} \quad \begin{pmatrix} 1 & -i \\ i & -1 \end{pmatrix},$$

so the operator is diagonalizable. The same is true in the 3×3 -case, with the geometric multiplicities of the eigenvalues in this case being $g_1 = 3, g_{-1} = 2, g_i = 2, g_{-i} = 2$.

3.7. Let $\xi = e^{\frac{2i\pi}{3}} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$. Then with $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \xi & 0 \\ 0 & 0 & \xi^2 \end{pmatrix}$ and $S = \begin{pmatrix} 1 & \xi^2 & 1 \\ 1 & \xi & \xi \\ 1 & 1 & \xi^2 \end{pmatrix}$ we have $SDS^{-1} =$

$$[P] = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

3.8. Every nonzero vector in the hyperplane $x + 2y + 3z + 4w = 0$ is an eigenvector with eigenvalue 0. Every vector parallel to $(1, 2, 3, 4)$ is an eigenvector of eigenvalue 30.

3.9. With notation as in the hint: the eigenspaces are $E_{1+i} = \text{span}(1+i, 1)$ and $E_{1-i} = \text{span}(1-i, 1)$, and $X_0 = (2-i) \begin{pmatrix} 1+i \\ 1 \end{pmatrix} + (2+i) \begin{pmatrix} 1-i \\ 1 \end{pmatrix}$, so

$$X_n = A^n X_0 = (2-i)(1+i)^n \begin{pmatrix} 1+i \\ 1 \end{pmatrix} + (2+i)(1-i)^n \begin{pmatrix} 1-i \\ 1 \end{pmatrix}.$$

Here the bottom coordinate is a_n , so

$$a_n = (2-i)(1+i)^n + (2+i)(1-i)^n \text{ for all } n \geq 0.$$

Note that despite how the expression looks, a_n is a real integer for each $n \in \mathbb{N}$.

3.10. $p(A) = A^4 + 2A^2 - 5A + 3I = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} + 2 \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} - 5 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + 3 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}.$

Since $B^2 = 0$ we get $p(B) = B^4 + 2B^2 - 5B + 3I = -5B + 3I = \begin{pmatrix} 3 & 0 & -5 \\ 0 & 3 & -5 \\ 0 & 0 & 3 \end{pmatrix}.$

3.11. We have $A^2 - 3A - 4I = 0$ (zero matrix), so $p(t) = t^2 - 3t - 4$ does the job. Note that $p(t)$ is in fact the characteristic polynomial of A .

3.12.

3.13. $p_A(A) = -(A + 2I)^2(A - 3I) = - \begin{pmatrix} 0 & 5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{pmatrix}^2 \begin{pmatrix} -5 & 5 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 0$ (easy when computed as a product of block-matrices)

3.14. $p(A) = 0, q(A) = 0$

3.15. The eigenvalues of A are $1+i$ and 2 , so the eigenvalues for B are $2^2 + 3 \cdot 2 - 5 = 5$ and $(1+i)^2 + 3(1+i) - 5 = 5i - 2$.

3.16. Since $p(t) = -t^3(t+1)^2$, we have

$$m(t) \in \{t^3(t+1)^2, t^2(t+1)^2, t(t+1)^2, t^3(t+1), t^2(t+1), t(t+1)\}$$

3.17. $3A + 2I$

3.18. a) $m_A(t) = (t-3)^2$ b) $m_B(t) = t(t-4)(t-7)(t-9)$ c) $m_C(t) = (t-1)(t-2)^2$

3.19. R^4 is the identity map on $\text{Mat}_n(\mathbb{C})$, so $m(t) = t^4 - 1$ annihilates R . It is clear that no lower degree polynomial can annihilate R (unless $n = 1$), so $m_R(t) = t^4 - 1$.

4.1. (b) and (c) are in Jordan form

4.2. With a_λ and g_λ as algebraic and geometric multiplicity of λ , and with $p(t)$ and $m(t)$ as characteristic and minimal polynomials:

a) $(a_2, g_2) = (2, 1)$, $(a_{-3}, g_{-3}) = (1, 1)$, $p(t) = (t-2)^2(t+3) = m(t)$ b) $(a_0, g_0) = (4, 2)$, $p(t) = t^4$, $m(t) = t^2$ c) $(a_5, g_5) = (3, 1)$, $(a_3, g_3) = (2, 1)$, $p(t) = (t-5)^3(t-3)^2 = m(t)$

4.3. $7 \cdot 3 = 21$

4.4. There are 11 different such Jordan forms. They corresponds to the 11 partitions of 6:

$6, 5+1, 4+2, 4+1+1, 3+3, 3+2+1, 3+1+1+1, 2+2+2, 2+2+1+1, 2+1+1+1+1, 1+1+1+1+1+1$

4.5. For example:

$$\text{a) } A = \begin{pmatrix} 3 & 1 & & \\ & 3 & & \\ & & 3 & \\ & & & 3 \end{pmatrix}, B = \begin{pmatrix} 3 & 1 & & \\ & 3 & & \\ & & 3 & 1 \\ & & & 3 \end{pmatrix}$$

$$\text{b) } A = \begin{pmatrix} 5 & 1 & & & \\ & 5 & 1 & & \\ & & 5 & & \\ & & & 5 & 1 \\ & & & & 5 \end{pmatrix}, B = \begin{pmatrix} 5 & 1 & & & \\ & 5 & 1 & & \\ & & 5 & & \\ & & & 5 & 1 \\ & & & & 5 \end{pmatrix}$$

4.6.

4.7.

4.8.

$$\mathcal{C}_J = \text{span} \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right) = \left\{ \begin{pmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{pmatrix} \mid a, b, c \in \mathbb{C} \right\}$$

In general, let N be the $n \times n$ -Jordan-block with eigenvalue 0. Then the commutant of any $n \times n$ -Jordan block J is $\mathcal{C}_J = \text{span}(I, N, N^2, \dots, N^{n-1})$

4.9. $J_4(0) \oplus J_2(0) \oplus J_1(0)$ is the only possibility (up to block-permutation).

4.10. a) $\dim \ker(F) = 5$ b) $\dim \text{Im}(F) = 9$ c) $\dim \ker(F^3) = 12$ d) $\dim \text{Im}(F^2) = 5$ e) $\dim \ker(F) \cap \text{Im}(F^2) = 3$

4.11. $p_F(t) = t^6$ and $m_F(t) = t^3$. Let $e'_4 = e_4 - e_1$, $e'_5 = e_5 - e_1$, and $e'_6 = e_6 - e_2$. Then $e_1, e_2, e_3, e'_4, e'_5, e'_6$ is a string basis for \mathbb{R}^6 :

$$e_1 \mapsto e_2 \mapsto e_3 \mapsto 0, \quad e'_4 \mapsto 0, \quad e'_5 \mapsto e'_6 \mapsto 0.$$

4.12. For example, with: $S = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ and $J = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ we have $SJS^{-1} = N$.

4.13. For example, with: $S = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 0 & -1 & 0 \end{pmatrix}$ and $J = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ we have $SJS^{-1} = M$.

4.14. For example, $S = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix}$ and $J = \begin{pmatrix} -1 & & \\ & 2 & \\ & & 2 \end{pmatrix}$.

4.15. For example, with

$$S = \begin{pmatrix} -1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ -2 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \text{and} \quad J = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

we have $SJS^{-1} = A$. Other S are possible, but J is unique up to switching the two blocks.

4.16. For example, with S and J as below we have $A = SJS^{-1}$:

$$S = \begin{pmatrix} 1 & 2 & 0 & 4 & 1 \\ 1 & 3 & 0 & 8 & 0 \\ 0 & 0 & 3 & -2 & 1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -2 & 1 & 0 \end{pmatrix} \quad J = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

Other options for S is possible, but not for J (except switching the two Jordan blocks).

4.17. For example, with $J = \begin{pmatrix} 3+i & 1 \\ 0 & 3+i \end{pmatrix}$ and $S = \begin{pmatrix} 3 & -1 \\ -2 & 1 \end{pmatrix}$ we have $A = SJS^{-1}$.

4.18. For example, $J = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$ and $S = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ gives $A = SJS^{-1} = \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix}$.

4.19. $[F] = J_4(1)$. $(F - \text{id})(p(x)) = p(x+1) - p(x)$ clearly decreases the degree of a polynomial by exactly 1, so x^3 is a first vector in a Jordan chain of length 4 for the eigenvalue 1, this determines the Jordan form.

4.20. The determinant is the product of diagonal in the Jordan-matrix, these are the eigenvalues of A , so $\det(A) = 0 \Leftrightarrow 0 \in \sigma(A)$.

4.21. With notation as in the hint, $A = S_1JS_1^{-1} = S_1(S_2^{-1}BS_2)S_1^{-1} = (S_1S_2^{-1})B(S_1S_2^{-1})^{-1}$, so $A = TBT^{-1}$ for $T = S_1S_2^{-1}$.

4.22. $J = J_3(1) \oplus J_1(i) \oplus J_1(-i)$. The functions $\cos(x) + i\sin(x) = e^{ix}$ and $\cos(x) - i\sin(x) = e^{-ix}$ are eigenvectors for the eigenvalues $\pm i$, and x^2e^x is the first vector of a Jordan chain of length 3 for the eigenvalue 1.

4.23. (A polynomial with simple roots would not be able to annihilate a Jordan block of size > 1 .)

4.24.

4.25.

$$A^2 \sim \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}, \quad \text{for } n \geq 3 \text{ we have } A^n \sim \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2^n & 1 \\ 0 & 0 & 0 & 0 & 2^n \end{pmatrix}$$

4.26. For example, with $J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $S = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ we have $A = SJS^{-1}$.

4.27. $x^2 + x + 1$ is the only irreducible polynomial of degree 2 with coefficients in \mathbb{Z}_2 (note that $x^2 + 1 = (x+1)(x+1)$). So take one of the matrices $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ or $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, these are the only matrices in $\text{Mat}_2(\mathbb{Z}_2)$ with the $p_A(t) = t^2 + t + 1$.

As a remark, if we extend our field of scalars from \mathbb{Z}_2 to the field $\mathbb{F} = \mathbb{Z}_2[x]/\mathbb{Z}_2[x](x^2 + x + 1) = \{0, 1, x, x+1\}$ (a field of four elements), the two matrices above actually do admit a Jordan-decomposition $A = SJS^{-1}$ where $S, J \in \text{Mat}_2(\mathbb{F})$. But constructions like this lie beyond the scope of this course.

4.28.

$$A^n = \begin{pmatrix} 2^n & n2^{n-1} & \frac{n(n-1)}{2}2^{n-2} & 0 & 0 \\ 0 & 2^n & n2^{n-1} & 0 & 0 \\ 0 & 0 & 2^n & 0 & 0 \\ 0 & 0 & 0 & (-6)^n & n(-6)^{n-1} \\ 0 & 0 & 0 & 0 & (-6)^n \end{pmatrix} = 2^{n-3} \begin{pmatrix} 8 & 4n & n(n-1) & 0 & 0 \\ 0 & 8 & 4n & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 8(-3)^n & 4n(-3)^{n-1} \\ 0 & 0 & 0 & 0 & 8(-3)^n \end{pmatrix}$$

4.29.

$$e^A = \begin{pmatrix} e^{-3} & 0 & 0 \\ 0 & \sqrt{e} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad e^B = I + B + \frac{B^2}{2} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & \frac{9}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

$$4.30. \quad e^J = \begin{pmatrix} e^2 & e^2 & 0 \\ 0 & e^2 & 0 \\ 0 & 0 & e^3 \end{pmatrix} = e^2 \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e \end{pmatrix}.$$

4.31. $e^A v = e^\lambda v$ so v is still an eigenvector but with eigenvalue e^λ .

4.32.

$$\begin{aligned} \sin(A_1) &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} & \sin(A_2) &= \begin{pmatrix} \sin(2) & \cos(2) & -\frac{\sin(2)}{2} & -\frac{\cos(2)}{6} \\ 0 & \sin(2) & \cos(2) & -\frac{\sin(2)}{2} \\ 0 & 0 & \sin(2) & \cos(2) \\ 0 & 0 & 0 & \sin(2) \end{pmatrix} & \sin(A_3) &= \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ \cos(A_1) &= \begin{pmatrix} 1 & 0 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & \cos(A_2) &= \begin{pmatrix} \cos(2) & -\sin(2) & -\frac{\cos(2)}{2} & \frac{\sin(2)}{6} \\ 0 & \cos(2) & -\sin(2) & -\frac{\cos(2)}{2} \\ 0 & 0 & \cos(2) & -\sin(2) \\ 0 & 0 & 0 & \cos(2) \end{pmatrix} & \cos(A_3) &= \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \end{aligned}$$

4.33. The linear system of differential equations of order two $X''(t) + A^2 X(t) = 0$ has the general solution $\sin(At)C + \cos(At)D$.4.34. If A has Jordan form with eigenvalues λ_i on the diagonal, then $\det(e^A) = \prod e^{\lambda_i} = e^{\sum \lambda_i} = e^{\text{tr} A} = e^0 = 1$.

$$4.35. \quad \begin{cases} a_n = (-3)^{n-1}(2n-6) \\ b_n = (-3)^{n-1}(-2n) \end{cases} \quad \text{so } \frac{a_n}{b_n} \rightarrow -1.$$

4.36. With $v_1 = (1, 1, 1)$, $v_2 = (0, 1, 2)$, and $v_3 = (0, 0, 1)$ we have $(A - 2I)v_1 = 0$, $(A - 2I)v_2 = v_1$, and $(A + I)v_3 = 0$, so (v_1, v_2, v_3) is a Jordan basis:

$$\text{With } S = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \text{ and } J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix} \text{ we have } A = SJS^{-1}.$$

4.37. With $X_0 = (0, 1, 0)^T$, and with $A = D + J$ from in the previous problem, the solution is $e^{tA}X_0 = S e^{tD} e^{tN} S^{-1}$

$$\begin{aligned} &= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} e^{2t} & 0 & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{-t} \end{pmatrix} \begin{pmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} te^{2t} \\ (t+1)e^{2t} \\ (t+2)e^{2t} - 2e^{-t} \end{pmatrix} \end{aligned}$$

5.1.

a) $3 + 2i$

b) $\sqrt{11}$

c) $t(i - 1, 1)$ where $t \in \mathbb{C}$

$$5.2. \quad \|x^2 + x + 1\|^2 = \int_0^1 (x^2 + x + 1)^2 dx = \int_0^1 x^4 + 2x^3 + 3x^2 + 2x + 1 dx = \left[\frac{x^5}{5} + \frac{2x^4}{4} + \frac{3x^3}{3} + \frac{2x^2}{2} + x \right]_0^1 = \frac{1}{5} + \frac{1}{2} + 1 + 1 + 1 = \frac{37}{10}, \text{ so the sought length is } \sqrt{\frac{37}{10}}.$$

5.3.

5.4. Only the first rule gives an inner product.

5.5. $\|v\|_2 = \sqrt{\langle v, v \rangle} = \sqrt{2\langle v, v \rangle} = \sqrt{2}\|v\|_1$, so in the new norm vectors are a factor of $\sqrt{2}$ longer. Similarly, in the new norm $\cos(\theta) = \frac{\langle u, v \rangle}{\|u\|_2 \|v\|_2} = \frac{2\langle u, v \rangle}{\sqrt{2}\|u\|_1 \sqrt{2}\|v\|_1} = \frac{\langle u, v \rangle}{\|u\|_1 \|v\|_1}$ so angles between vectors are the same with the two inner products.

5.6. For the first question the answer is yes: Since (u, v) is a basis for \mathbb{C}^2 we can define an inner product by declaring that $\langle u, v \rangle = 0$, $\langle u, u \rangle = 0$, and $\langle v, v \rangle = 1$.

For the second question the answer is no: Since $v' = (1 - i)u'$ we would then have $0 = \langle v', u' \rangle = \langle (1 + i)u', u' \rangle = (1 + i)\langle u', u' \rangle \Leftrightarrow \langle u', u' \rangle = 0$, but $u' \neq 0$ so this contradicts the inner product axioms.

5.7.

5.8.

5.9. For example, take $u = (1, 0)$ and $v = (0, 1)$. Then the parallelogram law for the maximum norm becomes $2 = 1 + 1 = \|(1, 1)\|^2 + \|(1, -1)\|^2 = 2\|(1, 0)\|^2 + 2\|(0, 1)\|^2 = 2 + 2 = 4$ which does not hold.

5.10. No, \mathcal{I} is closed under addition but not scalar multiplication: $(-1) \cdot \langle v, v \rangle \leq 0$ so it is not positive definite.

5.11.

a) $\|(2, 1, -4)\|_2 = \sqrt{2^2 + 1^2 + (-4)^2} = \sqrt{21}$

b) $\|(2, 1, -4)\|_{\max} = \max\{|2|, |1|, |-4|\} = 4$

c) $\|(2, 1, -4)\|_{\text{Mh}} = |2| + |1| + |-4| = 7$

d) $\|(2, 1, -4)\|_3 = (|2|^3 + |1|^3 + |-4|^3)^{\frac{1}{3}} = (73)^{\frac{1}{3}}$

5.12. $\|A\|_F = \sqrt{1^2 + (-1)^2 + 1^2 + 1^2} = 2$ and $\|A\|_{\sigma} = |1 \pm i| = \sqrt{2}$

5.13.

5.14. $p = \frac{4}{3}$ (In general, $(\cos^{\frac{2}{p}}(t), \sin^{\frac{2}{p}}(t))$ is a parametrization of the unit circle with respect to the p -norm)

5.15. Let e_1, e_2 be two standard basis vectors in \mathbb{C}^n . For $0 < p < 1$ we have $\|e_1 + e_2\|_p = (1^p + 1^p)^{\frac{1}{p}} = 2^{\frac{1}{p}} > 2 = 1 + 1 = \|e_1\|_p + \|e_2\|_p$.

5.16. $1 \cdot \|v\|_{\max} \leq \|v\| \leq \sqrt{2} \cdot \|v\|_{\max}$. Remark: On a finite dimensional vector space, all norms are in fact equivalent.

5.17.

5.18. No, for example: $(1, i, 0) \times (1, 1, 1)$ is not orthogonal to $(1, i, 0)$. In fact it is not possible to define a vector product on \mathbb{C}^3 with the same properties as the vector product on \mathbb{R}^3 .

5.19. The vectors of length zero corresponds to the surface $z^2 = x^2 + y^2$, a double cone in \mathbb{R}^3 (the "light-cone" in two-dimensional space time in physics).

5.20.

5.21.

5.22.

5.23.

5.24. By the Pythagorean theorem we get $\|v - u\|^2 = \|v - P_U(v)\|^2 + \|P_U(v) - u\|^2 \geq \|v - P_U(v)\|^2$, so the minimal distance between u and v is $\|v - P_U(v)\|$, and it is attained when $v - P_U(v) = 0$.

5.25. Following the method in the hint we obtain

$$(f_1, f_2, f_3) = \left(\frac{1}{\sqrt{2}}(1, i, 0), \frac{1}{\sqrt{3}}(-1, i, 1), \frac{1}{\sqrt{6}}(1, -i, 2) \right)$$

which is an orthonormal basis for \mathbb{C}^3 where (f_1, f_2) is an ON-basis for U .

5.26. With the orthogonal basis $(e_1, e_2) = (1, x - \frac{1}{2})$ of \mathcal{P}_1 we get

$$g(x) = P_{\mathcal{P}_1}(e^x) = \frac{\langle e^x, 1 \rangle}{\langle 1, 1 \rangle} 1 + \frac{\langle e^x, x - \frac{1}{2} \rangle}{\langle x - \frac{1}{2}, x - \frac{1}{2} \rangle} (x - \frac{1}{2}) = 6(3 - e)x + 4e - 10.$$

5.27.

5.28. $g(x) = \frac{\pi}{2} - \frac{4}{\pi} \cos(x)$

5.29. With $Q = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $R = \sqrt{2} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ we have $A = QR$ satisfying all the conditions.

5.30.

$$Q = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \quad R = \begin{pmatrix} \frac{3}{\sqrt{3}} & \frac{6}{\sqrt{3}} & \frac{6}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ 0 & 0 & \frac{3}{\sqrt{6}} \end{pmatrix}.$$

5.31. $A = QR = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 2 & 6 & 3 \\ 0 & -2 & 1 \end{pmatrix}$
 $B = Q'R' = \begin{pmatrix} \frac{1}{\sqrt{14}} & \frac{4}{\sqrt{21}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{14}} & \frac{1}{\sqrt{21}} & \frac{-2}{\sqrt{6}} \\ \frac{3}{\sqrt{14}} & \frac{-2}{\sqrt{21}} & \frac{1}{\sqrt{6}} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{14} & \frac{10}{\sqrt{14}} \\ 0 & \frac{\sqrt{21}}{\sqrt{21}} \\ 0 & 0 \end{pmatrix}.$

5.32.

5.33.

5.34.

5.35. a) Self adjoint: G b) Unitary: Only H c) G and H

5.36. For example: The operator on \mathbb{C}^2 with matrix $\begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$ is normal but neither self-adjoint nor unitary.

5.37. For $a, b \in \mathbb{C}$ we get a) $a = \bar{b}$ b) $|a| = 4, b = -\bar{a}$ c) $|a| = |b|, a - b \in \mathbb{R}$ For $a, b \in \mathbb{R}$ this becomes
a) $a = b$ b) $(a, b) = \pm(4, -4)$ c) $a = \pm b$

5.38.

5.39.

5.40.

5.41. For (a): $w \in \ker(F^*) \Leftrightarrow F^*(w) = 0 \Leftrightarrow \langle v, F^*(w) \rangle = 0 \forall v \in V \Leftrightarrow \langle F(v), w \rangle = 0 \forall v \in V \Leftrightarrow w \perp \text{Im}(F) \Leftrightarrow w \in \text{Im}(F)^\perp$. For b, just replace F by F^* and take the complement of both sides in (a).

5.42.

5.43.

5.44.

a) For example, $-I$ is Hermitian but not positive definite.

b) With notation as in the hint we have $X^*AX = X^*BX + iX^*CX$. Since B and C are Hermitian, $X^*BX \in \mathbb{R}$ and $X^*CX \in \mathbb{R}$, so in order for X^*AX to be positive it needs to be real, and thus $0 = C = \frac{i(A^* - A)}{2}$, and $A = A^*$.

c) For example, with $A = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}$, when $X \in \text{Mat}_{2 \times 1}(\mathbb{R})$ is nonzero we get

$$X^*AX = X^TAX = (x_1 \ x_2) \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + 2x_1x_2 + 2x_2^2 = (x_1 + x_2)^2 + x_2^2 > 0.$$

d) Our matrix $A = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}$ from the last problem is in fact not positive definite, because there exists nonzero *complex* X such that $X^*AX \not> 0$. For example, $X = \begin{pmatrix} 1 \\ i \end{pmatrix}$ gives $X^*AX = 3 + 2i \not> 0$.

5.45. $a = 1 - i$, $b > 1$, $c > \frac{b}{2(b-1)}$

5.46.

5.47.