Introduction to Linear Algebra

Jānis Lazovskis*

Updated December 3, 2021

These notes were created to accompany the course *Introduction to Linear Algebra* for the BITL program, in the Fall 2021 semester, at RTU Riga Business School.

The text may contain mistakes - please send any you find to my email at the bottom of this page.

Contents

*RTU Riga Business School, janis.lazovskis@rbs.lv

This course will use Gilbert Strang's *Introduction to Linear Algebra*. You are encouraged to read the Preface to the textbook, available at `math.mit.edu/linearalgebra` before the first lecture.

Part I
Vector spaces

Lecture 1: Vectors and matrices

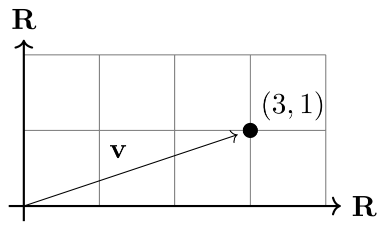
The first week will be a review of material you have seen before, but the setting may be broader, with different emphasis, and with different examples.

1.1 Vector review

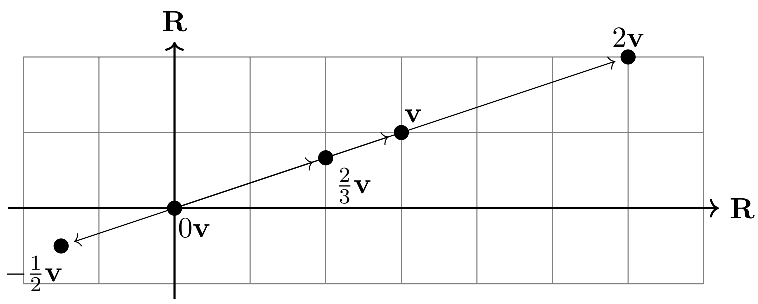
Definition 1.1. Let $n \in \mathbf{N}$. A *vector* in $\mathbf{R}^n$ is an ordered set of n elements.

The *zero vector*, or a *trivial vector*, denoted 0, is vector for which all elements are 0. Vectors that are not the zero vector are called *nontrivial*. A vector is usually thought of as a column of numbers, or a point in n-dimensional space, or the arrow to that point. All notions of a vector will be used interchangeably.

Example 1.2. The vector $\mathbf{v} = \left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$ in $\mathbf{R}^2$ can also be thought of as the arrow to $(3,1)$ or simply the point $(3,1)$ itself.



Multiplying the vector by elements of $\mathbf{R}$ we get other vectors "going in the same direction" as $\mathbf{v}$.



Vectors are combined together in *linear combinations*.

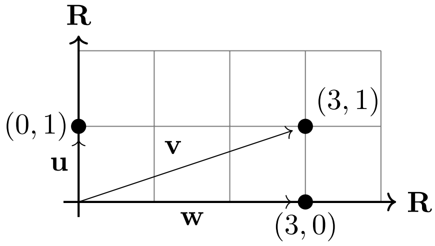
Definition 1.3. A *linear combination* of vectors is a vector $\mathbf{v} \in \mathbf{R}^n$ when it is expressed as a sum of other vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in \mathbf{R}^n$, and *scalars* $a_1, a_2, \dots, a_k \in \mathbf{R}$ multiplying them. That is,

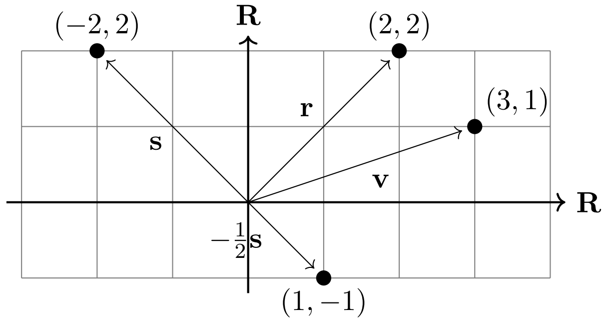
$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k.$$

For $k = 1$, a linear combination of vectors $v_1, \dots, v_k$ is called a *multiple* of the vector v_1.

Example 1.4. Every vector in the plane is a linear combination of (at most) two vectors, representing the x-direction and y-direction.

$$\mathbf{v} = \mathbf{w} + \mathbf{u} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3+0 \\ 0+1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \qquad \mathbf{v} = \mathbf{r} - \tfrac{1}{2}\mathbf{s} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - \tfrac{1}{2}\begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 - \frac{1}{2}\cdot(-2) \\ 2 - \frac{1}{2}\cdot 2 \end{bmatrix} = \begin{bmatrix} 2+1 \\ 2-1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$





The entries of vectors, and the numbers multiplying them, do not need to be numbers - they simply need to be elements of a *field*, a term which we will not define here.

Example 1.5. Some common examples of fields are $\mathbf{Q}, \mathbf{R}, \mathbf{C}$.

- The set $\mathbf{N}$ is not a field because although $1 \in \mathbf{N}$, there is no $x \in \mathbf{N}$ for which $1 + x = 1$ (the *additive identity* does not exist).
- The set $\mathbf{Z}$ is not a field because although $2 \in \mathbf{Z}$, there is no number $x \in \mathbf{Z}$ for which $2x = 1$ (*multiplicative inverses* do not exist).

Unless otherwise noted, we will use the field $\mathbf{R}$.

A key idea of vectors and their linear combinations that that they *fill a part of the space* in which they reside. The "part" of the space is another space itself.

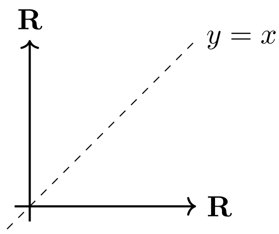
Example 1.6. Linear combinations can be described geometrically. For example:

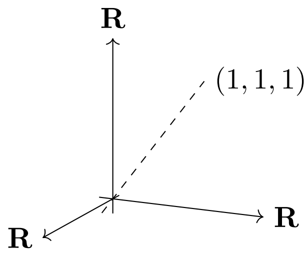
- Linear combinations of $(1, 1)$ and $(0, 0)$ form the line $y = x$ in the plane $\mathbf{R}^2$
- Multiples of $(1, 1, 1)$ form a line in 3-space $\mathbf{R}^3$
- Linear combinations of $(1, 1, 1)$ and $(1, 1, 0)$ form the plane $x - y = 0$ in $\mathbf{R}^3$
- Linear combinations of $(1, 1, 1)$, $(1, 1, 0)$, and $(0, 1, 1)$ fill all of $\mathbf{R}^3$. For example,

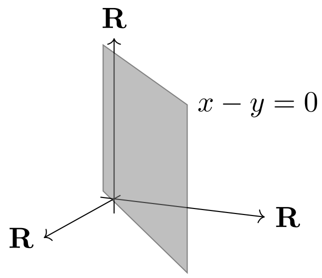
$$\begin{bmatrix} 7 \\ 9 \\ -5 \end{bmatrix} = -7\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 14\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

- Linear combinations of $(1, 1, 1)$, $(1, 1, 0)$, $(0, 1, 1)$, and $(1, 0, 1)$ still fill all of $\mathbf{R}^3$. For example,

$$\begin{bmatrix} 7 \\ 9 \\ -5 \end{bmatrix} = -7\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 14\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = -5\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 13\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$







Definition 1.7. The *dot product*, or *inner product* of two vectors $\mathbf{v} = (v_1, \dots, v_n), \mathbf{w} = (w_1, \dots, w_n) \in \mathbf{R}^n$, is the real number $\mathbf{v} \cdot \mathbf{w} := v_1 w_1 + \cdots + v_n w_n \in \mathbf{R}$.

In other words, the dot product is a function $\mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$. Working out the individual components, we see that $\mathbf{v} \cdot (\mathbf{u} + \mathbf{w}) = \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{w}$, a fact which you show in Exercise 1.2. The dot product of a vector $\mathbf{v}$ with itself is the square of the *norm*, or *length*, or *distance* of the vector $\mathbf{v}$. The norm is denoted $\|\mathbf{v}\|$, so we have

$$\|\mathbf{v}\| := \sqrt{\mathbf{v} \cdot \mathbf{v}}, \qquad \text{or} \qquad \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v} = v_1^2 + v_2^2 + \cdots + v_n^2.$$

We know the inside of the square root will be nonnegative, as we are summing squares. The norm satisfies the following properties, for any $\mathbf{v} \in \mathbf{R}^n$:

- Non-negative: $\|\mathbf{v}\| \geqslant 0$
- Positive definite: $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = 0$
- Multiplicative: $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$ for any $c \in \mathbf{R}$

These properties follow immediately from the properties of the real numbers and the definition of the norm above.

Definition 1.8. A vector $\mathbf{v} \in \mathbf{R}^n$ is a *unit vector* if $\|\mathbf{v}\| = 1$.

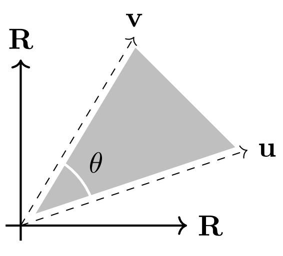
Proposition 1.9. For any $\mathbf{u}, \mathbf{v}$ nonzero in $\mathbf{R}^n$:

1. The vector $\frac{\mathbf{v}}{\|\mathbf{v}\|}$ is a unit vector.
2. The angle θ between $\mathbf{u}$ and $\mathbf{v}$ is computed by the relation $\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \cos(\theta)$
3. The *Cauchy–Schwarz inequality* holds: $|\mathbf{u} \cdot \mathbf{v}| \leqslant \|\mathbf{u}\|\|\mathbf{v}\|$
4. The *triangle inequality* holds: $\|\mathbf{u} + \mathbf{v}\| \leqslant \|\mathbf{u}\| + \|\mathbf{v}\|$

Proof. To prove 1., we need to show that the norm of $\frac{\mathbf{v}}{\|\mathbf{v}\|}$ is 1. This follows as

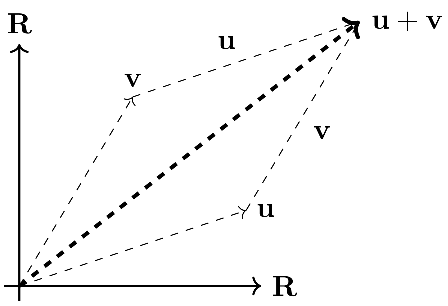
$$\left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\|^2 = \frac{\mathbf{v}}{\|\mathbf{v}\|} \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{1}{\|\mathbf{v}\|^2}(\mathbf{v} \cdot \mathbf{v}) = \frac{1}{\|\mathbf{v}\|^2}\|\mathbf{v}\|^2 = 1.$$

To prove 2., we use the law of cosines on the triangle formed by the origin 0, $\mathbf{u}$ and $\mathbf{v}$:



$$\begin{aligned}
\|\mathbf{u} - \mathbf{v}\|^2 &= \|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - 2\|\mathbf{v}\|\|\mathbf{u}\|\cos(\theta) \\
(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) &= \|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - 2\|\mathbf{v}\|\|\mathbf{u}\|\cos(\theta) \\
\mathbf{u} \cdot \mathbf{u} - 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} &= \|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - 2\|\mathbf{v}\|\|\mathbf{u}\|\cos(\theta) \\
\frac{-2\mathbf{u} \cdot \mathbf{v}}{-2\|\mathbf{v}\|\|\mathbf{u}\|} &= \cos(\theta)
\end{aligned}$$

To prove 3., use the fact that $\cos(\theta) \leqslant 1$, then take the absolute value of the equation from part 2. To prove 4., we can either draw a parallelogram and notice that the diagonal is $\mathbf{u} + \mathbf{v}$, and that it is shorter than the sum of the sides, which are $\mathbf{u}$ and $\mathbf{v}$. Or we can use algebra and part 3.



$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\
&= \mathbf{u} \cdot \mathbf{u} + 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \\
&\leqslant \|\mathbf{u}\|^2 + 2|\mathbf{u} \cdot \mathbf{v}| + \|\mathbf{v}\|^2 \\
&\leqslant \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2
\end{aligned}$$

□

As a result of part 2., if $\mathbf{u}$ is perpendicular to $\mathbf{v}$, then $\theta = \pi/2$, and so $\cos(\theta) = 0$. That is, $\mathbf{u}$ is perpendicular to $\mathbf{v}$ if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

Definition 1.10. Two non-zero vectors $\mathbf{v}$, $\mathbf{w}$ are *parallel* if there exists $c \in \mathbf{R}_{\neq 0}$ with $\mathbf{v} = c\mathbf{w}$. If $c = 1$, then the two vectors are *colinear*. In the opposite case, when the dot product $\mathbf{v}^T\mathbf{w} = 0$, the vectors are called *perpendicular*, or *orthogonal*.

Sometimes "parallel" is used when $c > 0$ and "anti-parallel" for $c < 0$. We will see orthogonality later in Lecture 7.

1.2 Matrix review

Definition 1.11. Let $m, n \in \mathbf{N}$. An $m \times n$ *matrix* over $\mathbf{R}$ is an ordered set of $m \cdot n$ elements. The space of all $m \times n$ matrices over $\mathbf{R}$ is denoted $\mathcal{M}_{m\times n}(\mathbf{R})$ or simply $\mathcal{M}_{m\times n}$, when the field is not relevant or clear from context.

Comparing Definition 1.11 with Definition 1.1, we see that a vector in $\mathbf{R}^n$ is just a $n \times 1$ (or $1 \times n$) matrix. Similarly to vectors, the elements of matrices may be over other fields, not necessarily $\mathbf{R}$. A two matrices of particular importance are the zero matrix 0, or $\mathbf{0}$ (all entries are zero) and the identity matrix I, or $\mathbf{1}$, (all entries are zero except the diagonal, which is all 1's), given by

$$0 := \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \qquad I := \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

These are square matrices. Sometimes to emphasize the size of the matrix, we write 0_n and I_n for matrices with n rows and n columns. For an $m \times n$ matrix A, the entry in row i and column j is denoted A_{ij} or $(A)_{ij}$ or $A(i, j)$ or a_{ij}. That is,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}.$$

Sometimes instead of giving specfic numbers, you are given specific submatrices. These are called *block matrices*. For example, if $A \in \mathcal{M}_{2\times 3}$, $B \in \mathcal{M}_{2\times 5}$, $C \in \mathcal{M}_{3\times 3}$, and $D \in \mathcal{M}_{3\times 5}$, then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathcal{M}_{5\times 8} \qquad \text{and} \qquad \begin{bmatrix} C & 0 \\ I & D \end{bmatrix} \in \mathcal{M}_{6\times 8}$$

are both block matrices. The identity I and zero 0 matrices are used without specifying their size as blocks in a block matrix. The matrix I will always be square, but 0 can be any shape. Finally, there are three special types of square matrices:

$$\begin{bmatrix} * & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & * \end{bmatrix} \qquad \begin{bmatrix} * & 0 & 0 & \cdots & 0 \\ * & * & 0 & \cdots & 0 \\ * & * & * & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \cdots & * \end{bmatrix} \qquad \begin{bmatrix} * & 0 & 0 & \cdots & 0 \\ 0 & * & 0 & \cdots & 0 \\ 0 & 0 & * & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & * \end{bmatrix}$$

upper triangular matrix $a_{ij} = 0$ if $i > j$ — *lower triangular matrix* $a_{ij} = 0$ if $i > j$ — *diagonal matrix* $a_{ij} = 0$ if $i \neq j$

The symbol "$*$" represents any number, and they do not all have to be the same. The two on the left are called *triangular matrices*. We will see several times over why these are special.

Definition 1.12. There are several common matrix operations.

- *sum*: the sum of $A \in \mathcal{M}_{m\times n}$ and $B \in \mathcal{M}_{m\times n}$ has ij-entry $(A+B)_{ij} = A_{ij} + B_{ij}$
- *product*: the product of $A \in \mathcal{M}_{m\times n}$ and $C \in \mathcal{M}_{n\times m}$ has ij-entry $(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}$
- *Hadamard product*: the Hadamard product, or entry-wise product, of $A \in \mathcal{M}_{m\times n}$ and $B \in \mathcal{M}_{m\times n}$ has ij-entry $(A \circ B)_{ij} = A_{ij}B_{ij}$

Remark 1.13. Matrix addition has the following properties:

- addition is *commutative*: $A + B = B + A$
- addition is *assocative*: $A + (B + C) = (A + B) + C$
- multiplication by a number is *distributive* over addition: $c(A + B) = cA + cB$

Multiplication does not have all these properties:

- multiplication is not always *commutative*: $AB \neq BA$
- multiplication is *assocative*: $A(BC) = (AB)C$ and $A(B\mathbf{x}) = (AB)\mathbf{x}$
- multiplication is *distributive* over addition: $C(A + B) = CA + CB$ and $(A + B)C = AC + BC$

Here A, B, C are matrices of the appropriate size, $c \in \mathbf{R}$, and $\mathbf{x}$ is a vector.

Example 1.14. The identity (also called the *multiplicative identity*) and zero (also called the *additive identity*) matrices have special properties with addition and multiplication. For any $A \in \mathcal{M}_{m\times n}$:

- the product of A with I is A itself: $AI = IA = I$
- the product of A with 0 is is 0: $A0 = 0A = 0$
- the sum of A and 0 is A itself: $A + 0 = 0 + A = A$

In the second property, the zero matrix 0 does not have the same size every time it is used.

Definition 1.15. Let A be an $n \times n$ matrix. The *inverse* of A is a matrix B for which $AB = BA = I$.

Note that the inverse of a matrix A does not always exist. When it does, it is usually denoted A^{-1}. As a result of the first property from Example 1.14, the inverse of the identity matrix is itself: $II = I$, so $I^{-1} = I$.

Moreover, if $A \in \mathcal{M}_{m\times n}$ and $m \neq n$, then there may be a matrix $B \in \mathcal{M}_{n\times m}$ for which $AB = I$, but not necessarily $BA = I$, in which case B is called a *right inverse* of A. We will later see algorithms that compute the inverse, for now we just look at some examples.

Example 1.16. The inverse of the *difference matrix* is a *sum matrix*. That is, for

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

we have $AB = I$. Both of these matrices are triangular, or more specifically, lower triangular. These matrices get their names from what they do to a vector $\mathbf{x} = (x_1, x_2, x_3, x_4)$:

$$A\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \\ x_4 - x_3 \end{bmatrix}, \qquad B\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 + x_1 \\ x_3 + x_2 + x_1 \\ x_4 + x_3 + x_2 + x_1 \end{bmatrix}.$$

Example 1.17. The *cyclic matrix* C does not have an inverse. That is, there is no vector $\mathbf{x}$ for which

$$C\mathbf{x} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \mathbf{a},$$

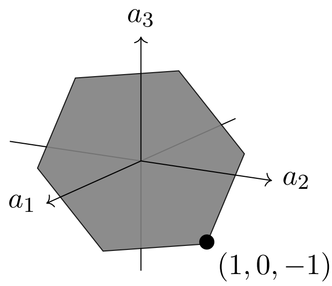
for any chosen $\mathbf{a}$. It is immediate that $\mathbf{a} = 0$ has a solution, when $x_1 = x_2 = x_3$. But it is also immediate that $\mathbf{a} = (1, 2, 3)$ is not a solution, because adding the three equations

$$x_1 - x_3 = 1, \qquad x_2 - x_1 = 2, \qquad x_3 - x_2 = 3,$$

gives 0 on the left side and 6 on the left. In this situation, we say:

- when $a_1 + a_2 + a_3 = 0$, there is a solution to $C\mathbf{x} = \mathbf{a}$, or equivalently,
- all linear combinations $x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + x_3\mathbf{c}_3$ lie on the plane given by $a_1 + a_2 + a_3 = 0$,

where $C = [\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3]$. If we consider a_1, a_2, a_3 as changing along the x, y, z axes,respectively, we see the collection of linear combinations $x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + x_3\mathbf{c}_3$ is indeed a plane:



1.3 Exercises

Exercise 1.1. Consider the four vectors $\mathbf{v} = \begin{bmatrix} 0 \\ 6 \\ -1 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} -3 \\ -4 \\ -5 \end{bmatrix}, \mathbf{z} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} -5 \\ 5 \\ -4 \end{bmatrix}$.

1. Find $a, b, c \in \mathbf{R}$ with $a\mathbf{v} + b\mathbf{w} + c\mathbf{z} = \mathbf{y}$.
2. Write your solution from part (a) as an equation $A\mathbf{x} = \mathbf{y}$, where A is a 3×3 matrix and $\mathbf{x}$ is a vector in $\mathbf{R}^3$.

Exercise 1.2. Check that the dot product from Definition 1.7 is *distributive* over vector addition. That is, show that $\mathbf{v} \cdot (\mathbf{u} + \mathbf{w}) = \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{w}$, for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^n$.

Exercise 1.3. Let $\mathbf{v} \in \mathbf{R}^3$ be non-trivial, and let $\mathbf{w}, \mathbf{z} \in \mathbf{R}^3$ be non-trivial vectors perpendicular to $\mathbf{v}$. Show that the halfway point between $\mathbf{w}$ and $\mathbf{z}$ is also perpendicular to $\mathbf{v}$.

Exercise 1.4. A non-square matrix A may have (non-square) matrices B, C for which $AB = I$ and $CA = I$, in which case we call B a *right inverse* and C a *left inverse* for A. Let $A = \left[\begin{smallmatrix} 1 & 0 & -2 \\ 3 & -1 & 1 \end{smallmatrix}\right]$.

1. Construct a right inverse for A, that is, a 3×2 matrix B for which $AB = I$. Make it so that $BA \neq I$.
2. Try to construct a left inverse for A, that is, a 3×2 matrix C for which $CA = I$. Is it possible?

Exercise 1.5. Recall the definition of the inverse of a matrix A, which is a matrix B for which $AB = BA = I$. Show that B is *unique*. That is, show that if there exists a matrix C with $AC = CA = I$, then $C = B$.

Exercise 1.6. This question is about *triangular* matrices.

1. Show that the product of two lower trinagular matrices is lower triangular.
2. Show that the product of two upper trinagular matrices is upper triangular. The concept of a *transpose*, introduced in the next lecture, will make this computation easier, given your work from part (a).
3. What form will the product of a lower triangular with an upper triangular matrix have? Can you come up with an example where the result is a diagonal matrix, but the original matrices are not diagonal?

Lecture 2: Elimination and inverses

This lecture reviews how to solve linear systems, and goes into more detail. Recall the three *elementary row operations*:

1. add a multiple of one row to another row
2. swap two rows
3. multiply a row by a nonzero number

These are not all equal operations: the third is a special case of the first, and the second changes some key aspects of the linear system. We will understand these operations as matrix multiplication.

2.1 Gaussian elimination

The main object of study for this lecture is the *matrix equation* $A\mathbf{x} = \mathbf{b}$, where $A \in \mathcal{M}_{m\times n}$, $\mathbf{b} \in \mathbf{R}^n$ and $\mathbf{x}$ is a column of n variables $x_1, \dots, x_n$. You should understand this equation in two ways:

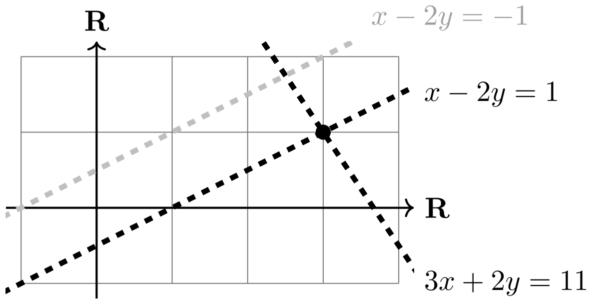
- by the *columns* of A: a linear combination of the n columns of A produces the vector $\mathbf{b}$
- by the *rows* of A: the m equations from the m rows of A describe m planes meeting at the point $\mathbf{x} \in \mathbf{R}^n$

Note that the word *plane* comes from a flat surface living in space (that is, $\mathbf{R}^3$). It is more precise to say *hyperplane* to describe all the points in $\mathbf{R}^n$ satisfying a single equation.

Example 2.1. Let $A = \left[\begin{smallmatrix} 3 & 2 \\ 1 & -2 \end{smallmatrix}\right]$ and $\mathbf{b} = \left[\begin{smallmatrix} 11 \\ 1 \end{smallmatrix}\right]$, with $\mathbf{x} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$. As columns of A, we have a linear combination

$$x\mathbf{a}_1 + y\mathbf{a}_2 = \mathbf{b}, \qquad \text{or} \qquad x\begin{bmatrix} 3 \\ 1 \end{bmatrix} + y\begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 11 \\ 1 \end{bmatrix}.$$

As rows of A, we have two equations $3x + 2y = 11$ and $x - 2y = 1$, giving the lines in the following picture.



The point $(3, 1)$ where both lines meet the single solution $\mathbf{x}$ that solves the given matrix equation $A\mathbf{x} = \mathbf{b}$. Observe that:

- If the lines were parallel and not colinear, there would be *no solutions*, because the lines would not intersect. For example, if instead of $3x + 2y = 11$ we had $x - 2y = -1$.
- If the lines were parallel and colinear, there would be *infinitely many solutions*, because the lines would intersect at all points. For example, if instead of $3x + 2y = 11$ we had $2x - 4y = 2$.

Example 2.2. two special cases: $A = I$ and $\mathbf{b} = 0$.

Previously *Gaussian elimination* was presented in no particular order. The algorithm we use here has a specific order. We *start from the top* and *work downwards* to get zeros *below all the pivots*. For a matrix equation $A\mathbf{x} = b$, where $A \in \mathcal{M}_{m\times n}$, we apply this process to the *augmented matrix* $[A\ \mathbf{b}] \in \mathcal{M}_{m\times(n+1)}$, which is just A with the last column $\mathbf{b}$.

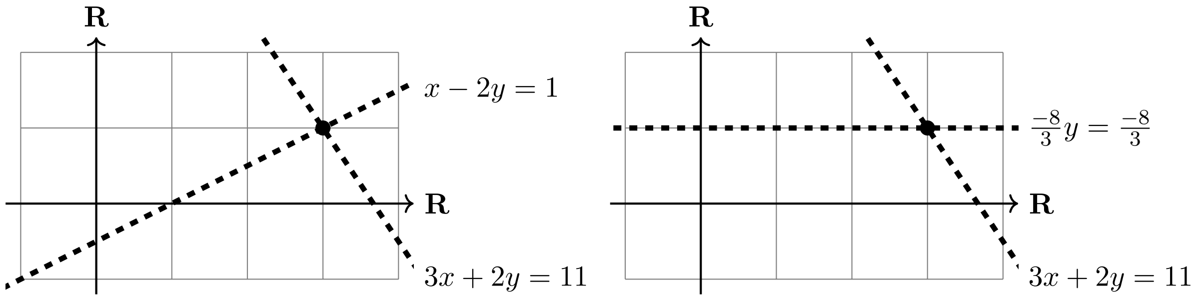
Example 2.3. We use the augmented matrix $\left[\begin{smallmatrix} 3 & 2 & 11 \\ 1 & -2 & 1 \end{smallmatrix}\right]$ from Example 2.1.

$$\begin{bmatrix} 3 & 2 & 11 \\ 1 & -2 & 1 \end{bmatrix} \quad 3 \text{ is the first } \textit{pivot}$$

$$\begin{bmatrix} 3 & 2 & 11 \\ 0 & -2-\frac{2}{3} & 1-\frac{11}{3} \end{bmatrix} \quad \frac{1}{3} \text{ is the } \textit{multiplier } \ell_{21} \qquad \text{previous matrix multiplied by } \begin{bmatrix} 1 & 0 \\ \frac{-1}{3} & 1 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 2 & 11 \\ 0 & \frac{-8}{3} & \frac{-8}{3} \end{bmatrix} \quad \frac{-8}{3} \text{ is the second pivot}$$

The geometric interpretation of applying the row operation comes from $x - 2y = 1$ becoming $\frac{-8}{3}y = \frac{-8}{3}$, which can be drawn the same as $y = 1$. Note that the point of intersection does not change.



Definition 2.4. The previous and next examples introduce key concepts for elimination on an $m \times n$ matrix A and on the augmented $m \times (n+1)$ matrix $[A\ \mathbf{b}]$.

- The *pivot* in row i is the first nonzero value that appears in row i. If the first nonzero value is the last entry (in the vector $\mathbf{b}$) or all entries of row i are 0, the pivot in row i does not exist.
- Row j is multiplied by the *multiplier* ℓ_{ij}, and the resulting row is subtracted from row i to get a zero below the pivot of row j.
- The row operation of subtracting ℓ_{ij} times row j from row i is done by the *elimination matrix* E_{ij}. This is the identity matrix I with $-\ell_{ij}$ in ij-position.
- The row operation of swapping row i with row j is done by the *permutation matrix* P_{ij}. This is the identity matrix I with the row i and row j swapped.

Both elimination and permuatation matrices are called *elementary matrices*. A permuatation matrix has a slightly more general definition.

Definition 2.5. A *permutation matrix* is the identity matrix I with a different rearrangement of rows. For example,

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

are all 4×4 permutation matrices. The identity matrix I is a trivial permutation matrix, or not considered a permutation matrix.

Example 2.6. Let $A = \left[\begin{smallmatrix} 0 & 6 & -2 \\ 4 & 8 & -4 \\ -2 & 2 & 7 \end{smallmatrix}\right]$ and $\mathbf{b} = \left[\begin{smallmatrix} 2 \\ 8 \\ 12 \end{smallmatrix}\right]$. For this augmented matrix, the first pivot seems to be zero, but we cannot have that, so we swap the second row with the first row. Elementary matrices are given on the right.

$$\begin{bmatrix} 0 & 6 & -2 & 2 \\ 4 & 8 & -4 & 8 \\ -2 & 2 & 7 & 12 \end{bmatrix} \quad \text{0 can not be a pivot}$$

$$\begin{bmatrix} 4 & 8 & -4 & 8 \\ 0 & 6 & -2 & 2 \\ -2 & 2 & 7 & 12 \end{bmatrix} \quad \text{swap first two rows, 4 is first pivot} \qquad \text{previous matrix multiplied by } \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 8 & -4 & 8 \\ 0 & 6 & -2 & 2 \\ 0 & 6 & 5 & 16 \end{bmatrix} \quad \frac{-1}{2} \text{ is multiplier } \ell_{31}\text{, 6 is second pivot} \qquad \text{previous matrix multiplied by } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 8 & -4 & 8 \\ 0 & 6 & -2 & 2 \\ 0 & 0 & 7 & 14 \end{bmatrix} \quad \text{1 is multiplier } \ell_{32}\text{, 7 is third pivot} \qquad \text{previous matrix multiplied by } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

This is now a system $U\mathbf{x} = \mathbf{c}$, for $U = \left[\begin{smallmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ 0 & 0 & 7 \end{smallmatrix}\right]$ and $\mathbf{c} = \left[\begin{smallmatrix} 8 \\ 2 \\ 14 \end{smallmatrix}\right]$. Equivalently, we have three equations:

$$\begin{aligned} 4x + 8y - 4z &= 8, \\ 6y - 2z &= 2, \\ 7z &= 14. \end{aligned}$$

Note that U is upper triangular. To find the vector $\mathbf{x}$ which solves this system, use back substitution from the bottom row up to find $z = 2$, $y = 1$, $x = 2$.

Remark 2.7. There are two cases which we have not considered for a matrix equation $A\mathbf{x} = \mathbf{b}$, but which were mentioned in Example 2.1. Both of these happen when an $n \times n$ matrix we get less than n pivots. If elimination produces in the augmented matrix $[A \ \mathbf{b}]$ a row of:

- all zeros except the last entry: then there are *no solutions*, because it implies an equation such as $0x + 0y + 0z = 1$, or $0 = 1$.
- all zeros: then there are *inifinitely many solutions*, because we then only have $n - 1$ equations but still n unknowns, so one of the unknowns can be freely chosen.

In both of these cases the matrix A is called *singular*.

2.2 Inverses and factorization

In Lecture 1.2 an inverse matrix in Definition 1.15 for a matrix $A \in \mathcal{M}_{n\times n}$ was simply described as a matrix B for which $AB = I$. However, this matrix B exists if and only if, equivalently:

- A has n pivots, or
- the only solution to $A\mathbf{x} = 0$ is $\mathbf{x} = 0$.

Then we write A^{-1} instead of B.

Remark 2.8. We note some common inverses.

- The inverse of a 2×2 matrix exists if and only if $ad - bc \neq 0$:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

- The inverse of a diagonal matrix exists iff the entries on the diagonal are nonzero:

$$\begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}^{-1} = \begin{bmatrix} 1/d_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/d_2 & 0 & \cdots & 0 \\ 0 & 0 & 1/d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/d_n \end{bmatrix}$$

- Similarly, the inverse of an upper triangular matrix exists iff the entries on the diagonal are nonzero. If some are zero, it immediately means we are missing some pivots (as everything below the diagonal is zero).

Taking the inverse of a product of matrices reverses their order: $(AB)^{-1} = B^{-1}A^{-1}$. This follows as

$$\begin{aligned} AB(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} && \text{(commutativity of multiplication)} \\ &= AIA^{-1} && \text{(definition of inverse)} \\ &= (AI)A^{-1} && \text{(commutativity of multiplication)} \\ &= AA^{-1} && \text{(property of identity matrix)} \\ &= I && \text{(definition of inverse)} \end{aligned}$$

Finally we come to a constructive definition of the inverse of a matrix, which, instead of *Gaussian* elimination, uses *Gauss–Jordan* elimination. This starts with the usual Gaussian elimination procedure, but then runs it in reverse, upside down: start *from the bottom* and *work upwards* to get zeros *above all the pivots*.

Example 2.9. Let $A = \begin{bmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ -2 & 2 & 1 \end{bmatrix}$, for which we want to find the inverse. To do this, we work with the block matrix $[A\ I]$, and on it we do not only Gaussian elimination on the matrix, as in Example 2.6, but also Gauss–Jordan elimination, which clears the matrix above the pivots. Elementary matrices are given on the left.

$$\begin{bmatrix} 4 & 8 & -4 & 1 & 0 & 0 \\ 0 & 6 & -2 & 0 & 1 & 0 \\ -2 & 2 & 1 & 0 & 0 & 1 \end{bmatrix} \quad \text{4 is first pivot}$$

$$\begin{bmatrix} 4 & 8 & -4 & 1 & 0 & 0 \\ 0 & 6 & -2 & 0 & 1 & 0 \\ 0 & 6 & -1 & 1/2 & 0 & 1 \end{bmatrix} \quad \frac{-1}{2} \text{ is multiplier } \ell_{31}\text{, 6 is second pivot} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 8 & -4 & 1 & 0 & 0 \\ 0 & 6 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1/2 & -1 & 1 \end{bmatrix} \quad \text{1 is multiplier } \ell_{32}\text{, 1 is third pivot} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 8 & -4 & 1 & 0 & 0 \\ 0 & 6 & 0 & -1 & -1 & -2 \\ 0 & 0 & 1 & 1/2 & -1 & 1 \end{bmatrix} \quad \text{0 above third pivot, second line} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 8 & 0 & 3 & -4 & 4 \\ 0 & 6 & 0 & -1 & -1 & -2 \\ 0 & 0 & 1 & 1/2 & -1 & 1 \end{bmatrix} \quad \text{0 above third pivot, second line} \quad \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 0 & 0 & 13/3 & -8/3 & 20/3 \\ 0 & 6 & 0 & -1 & -1 & -2 \\ 0 & 0 & 1 & 1/2 & -1 & 1 \end{bmatrix} \quad \text{0 above second pivot} \quad \begin{bmatrix} 1 & -8/6 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 13/12 & -2/3 & 5/3 \\ 0 & 1 & 0 & -1/6 & -1/6 & -1/3 \\ 0 & 0 & 1 & 1/2 & -1 & 1 \end{bmatrix} \quad \text{multiply by the pivot reciprocals} \quad \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/6 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

We have now reached the matrix $[I\ A^{-1}]$. To see the submatrix on the right is really the inverse, first multiply the elementary matrices together to get E. Above we showed that

$$E[A\ I] = [I\ B]$$

for some matrix B (which we are trying to show is the inverse of A). Block multiplication tells us that

$$E[A\ I] = [EA\ EI] = [EA\ E] \quad \Longrightarrow \quad EA = I \text{ and } E = B.$$

It follows that $BA = I$, which means that B is the inverse of A.

Remark 2.10. We now have a new, equivalent definitionf of $A \in \mathcal{M}_{n\times n}$. not having an inverse: If Gauss–Jordan elimination of $[A\ I]$ results in $[J\ B]$, where J is almost I, but has some zeros on the diagonal, then A has no inverse.

We are now at the final theme of this lecture: *factorization*, or *decomposition*. We want to describe a matrix A in terms of simpler, triangular matrices. We will do this in four ways:

$$A = LU, \qquad A = LDU, \qquad PA = LU, \qquad PA = LDU.$$

A is the original matrix, L is a lower triangular matrix, U is upper triangular, D is diagonal, and P is a permutation matrix. The first two ways are for matrices that do not require row swaps when doing elimination, otherwise row swaps are caputred in the permutation matrix P. Note that we must do all the permutations first.

Example 2.11. Consider the matrix $A = \begin{bmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ -2 & 2 & 1 \end{bmatrix}$ from Example 2.9. Using the elementary matrices from the first two steps, we find:

$$E_{32}E_{31}A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ -2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ 0 & 0 & 1 \end{bmatrix} = U,$$

since U is upper triangular. Next, observe that the inverse E_{ij}^{-1} of an eliminatoin matrix E_{ij} is almost the same as E_{ij}, except: where E_{ij} had $-\ell_{ij}$ in the ij-position, E_{ij}^{-1} has ℓ_{ij} in the ij-position. That is, for this example we have

$$E_{31}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/2 & 0 & 1 \end{bmatrix}, \qquad E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \qquad E_{31}^{-1}E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/2 & 1 & 1 \end{bmatrix} = L,$$

because L is upper triangular. Now we have the desired result

$$A = LU, \qquad \text{or} \qquad \underbrace{\begin{bmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ -2 & 2 & 1 \end{bmatrix}}_{\text{original matrix}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/2 & 1 & 1 \end{bmatrix}}_{\text{lower triangular factor}} \underbrace{\begin{bmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{upper triangular factor}}.$$

We can also factor U further as DU, placing the pivots of U in a separate matrix:

$$A = LDU \qquad \text{or} \qquad \begin{bmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ -2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/2 & 1 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 4 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{diagonal factor}} \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -1/3 \\ 0 & 0 & 1 \end{bmatrix}.$$

We finish off the lecture with some useful types of matrices.

Definition 2.12. Let A be an $n \times m$ matrix.

- The *transpose* of A is an $m \times n$ matrix denoted A^T, with $(A^T)_{ij} = A_{ji}$.
- The matrix A is *symmetric* if $m = n$ and $A_{ij} = A_{ji}$ for all i, j.
- The matrix A is *skew-symmetric* if $m = n$ and $A_{ij} = -A_{ji}$ for all i, j.

Observe that another way to express that A is symmetric is to say that $A = A^T$, and another way to express that A is skew-symmetric is to say $A = -A^T$.

Remark 2.13. The transpose can be thought of as a function $\mathcal{M}_{m\times n} \to \mathcal{M}_{n\times m}$. It plays nicely with the addition, multiplication, and inverse functions we have seen in Definitions 1.12 and 1.15:

- the transpose of a sum is the sum of the transposes: $(A+B)^T = A^T + B^T$
- the transpose of a product is the product oif the transposes, but reversed: $(AB)^T = B^T A^T$
- the transpose of an inverse is the inverse of the tranpose: $(A^{-1})^T = (A^T)^{-1}$

Moreover, the dot product of two vectors from Definition 1.7 can be thought of as matrix multiplication, if we use the transpose:

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} \tag{1}$$

$\mathbf{v} \in \mathbf{R}^n$, $\mathbf{w} \in \mathbf{R}^n$, $\mathbf{v}^T \in \mathcal{M}_{1\times n}$, $\mathbf{w} \in \mathcal{M}_{n\times 1}$

This is why we need to be careful with the multiplication symbol $\cdot$, always being aware of the sizes of objects we are working with. That is because multiplying the other way $\mathbf{w}\mathbf{v}^T$ gives an $n \times n$ matrix, which is called the *outer product*:

$$\mathbf{v} = \begin{bmatrix}1\\2\\3\\4\end{bmatrix} \in \mathbf{R}^4, \qquad \mathbf{v}^T\mathbf{w} = \begin{bmatrix}1 & 2 & 3 & 4\end{bmatrix}\begin{bmatrix}1\\-1\\2\\-2\end{bmatrix} = 1\cdot 1 + 2\cdot(-1) + 3\cdot 2 + 4 \cdot (-2) = -3 \quad \in \mathbf{R} = \mathcal{M}_{1\times 1}$$

$$\mathbf{w} = \begin{bmatrix}1\\-1\\2\\-2\end{bmatrix} \in \mathbf{R}^4, \qquad \mathbf{w}\mathbf{v}^T = \begin{bmatrix}1\\2\\3\\4\end{bmatrix}\begin{bmatrix}1 & -1 & 2 & -2\end{bmatrix} = \begin{bmatrix}1 & -1 & 2 & -2\\2 & -2 & 4 & -4\\3 & -3 & 6 & -6\\4 & -4 & 8 & -8\end{bmatrix} \qquad \in \mathcal{M}_{4\times 4}$$

Example 2.14. Taking the transpose of a product of a matrix with a vector is just like taking the tranpose of two matrices. Using the property from Equation (1) and the observations in Remark 2.13, we see some interesting results. For $A \in \mathcal{M}_{m\times n}$ and $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, we have

$$A\mathbf{x} \cdot \mathbf{y} = (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^T A^T \mathbf{y} = \mathbf{x}^T(A^T\mathbf{y}) = \mathbf{x} \cdot (A^T\mathbf{y}).$$

Remark 2.15. We end this lecture with an observation about symmetric matrices: If $A \in \mathcal{M}_{n\times n}$ is symmetric, then its decomposition into $A = LDU$ has $L = U^T$.

2.3 Exercises

Exercise 2.1. Construct a 3×3 matrix A which has:

1. pivots 1,2,3
2. pivots 1,2,3 and multipliers $\ell_{32} = 4$, $\ell_{31} = 5$ and $\ell_{21} = 6$
3. only two pivots 1 and 2, but no zeros in any positions

Exercise 2.2. Let A be a 3×3 matrix.

1. Find the pivots when A has each of the following forms. The numbers $a, \ldots, i$ are all nonzero.

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \qquad \begin{bmatrix} 0 & b & c \\ 0 & e & f \\ 0 & h & i \end{bmatrix} \qquad \begin{bmatrix} a & b & c \\ d & bd/a & f \\ d & bd/a & i \end{bmatrix} \qquad \begin{bmatrix} 0 & b & c \\ 0 & e & ce/b \\ 0 & e & ce/b \end{bmatrix}$$

all pivots | no first pivot | no second pivot | no first or third pivot

⋈ 2. Write a function that takes in such a matrix and returns a list of the three pivots. You may assume that all of the pivots exist.

⋈ 3. Run your function on 1000 random 3×3 matrices with entries in the range $[-1, 1]$. What is the range and the average of all the pivots? How often do you get a zero?

In Python, you may use consider A as a list of lists `[[a,b,c],[d,e,f],[g,h,i]]`.

Exercise 2.3. This question is about the three permutation matrix examples given in Definition 2.5.

1. Is the product of all three a permutation matrix?
2. Are the inverses of each still permutation matrices?

Exercise 2.4. Suppose that $A_i \in \mathcal{M}_{n\times n}$ has an inverse A_i^{-1}, for $i = 1, \ldots, k$. What is the inverse of the k-fold product $A_1 A_2 \cdots A_k$?

Exercise 2.5. Using Gauss–Jordan elimination, find the inverse matrix of $A = \left[\begin{smallmatrix} 0 & 2 & -1 \\ 1 & 0 & -4 \\ 2 & 2 & 2 \end{smallmatrix}\right]$.

Exercise 2.6. Decompose the matrix A from Example 2.6 as $PA = LDU$.

Lecture 3: The column space and the nullspace

This lecture will introduce what it means to be "in the solution space" to an equation $A\mathbf{x} = \mathbf{b}$.

3.1 Vector spaces and the column space

Recall that a *field* is a set with nice properties, such as $\mathbf{R}, \mathbf{Q}, \mathbf{C}$. Fields have addition and multiplication built into them.

Definition 3.1. Let V be a set and F a field. The elements of F are called *scalars*. The set V is a *vector space* if there are two operations

- addition $+\colon V \times V \to V$,
- scalar multiplication $\cdot : F \times V \to V$,

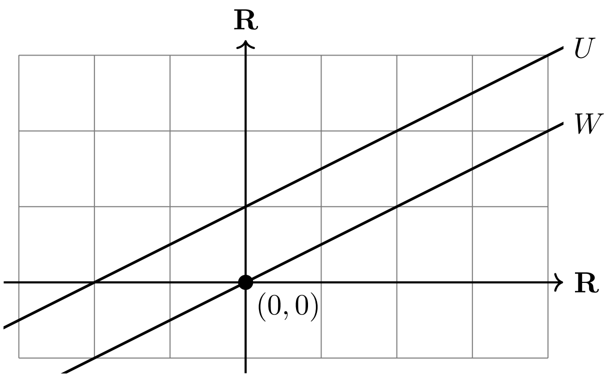
that satisfy the follow properties, for every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a, b \in F$:

1. addition has an identity element: there exists $0 \in V$ with $0 + \mathbf{v} = \mathbf{v}$
2. addition has inverse elements: there exists $-\mathbf{v} \in V$ with $\mathbf{v} + (-\mathbf{v}) = 0$
3. scalar multiplication has an identity element: there exists $1 \in F$ with $1\mathbf{v} = \mathbf{v}$
4. addition is commutative: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
5. addition is associative: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
6. scalar multiplication is distributive over addition: $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
7. scalar multiplication is distrbutive over field addition: $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$
8. field multiplication is compatible with scalar multiplication: $(ab)\mathbf{v} = a(b\mathbf{v})$

If V is a vector space and $W \subseteq V$ is a subset of V and is a vector space on its own, with the same two operations satisfying the same properties, then W is a *subspace* of V. It is immediate that every vector space is a subspace of itself, so whenever $W \subseteq V$ is a subspace and $W \neq V$, we say W is a *proper subspace* of V.

Example 3.2. We consider some basic examples of vector spaces.

- The empty set $\emptyset$ is not a vector space, because vector space must contain the zero vector.
- The set $V = \{0\}$ is a vector space, called the *trivial* or *zero* vector space.
- The space $\mathcal{M}_{2\times 2}$ is a vector space, with addition being matrix addition, and scalar multiplication the usual scalar multiplication over $\mathbf{R}$. This space is 4-dimensional, though we will see the notion of dimension next lecture.
- For $V = \mathbf{R}^2$, the set $W = \{c(2,1) \ : \ c \in \mathbf{R}\} \subseteq V$, which is all the multiples of $\mathbf{v} = (2,1)$, is a subspace of $\mathbf{R}^2$. The set $U = \{c(2,1) + (0,1) \ : \ c \in \mathbf{R}\} \subseteq V$, which is the same as W but shifted up by 1 unit, is not a vector space, as $(0,0) \notin U$.



Remark 3.3. We make some observations about vector spaces and subspaces.

- Every vector space and subspace must contain the zero vector.
- Any line through the origin is a subspace of $\mathbf{R}^n$.
- A subspace containing $\mathbf{u}$ and $\mathbf{v}$ must contain every linear combination $a\mathbf{u} + b\mathbf{v}$.

Example 3.4. Combining the above remark and Example 3.2, we see that $U = \{$all upper triangular matrices $\left[\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right]\} \subseteq \mathcal{M}_{2\times 2}$ is a subspace of $\mathcal{M}_{2\times 2}$, as is $D = \{$all diagonal matrices $\left[\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right]\} \subseteq \mathcal{M}_{2\times 2}$. Moreover, D is a subspace of U.

Example 3.5. Let V be any vector space, such as $\mathbf{R}^n$, and $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subseteq V$ any collection of elements of V. Then the space of all linear combinations of elements of S, written $W = \{\sum_{i=1}^n c_i \mathbf{v}_i \ : \ c_i \in F\}$, is a subspace of V. This space is called the *span* of the vectors in S, and we say V is *spanned* by those vectors.

Definition 3.6. Let V, W be two vector spaces. Their *direct sum*, or simply *sum*, is the vector space

$$V \oplus W := \{(\mathbf{v}, \mathbf{w}) \ : \ \mathbf{v} \in V, \mathbf{w} \in W\},$$

with vector addition and scalar multiplication defined component-wise. That is, $c(\mathbf{v}, \mathbf{w}) = (c\mathbf{v}, c\mathbf{w})$. This general approach is taken in case V and W are not subspaces of a common space. If there exists a vector space U with $V, W \subseteq U$, then we have the vector space

$$V + W := \{\mathbf{v} + \mathbf{w} \ : \ \mathbf{v} \in V, \mathbf{v} \in W\}.$$

In this case, we have all linear combinations of vectors from both spaces. This is called the subspace *generated by* U and V. It is the smallest subspace containing $U \cup V$, which itself is not necessarily a subspace.

Note that $V \oplus W$ and $V + W$ are subspaces, but $V \cup W$ is not. These three spaces are not the same, in fact $V \oplus W$ is never equal to $V + W$ (though there may be a nice function between the two.

Example 3.7. We note some common exmples of vector spaces generated by other spaces:

- The vector space generated by V and any of its subspaces W is the original space: $V + W = V$
- The vector space generated by two spans is the span of the union:

$$\text{span}(\{\mathbf{v}_1, \mathbf{v}_2\}) \ + \ \text{span}(\{\mathbf{w}_1, \mathbf{w}_2\}) = \text{span}(\{\mathbf{v}_1, \mathbf{v}_2\} \cup \{\mathbf{w}_1, \mathbf{w}_2\}) = \text{span}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{w}_1, \mathbf{w}_2\})$$

See Exercise 3.3 for more details on why the union of two vector spaces $V \cup W$ is not the same as $+$.

The reason we are talking about vector spaces is that the matrix product $A\mathbf{x}$ from the matrix equation $A\mathbf{x} = \mathbf{b}$, over all possibilities $\mathbf{x}$, describes a vector space. This space has a particular name.

Definition 3.8. For an $m \times n$ matrix A, the *column space* of A, denoted $C(A)$ or $\text{col}(A)$, is the set of all vectors $\mathbf{v} \in \mathbf{R}^m$ that are linear combinations of the columns of A. That is, it consists of the vectors

$$\mathbf{v} = c_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + c_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + c_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix},$$

for $c_i \in \mathbf{R}$ for all i. Since this is a linear combination of vectors, $\text{col}(A)$ is a subspace of $\mathbf{R}^m$.

In other words, the equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b} \in \text{col}(A)$. Note that $\in \text{col}(A)$ always.

Example 3.9. Consider the following matrices:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

The column space $\text{col}(I)$ is all of $\mathbf{R}^2$, since any vector $(a, b) \in \mathbf{R}^2$ can be described as $a \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] + b \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$, which is a linear combination of the columns of I. The column space of A is all multiples of the vector $\left[\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right]$, since the second and third rows are multiples of the first row.

3.2 The nullspace of a matrix

Definition 3.10. For an $m \times n$ matrix A, the *nullspace* of A is the set of all vectors $\mathbf{x} \in \mathbf{R}^n$ with $A\mathbf{x} = 0$. It is denoted $N(A)$ or $\text{null}(A)$.

The nullspace is a vector space, but it is not always a subspace of the column space. Indeed, the nullspace lives inside $\mathbf{R}^n$, but the column space lives in $\mathbf{R}^m$. When $m = n$, the n-dimensional 0 vector is always in the intersection of the two spaces. This is not saying much, however, as the interection of any two vector spaces contains at least the zero vector.

Example 3.11. The nullspace of the matrix $A = \left[\begin{smallmatrix} 2 & -1 \\ 4 & -2 \end{smallmatrix}\right]$ consists of the vectors in $\mathbf{x} \in \mathbf{R}^2$ for which $A\mathbf{x} = 0$. By elimination, we see the second line is a multiple of the first, so the nullspace is all pairs (x_1, x_2) for which $2x_1 - x_2 = 0$, or $x_1 = x_2/2$. Choosing $x_2 = 1$ (though we could choose any other value) we get $x_1 = 1/2$, so the nullspace is all multiples of the vector $(1/2, 1)$.

The choice $(1/2, 1)$ was a *special solution*, but there are many other solutions.

Remark 3.12. Elimination on a matrix does not change its nullspace. We can see this by considering the original equation $A\mathbf{x} = 0$ and the elminiated equation $EA\mathbf{x} = 0$. Since E is an elmentary matrix, it has an inverse, so $A\mathbf{x} = E^{-1}0 = 0$. Hence $\mathbf{x}$ satisfies the first equation iff it satisfies the second equation.

To compute the nullspace in general, we do Gaussian and Gauss-Jordan elmination, and end up with 1 on the diagonal.

Example 3.13. Compute the nullspace of the matrix $A = \left[\begin{smallmatrix} 2 & -2 & 2 & 4 & 8 \\ 1 & 5 & -3 & 0 & 1 \\ 3 & 3 & -1 & -5 & 6 \end{smallmatrix}\right]$. We begin with Gaussian elmination to get zeros below the first pivot:

$$\ell_{21} = \frac{1}{2}, \quad \ell_{31} = \frac{3}{2} : \quad \begin{bmatrix} 2 & -2 & 2 & 4 & 8 \\ 0 & 6 & -4 & -2 & -3 \\ 0 & 6 & -4 & -11 & -6 \end{bmatrix}.$$

We continue to get a zero below the second pivot:

$$\ell_{32} = 1 : \quad \begin{bmatrix} 2 & -2 & 2 & 4 & 8 \\ 0 & 6 & -4 & -2 & -3 \\ 0 & 0 & 0 & -9 & -3 \end{bmatrix}.$$

The third pivot is -9. Now we move upward and clear the entries above the third pivot:

$$\begin{bmatrix} 2 & -2 & 2 & 0 & 20/3 \\ 0 & 6 & -4 & 0 & -7/3 \\ 0 & 0 & 0 & -9 & -3 \end{bmatrix}.$$

Next, get a zero above the second pivot:

$$\begin{bmatrix} 2 & 0 & 2/3 & 0 & 53/9 \\ 0 & 6 & -4 & 0 & -7/3 \\ 0 & 0 & 0 & -9 & -3 \end{bmatrix}.$$

Finally, multiply through by the pivot reciprocals to get pivots that are 1:

$$\begin{bmatrix} 1 & 0 & 1/3 & 0 & 53/18 \\ 0 & 1 & -2/3 & 0 & -7/18 \\ 0 & 0 & 0 & 1 & 1/3 \end{bmatrix}.$$

This is called the *reduced row echelon form*, or *RREF*, of A. We continue solving for the nullspace after the following definitions.

Definition 3.14. In the example above, the columns 1,2,4 are the *pivot columns* and 3,5 are the *free columns*. The variables x_1, x_2, x_4 are the *pivot variables* and x_3, x_5 are the *free variables*.

The nullspace null(A) from Example 3.13 is defined as a linear combination of as many vectors as there are free columns. Each free column gives a nonzero $\mathbf{x}$ that will be in the nullspace, by setting that free variable to 1, all other free variables to 0, and choosing the earlier pivot variables to be the negative entries in those rows:

$$\begin{bmatrix} 1 & 0 & 1/3 & 0 & 53/18 \\ 0 & 1 & -2/3 & 0 & -7/18 \\ 0 & 0 & 0 & 1 & 1/3 \end{bmatrix} \underbrace{\begin{bmatrix} -1/3 \\ 2/3 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{s}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 1/3 & 0 & 53/18 \\ 0 & 1 & -2/3 & 0 & -7/18 \\ 0 & 0 & 0 & 1 & 1/3 \end{bmatrix} \underbrace{\begin{bmatrix} -53/18 \\ 7/18 \\ 0 \\ -1/3 \\ 1 \end{bmatrix}}_{\mathbf{r}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The two vectors $\mathbf{s}, \mathbf{r}$ are the *special solutions* for the nullspace of A. Hence the nullspace is

$$\text{null}(A) = \{a\mathbf{s} + b\mathbf{r} \ : \ a, b \in \mathbf{R}\} = \left\{ a \begin{bmatrix} -1/3 \\ 2/3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} -53/18 \\ 7/18 \\ 0 \\ -1/3 \\ 1 \end{bmatrix} \ : \ a, b \in \mathbf{R} \right\},$$

so for example, something like

$$\begin{bmatrix} -108 \\ 18 \\ 6 \\ -13 \\ 36 \end{bmatrix} = 6 \begin{bmatrix} -1/3 \\ 2/3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + 36 \begin{bmatrix} -53/18 \\ 7/18 \\ 0 \\ -1/3 \\ 1 \end{bmatrix}$$

is in the nullspace.

Remark 3.15. Note that the pivot columns create an identity matrix in RREF of A, as do the free variable rows in the special solutions.

If a square matrix has full rank, its nullspace contains only the zero vector.

3.3 Exercises

Exercise 3.1. Check that the subspace $W \subseteq V$ from in the fourth example in Example 3.2 satisfies the conditions of being a vector space from Definition 3.1.

Exercise 3.2. Let $V = \text{span}(\{\mathbf{u}, \mathbf{v}, \mathbf{w}\})$ and $W = \text{span}(\{\mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w}\})$. Show that $W \subseteq V$.

Exercise 3.3. Consider the following vector spaces:

$$V = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}, \qquad W = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

1. Show that $\mathbf{R}^3$ is a subspace of $V + W$ by describing an arbitrary vector $(x, y, z) \in \mathbf{R}^3$ as a linear combination of the elements of V and W.
2. Show that $V \cup W \neq V + W$ by finding a vector in $V + W$ that is not in $V \cup W$.

Exercise 3.4. Create a matrix with no zero columns that has:

1. size 3×3 and column space the xy-plane (that is, all linear combinations of $(1, 0, 0)$ and $(0, 1, 0)$)
2. size 3×4 and column space the xy-plane
3. size 2×2, column space all of $\mathbf{R}^2$, and no zero entries. Describe $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ as linear combinations of the columns.

Exercise 3.5. Let I be the 2×2 identity matrix. For each of the following matrices, bring it to RREF and drescribe its nullspace as a span of vectors.

$$A = \begin{bmatrix} I & I \end{bmatrix} \qquad B = \begin{bmatrix} I & I \\ 0 & I \end{bmatrix} \qquad C = \begin{bmatrix} I & I \\ I & I \end{bmatrix}$$

Lecture 4: Completely solving $A\mathbf{x} = \mathbf{b}$

Previously we saw how to solve $A\mathbf{x} = 0$, by doing elimination until we get an upper triangular matrix $R\mathbf{x} = 0$, whose solutions $\mathbf{x}$ are the same solutions that solve the first equation.

4.1 Rank and the particular solution

We begin with the example from the previous lecture,

$$A = \begin{bmatrix} 2 & -2 & 2 & 4 & 8 \\ 1 & 5 & -3 & 0 & 1 \\ 3 & 3 & -1 & -5 & 6 \end{bmatrix}, \qquad R = \begin{bmatrix} 1 & 0 & 1/3 & 0 & 53/18 \\ 0 & 1 & -2/3 & 0 & -7/18 \\ 0 & 0 & 0 & 1 & 1/3 \end{bmatrix}, \qquad EA = R$$

for some product of elimination matrices E. The columns 1,2,4 are the *pivot columns* and the columns 3,5 are the *free columns* (this is true for both R and A). It is immediate that columns 1,2,4 of R can not be written one as a linear combination of the others - that is, these three columns are *linearly independent*. Again, this is true for both R and A.

Definition 4.1. The *rank* of a matrix A is the number of pivots of A, and is denoted $\operatorname{rank}(A)$.

The rank can also be though of as the number of columns in A that are not linear combinations of the others. Reducing the matrix A to RREF reveals which columns are combinations of others. Since only row operations were performed, any linear (in)dependence among the columns is preserved.

Example 4.2. When a matrix has rank 1, all the columns are multiples of the first one. For example,

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

has rank one, and its column space is all the multiples of $(1, 1, 1)$. To find its nullspace, we first get its RREF, which is $\left[\begin{smallmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right]$, which has special solutions

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

hence its nullspace is the span of $\left[\begin{smallmatrix} -2 \\ 1 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} -3 \\ 0 \\ 1 \end{smallmatrix}\right]$.

Remark 4.3. A rank 1 square $n \times n$ matrix may be expressed as a product of a $n \times 1$ vector with a $1 \times n$ vector, since all the columns are multiples of the first column. For example,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \mathbf{v}\mathbf{w}^T.$$

Example 4.4. The identity matrix I has full rank. The zero matrix 0 has rank 0.

Definition 4.5. The number of special solutions to $A\mathbf{x} = 0$ is called the *nullity* of A.

The nullity is the number of free columns of A, and the smallest number of vectors used to define $\operatorname{null}(A)$ as a span. If $A \in \mathcal{M}_{m\times n}$, then we have a very powerful equation, which we will see later:

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n.$$

This is called the *rank-nullity theorem*.

So far we have constructed all solutions to the matrix equation $A\mathbf{x} = 0$. Now we do the same for the equation $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} \neq 0$.

Example 4.6. Recall Example 3.13 from the previous lecture, which we revisited at the beginning of this lecture. Instead of $A\mathbf{x} = 0$, we consider $A\mathbf{x} = \mathbf{b}$, which, after elimination, becomes $R\mathbf{x} = \mathbf{d} = [d_1\ d_2\ d_3]^T$. The vector $\mathbf{x} = 0$ is not a solution anymore, but we can find a quick solution by setting the variables corresponding to the free columns equal to 0:

$$\begin{bmatrix} 1 & 0 & 1/3 & 0 & 53/18 \\ 0 & 1 & -2/3 & 0 & -7/18 \\ 0 & 0 & 0 & 1 & 1/3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 0 \\ x_4 \\ 0 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}, \qquad \text{or} \qquad \begin{aligned} x_1 &= d_1, \\ x_2 &= d_2, \\ x_3 &= d_3. \end{aligned}$$

This is called a *particular solution* to $A\mathbf{x} = \mathbf{b}$. This particular solution solution $\mathbf{x} = [d_1\ d_2\ 0\ d_3\ 0]$ will also solve $A\mathbf{x} = \mathbf{b}$, because if $A = ER$, for some elimination matrix E, then $\mathbf{d} = E\mathbf{b}$.

Remark 4.7. What we have done so far can be summarized as follows:

- The special solutions $\mathbf{x} = \mathbf{s}, \mathbf{r}$ solve $A\mathbf{x} = 0$
- The particular solution $\mathbf{x} = \mathbf{p}$ solves $A\mathbf{x} = \mathbf{b}$

Finally, we get the *complete solution* to the system $A\mathbf{x} = \mathbf{b}$ is the sum of the particular and special solutions. That is, $\mathbf{x} = \mathbf{p} + x_3\mathbf{s} + x_5\mathbf{r}$ solves the system, for any $x_3, x_5 \in \mathbf{R}$, because

$$A(\mathbf{p} + x_3\mathbf{s} + x_5\mathbf{r}) = A\mathbf{p} + x_3A\mathbf{s} + x_5A\mathbf{r} = \mathbf{b} + x_3 \cdot 0 + x_5 \cdot 0 = \mathbf{b}.$$

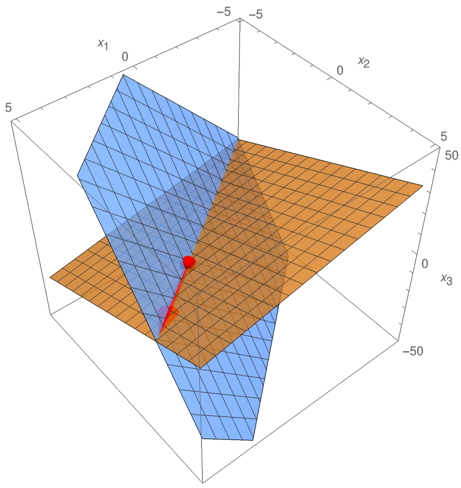
The suggested approach is to first find the special solution by doing elimination on the augmented matrix $[A\ \mathbf{b}]$ to get the matrix $[R\ \mathbf{d}]$. Then the particular solution is immediate from $\mathbf{d}$, and the special solutions can be found by solving $R\mathbf{s} = 0$, where $\mathbf{s}$ among all free variables has a single 1 (all others are 0).

Example 4.8. Consider the matrix equation

$$\begin{bmatrix} 4 & -8 & 2 \\ -10 & 12 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -16 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -3/4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7/4 \\ 1/8 \end{bmatrix}.$$

The complete solution to this equation is

$$\mathbf{x} = \begin{bmatrix} 7/4 \\ 1/8 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 3/4 \\ 1 \end{bmatrix},$$



for any $x_3 \in \mathbf{R}$. This equation represents two planes intersecting in space, the particular solution is a point on the line of intersecton, and the special solution is a vector in the direction of the line. The line is the nullspace.

4.2 Different types of complete solutions

Now we consider the implications for the complete solution given the rank of the matrix.

Definition 4.9. Let $A \in \mathcal{M}_{m \times n}$.

- If $m \geqslant n$ and A has n pivots, then A has *full column rank*.
- If $m \leqslant n$ and A has m pivots, then A has *full row rank*.
- If $m = n$ and A has n pivots, then A has *full rank*.

If $A \in \mathcal{M}_{m \times n}$ has full column rank, then in row reduced echelon form it looks like the block matrix $\left[\begin{smallmatrix} I \\ 0 \end{smallmatrix}\right]$, where I is of size $n \times n$ and the zero matrix 0 has size $(m - n) \times n$. Then:

- all columns of A are pivot columns,
- there are no free variables, so there are no special solutions,
- the nullspace contains only the zero vector $\text{null}(A) = \{0\}$,
- if $A\mathbf{x} = \mathbf{b}$ has a solution, there is one unique solution.

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

Example 4.10. The equation $A\mathbf{x} = \mathbf{b}$ with A an $m \times 3$ matrix with full column rank represents m planes intersecting in space. If the planes all intersect in one point, there is a solution to this equation.

If $A \in \mathcal{M}_{m \times n}$ has full row rank, then in row reduced echelon form it looks like the block matrix $[I\ 0]$, where I is of size $m \times m$ and the zero matrix 0 has size $m \times (n - m)$. Then:

- all rows of A have pivots, so there are no zero rows,
- there are $n - m$ special solutions,
- the column space is all of $\mathbf{R}^m$,
- $A\mathbf{x} = \mathbf{b}$ has a solution for any vector $\mathbf{b}$

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

Remark 4.11. We can summarize every matrix $A \in \mathcal{M}_{m \times n}$ as one of the following four situations.

- $\text{rank}(A) = m$, $\text{rank}(A) = n$: Then A is square and invertible, and $A\mathbf{x} = \mathbf{b}$ has exactly 1 solution.
- $\text{rank}(A) = m$, $\text{rank}(A) < n$: Then A is wider than it is taller, and $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions.
- $\text{rank}(A) < m$, $\text{rank}(A) = n$: Then A is taller than it is wider, and $A\mathbf{x} = \mathbf{b}$ has 0 or 1 solution.
- $\text{rank}(A) < m$ and $\text{rank}(A) < n$: Then A can have any shape, but it is not full rank, and $A\mathbf{x} = \mathbf{b}$ has either 0 or infintely many solutions.

4.3 Exercises

Exercise 4.1. Consider the two vectors $\mathbf{v} = [a\ a\ a\ a]^T$ and $\mathbf{w} = [1\ 1\ 1\ 1]^T$. What will be the rank of the 4×4 matrix $\mathbf{v}\mathbf{w}^T$? Your answer should depend on a.

Exercise 4.2. Find the complete solution to $A\mathbf{x} = \mathbf{b}$, for

$$A = \begin{bmatrix} 3 & 0 & -9 & -3 & 0 \\ 6 & 0 & -21 & 0 & 2 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 9 \\ -1 \end{bmatrix}.$$

Exercise 4.3. Suppose you know that the solution to a matrix equation $A\mathbf{x} = \mathbf{b}$, where $A \in \mathcal{M}_{2\times 3}$, is the vector

$$\mathbf{x} = \begin{bmatrix} 7 \\ 4 \\ -2 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix},$$

for any $x_2 \in \mathbf{R}$. Construct one possible matrix A and vector $\mathbf{b}$ for which this could be the solution.

Exercise 4.4. For the following matrices A, B, find the ranks of $A^T A$, AA^T, $B^T A$, BB^T:

$$A = \begin{bmatrix} 2 & 0 & 3 \\ -1 & 1 & 3 \end{bmatrix}, \qquad B = \begin{bmatrix} -1 & 3 \\ 9 & 0 \\ 7 & 0 \\ -3 & 1 \end{bmatrix}.$$

Lecture 5: Independence, basis, dimension

We have now arrived at the next big theme of this course: *dimension*.

5.1 Linear independence

Recall that the rank of a matrix was the number of pivots the matrix had. Another way to describe the rank is to use *linear independence* of the columns.

Definition 5.1. Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\} \subseteq \mathbf{R}^n$ be the columns of a matrix $A \in \mathcal{M}_{n\times k}$. These vectors are *linearly independent* if

- the only solution to $A\mathbf{x} = 0$ is $\mathbf{x} = 0$, or equivalently, if
- $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_k\mathbf{v}_k = 0$ implies $x_i = 0$ for all i.

If a set of vectors is not linearly independent, then the set is *linearly dependent*. Every set of vectors is either linearly independent or linearly dependent. We often say "the vectors are linearly independent" instead of "the set of vectors is linearly independet", but both are correct uses of the term.

Example 5.2. The vectors $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}2\\2\end{smallmatrix}\right]$ are linearly dependent, because $\left[\begin{smallmatrix}1&2\\1&2\end{smallmatrix}\right]\left[\begin{smallmatrix}-2\\1\end{smallmatrix}\right] = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$. The vectors $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}2\\2.1\end{smallmatrix}\right]$ are linearly independent, because attempting to solve $A\mathbf{x} = 0$ will lead to $\mathbf{x} = 0$.

Another equivalent way of saying that the vectors in the columns of A are linearly independent is to say that $(A) \neq \{0\}$. This is equivalent to the first definition given in Definition 5.1.

Remark 5.3. If there are more than 3 vectors in a collection of vectors from $\mathbf{R}^3$, the set must be linearly dependent. This follows from the fact that a 3×4 matrix A can have rank at most 3, so there will be at least one special solution $A\mathbf{x} = 0$.

Recall the *span* of a collection of vectors from Example 3.5, and the columns of a matrix *spanning* its column space, as well as the vectors from special solutions *spanning* the nullspace.

Definition 5.4. Let V be a vector space and $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subseteq V$. The set S *spans* V if $V = \operatorname{span}(S)$. Then S is called a *spanning set* of V. If for every other spanning set S' of V, the size of S is less than or equal to the size of S'. then S is called a *minimal spanning set*.

Example 5.5.

- The vector space $\mathbf{R}^3$ is spanned by the vectors $\left[\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right]$.
- The column space of a matrix is spanned by its columns. We can also talk about the *row space* of a matrix A, which is defined simply as the column space of the matrix A^T, and written $\operatorname{row}(A) = \operatorname{col}(A^T)$.

The idea of a minimal spanning set from Definition 5.4 can be made more precise with the idea of linear independence from Definition 5.1.

Definition 5.6. Let V be a vector space and $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subseteq V$. The set S is a *basis* for V if S spans V and V is linearly independent.

Example 5.7. The *standard basis* for $\mathbf{R}^3$ consists of the vectors $\left[\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right]$. In general, the *standard basis* for $\mathbf{R}^n$ consists of the n vectors that have zeros everywhere except in position i, for each $i = 1, \dots, n$. These vectors are often denoted $\mathbf{e}_1, \dots, \mathbf{e}_n$.

The standard basis is not the only basis for $\mathbf{R}^n$, and every $n \times n$ matrix will full rank has columns that give a basis for $\mathbf{R}^n$. We end this part with an example of how to directly relate a basis with a span.

Example 5.8. Consider the following vectors in $\mathbf{R}^4$. We know they span a subpsace of $\mathbf{R}^4$, because the span of any vectors is a vector space, but what is this space? In other words, what is a basis for the space that there four vectors span?

$$\mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ 7 \\ 1 \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 3 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 5 \\ 5 \\ 12 \\ 5 \end{bmatrix}$$

The first approach is to make them rows of a matrix and do Gaussian elimination. That will give us *zero rows*, which will correspond to linearly dependent vectors.

$$\begin{bmatrix} 3 & 2 & 7 & 1 \\ 1 & -1 & 2 & 3 \\ 5 & 5 & 12 & 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & 2 & 7 & 1 \\ 0 & -5/3 & -1/3 & 8/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We see immediately that there are two pivots, so the first two rows (vectors $\mathbf{u}$, $\mathbf{v}$) are linerarly independent. That is, $\mathbf{u}$ and $\mathbf{v}$ form a basis for $\text{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$. The second approach is to make them columns of a matrix and do Gaussian elimination. That will give us *free columns*, which will correspond to linearly dependent vectors.

$$\begin{bmatrix} 3 & 1 & 5 \\ 2 & -1 & 5 \\ 7 & 2 & 12 \\ 1 & 3 & 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & 1 & 5 \\ 0 & -5/3 & 13/3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that the non-free (pivot) columns indicate which of the original columns are basis vectors. We have reached the same conclusion as in the first approach.

5.2 Change of basis and dimension

The word *basis* is another name for *minimal spanning set*. It is often difficult to consider all possible spanning sets, so we use the basis definition more often. We now observe three key conclusions:

- bases are not unique
- every basis of a vector space must have the same number of vectors
- defining a vector space is equivalent to defining a basis for that vector space

The last conclusion comes from the fact that a basis B spans V, that is, $\text{span}(B) = V$.

Remark 5.9. Given a basis for a vector space V, every vector in V can be expressed as a linear combination of of vectors of that basis. For some vector space it is very obvious:

$$\begin{bmatrix} 4 \\ -2 \\ 8 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - 2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 8 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

However, if we have a different basis, how can we figure out what the linear combination is in the other basis? This is where the *change of basis matrix* appears. Suppose that B and B' are bases for V, with

$$B = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}, \qquad B' = \{\mathbf{w}_1, \dots, \mathbf{w}_k\}, \qquad \mathbf{v} \in V.$$

The coefficients for expressing $\mathbf{v}$ in the basis B are in the solution vector $\mathbf{x}$ to $[\mathbf{v}_1 \ \cdots \ \mathbf{v}_k]\mathbf{x} = \mathbf{v}$. Similarly, the coefficients for expressing $\mathbf{v}$ in the basis B' are in the solution vector $\mathbf{y}$ to $[\mathbf{w}_1 \ \cdots \ \mathbf{w}_k]\mathbf{y} =$

$\mathbf{v}$. These two vectors are related by the equation

$$\underbrace{\begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_k \\ | & & | \end{bmatrix}}_{\text{change of basis matrix}} \mathbf{x} = \mathbf{y}, \qquad \text{where} \qquad \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_k \\ | & & | \end{bmatrix} \mathbf{a}_i = \mathbf{w}_i.$$

Definition 5.10. Let V be a vector space. The *dimension* of V is the number of vectors in any basis of V. It is denoted $\dim(V)$.

This assumes a very important point: all vector spaces have a basis. The proof of this statement relies on something called the "axiom of choice", which is a foundational assumption of mathematics that has not been proved. The axiom of choice says that given an infinite collection of sets, you may take a single element from every set.

Example 5.11. We have already seen dimension, but under different names.

- The dimension of $\mathbf{R}^n$ is n
- The dimension of the column space of A is the rank of A

The dimension of the nullspace has a particular name, we call it the *nullity* of A, and write $\dim(\text{null}(A)) = \text{nullity}(A)$.

Recall the definition of $U \oplus V$ and $U + V$ from Definiton 3.6. There we saw that if $U = \text{span}(B)$ and $V = \text{span}(B')$, then $U + V = \text{span}(B \cup B')$. A similar statement holds for dimension.

Remark 5.12. Let V be a vector space with subspaces U, W.

- The intersection $U \cap W$ is a subspace of V
- The sum $+$ of vector spaces satisfies $\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W)$
- The sum $\oplus$ of vector spaces satisfies $\dim(U \oplus W) = \dim(U) + \dim(W)$

The third statement does not need that U, W be subspaces of the same space. Statements like this do not exist for the union of vector spaces spaces, because that is not necessarily a vector space.

Remark 5.13. Let V be a vector space and $U \subseteq V$. If $\dim(U) = \dim(V)$, then $U = V$. This follows by taking the basis $\mathbf{u}_1, \dots, u_n$ of U, and asking if there are any vectors in V which cannot be expressed as linear combinations of the $\mathbf{u}_i$. If no, then the spaces are the same. If there exists some $\mathbf{v}$, then $\mathbf{u}_1, \dots, \mathbf{u}_n, \mathbf{v}$ is a linearly independent set of $n+1$ vectors in V, which is impossible.

Definition 5.14. Let V be a vector space with $\dim(V) = n$, and $U \subseteq V$ a subspace of dimension $\dim(U) = k$. The *codimension* of U in V is $\text{codim}(U) = n - k$.

For example, lines are codimension 1 in $\mathbf{R}^2$, but codimension 2 in $\mathbf{R}^3$. The set of points in $\mathbf{R}^n$ that satisfy one linear equation (that goes through the origin) is codimension 1.

Example 5.15. The space of $n \times n$ matrices has dimension n^2. It has as a subspace the space of $n \times n$ upper triangular matrices, which has dimension $n(n+1)/2$. For $n = 2$, a basis for each of these spaces is

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

in the first case, and

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

in the second case.

5.3 Exercises

Exercise 5.1. Find two different sets of linearly independent vectors from the vectors

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

Exercise 5.2. For a 2×2 matrix, linear independence on the columns only depends on if one column is a multiple of the other.

$\bowtie$ (a) Generate 10 000 random 2×2 matrices, with real number entries in the range $[-5, 5]$. How many have column space dimension 1?

$\bowtie$ (b) Repeat the same as in part (a), but use integer entries in the range $[-5, 5]$. How many have column space dimension 1? **Bonus:** How many would you expect to have dimension 1?

Exercise 5.3. Consider the basis B for $\mathbf{R}^3$ and a vector $\mathbf{v}$,

$$B = \left\{ \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\1\\-1\end{bmatrix}, \begin{bmatrix}3\\0\\6\end{bmatrix} \right\}, \qquad \mathbf{v} = \begin{bmatrix}-3\\-1\\5\end{bmatrix}.$$

Express $\mathbf{v}$ in terms of B.

Exercise 5.4. Consider the plane $P = \{(x, y, z) \in \mathbf{R}^3 \ : \ 2x - 4y - 5z = 0\}$, which is a subspace of $\mathbf{R}^3$. What is its basis?

Exercise 5.5. Find the change of basis matrix from $\left\{ \left[\begin{smallmatrix}3\\2\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\1\end{smallmatrix}\right] \right\}$ to $\left\{ \left[\begin{smallmatrix}-2\\3\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\5\end{smallmatrix}\right] \right\}$.

Exercise 5.6. Prove the claims from Remark 5.12.

Lecture 6: The rank-nullity theorem

With this lecture we will finish constructing the row / column / null spaces, and move on to doing something with these spaces.

6.1 Looking ahead

Before we do that, we take a quick look ahead and cover some topics necessary for your machine learning course.

6.1.1 Planes and hyperplanes

In $\mathbf{R}^n$, a *hyperplane* is $H = \{\mathbf{v} \in \mathbf{R}^n \ : \ \mathbf{a}^T\mathbf{v} = [a_1 \ a_2 \ \cdots \ a_n]\mathbf{v} = a_0, a_i \in \mathbf{R}\}$, or all the vectors in $\mathbf{R}^n$ that satisfy a single linear equation. If $a_0 = 0$, it goes through the origin and has the same vector operations as $\mathbf{R}^n$, so is a subspace. If $a_0 \neq 0$, it is still called an (affine) subspace, and every vector in it may be expressed as

$$\mathbf{v} = \mathbf{a} + \mathbf{v}'.$$

Vector addition and scalar multiplication are defined by, for $\mathbf{v} = \mathbf{a} + \mathbf{v}'$ and $\mathbf{w} = \mathbf{a} + \mathbf{w}'$, and $c \in \mathbf{R}$:

- $\mathbf{v} + \mathbf{w} = \mathbf{a} + (\mathbf{v}' + \mathbf{w}')$
- $c\mathbf{v} = \mathbf{a} + c\mathbf{v}'$

6.1.2 Partial derivatives

We have seen functions $f \colon \mathbf{R} \to \mathbf{R}$, such as $f(x) = x^2$, but there are functions $f \colon \mathbf{R}^3 \to \mathbf{R}$, such as $f(\mathbf{x}) = x_1^2 - 2x_2 + x_2x_3^3$. The *partial derivative* of f with respect to x_i is

$$\begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \dfrac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2 + x_3^3 \\ 3x_2x_3^2 \end{bmatrix},$$

where in each case f is considered as a function of a single variable, and the other variables are treated as constants. IN the case of a single variable function, the derivative at a point was the *slope* of the function at that point. In functions of multiple variables, such as the f above, the vector above is the *gradient* of f, or the *vector of partial derivatives*.

6.1.3 Projections

Let $\mathbf{v} \in \mathbf{R}^n$. The vector $\mathbf{v}$ can be *projected* onto:

- another vector $\mathbf{w} \in \mathbf{R}^n$, resulting in $\text{proj}_{\mathbf{w}}(\mathbf{v}) = \frac{\mathbf{w}^T\mathbf{v}}{\mathbf{w}^T\mathbf{w}}\mathbf{w}$
- a hyperplane $U \subseteq \mathbf{R}^n$
- any subspace $V \subseteq \mathbf{R}^n$, resulting in $\text{proj}_V(\mathbf{v}) = A(A^TA)^{-1}A^T\mathbf{v}$, where the columns of A are linearly independent basis vectors of V

The simplest projections are onto the individual axes or planes in 3 dimensional space. For example, the projecion onto the xy-plane is:

$$\text{proj}_{xy\text{-plane}}\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

6.2 The four fundamental spaces

Given a matrix $A \in \mathcal{M}_{m\times n}$ and what results after applying row operations $R \in \mathcal{M}_{m\times n}$, we have seen three related vector spaces:

- the *column space* $\text{col}(A) \neq \text{col}(R)$, which is the span of the columns
- the *null space* $\text{null}(A) = \text{null}(R)$, which is the span of the special solutions to $A\mathbf{x} = 0$ or $R\mathbf{x} = 0$
- the *row space* $\text{row}(A) = \text{row}(R)$, which is the span of the rows

Note that $\text{row}(A^T) = \text{col}(A)$ and $\text{col}(A^T) = \text{row}(A)$. So what is $\text{null}(A^T)$?

Definition 6.1. Let $A \in \mathcal{M}_{m\times n}$. The *left nullspace* of A is the nullspace of A^T.

These four vector spaces, called the *four fundamental subspaces*, come together in very nice ways, which we now discuss. Let $A \in \mathcal{M}_{m\times n}$:

1. there are subspace relations:
   - $\text{col}(A) \subseteq \mathbf{R}^m$ and $\text{null}(A^T) \subseteq \mathbf{R}^m$ are subspaces
   - $\text{col}(A^T) \subseteq \mathbf{R}^n$ and $\text{null}(A) \subseteq \mathbf{R}^n$ are subspaces
2. there are dimension relations:
   - $\dim(\text{col}(A)) = \dim(\text{col}(A^T)) = \text{rank}(A) = \text{rank}(A^T)$
   - $\dim(\text{col}(A)) + \dim(\text{null}(A^T)) = m$
   - $\dim(\text{col}(A^T)) + \dim(\text{null}(A)) = n$

The last statement is called the *rank-nullity theorem*. We now look at more relations among these vector spaces.

Remark 6.2. Vectors in the column space of A are perpendicular to vectors in the left nullspace. For example, consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}, \qquad \text{col}(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix} \right\} = \text{span}\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}.$$

The left nullspace is

$$\text{null}(A^T) = \text{null}\left(\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix} \right) = \text{null}\left(\begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \right) = \text{span}\left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}.$$

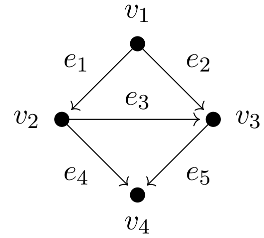
Taking the dot product of the basis vectors, we find

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

and so every vector in $\text{col}(A)$ is perpendicular to every vector in $\text{null}(A^T)$.

Example 6.3. For a practical application of these spaces, consider the following two matrices, both representations of the directed graph below. In A, the rows correspond to edges, and the columns correspond to vertices: each row has a -1 for the vertex where the edge starts and a 1 for the vertex where the edge ends. In B, entry B_{ij} is 1 if there is a directed edge from v_i to v_j, and 0 otherwise.

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



Bringing the matrix A^T to row reduced echelon form gives informsation about the row space of A and the left nullspace of A. The linearly independent rows of A are the rows corresponding to the edges e_2, e_2, e_4, and these edges form a *spanning tree* of the graph. The dependent row 3, corresponding to edge e_3, is dependent because adding it would create a *cycle* in the graph (among v_1, v_2, v_3), and cycles contain redundant information, so we want to get rid of cycles. Similarly we get a cycle if we add row 5, corresponding to edge e_5, because then we have a cycle of four edges.

6.3 Exercises

Exercise 6.1. Find a basis for the column space, nullspace, row space, and left nullspace of

$$A = \begin{bmatrix} 0 & 1 & a & a & a & a \\ 0 & 0 & 1 & b & b & b \\ 0 & 0 & 0 & 1 & c & c \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

for $a, b, c \in \mathbf{R}$. Be sure to consider situations when each of a, b, c are zero and when they are not.

Lecture 7: Orthogonality

The vector space pairs column space / nullspace and row space / left nullspace are special because of the relationship of each element of the pair to the other. In this lecture we will generalize this relationship.

7.1 Orthogonal spaces

Recall from Lecture 1 that two vectors $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$ are *orthogonal* if $\mathbf{u}^T\mathbf{v} = 0$. Note that in this case we have something that looks like the Pythagorean theorem:

$$\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = \mathbf{u}\cdot\mathbf{u} + 2\underbrace{\mathbf{u}\cdot\mathbf{v}}_{0} + \mathbf{v}\cdot\mathbf{v} = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

If $\mathbf{u}, \mathbf{v}$ are orthogonal and both have length 1, then they are called *orthonormal*

Definition 7.1. Two subspaces $U, V \subseteq \mathbf{R}^n$ are *orthogonal* if every pair of vectors $\mathbf{u} \in U, \mathbf{v} \in V$ is orthogonal.

Example 7.2. For $A \in \mathcal{M}_{m\times n}$, the nullspace null$(A)$ and the row space row(A) are orthognal to each other. Recall that $\mathbf{x} \in \text{null}(A)$ if $A\mathbf{x} = 0$. Another way of saying this is, for $\mathbf{r}_i \in \mathbf{R}^n$ a row of A, that

$$\begin{bmatrix} - & \mathbf{r}_1 & - \\ - & \mathbf{r}_2 & - \\ & \vdots & \\ - & \mathbf{r}_m & - \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{r}_1\cdot\mathbf{x} \\ \mathbf{r}_2\cdot\mathbf{x} \\ \vdots \\ \mathbf{r}_m\cdot\mathbf{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

and since the rowspace $\text{row}(A) = \text{span}(\{\mathbf{r}_1, \dots, \mathbf{r}_m\})$, we see that $\mathbf{v}\cdot\mathbf{x} = 0$ for any $\mathbf{v} \in \text{row}(A)$ and for any $\mathbf{x} \in \text{null}(A)$. Applying the same observation to the transpose A^T, we see that the left nullspace of A (which the nullspace of A^T) is orthogonal to the column space of A (which is the row space of A^T).

Remark 7.3. To check that two vector spaces are orthogonal, it suffices to check that every pair of elements $\mathbf{u} \in B$, $\mathbf{v} \in B'$ are orthogonal, for B a basis of U and B' a basis for V.

We now consider orthogonality in the context of particular matrices.

Example 7.4. The matrix $R_\theta := \left[\begin{smallmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{smallmatrix}\right]$ is called the *rotation matrix*, since the angle between $\mathbf{v} \in \mathbf{R}^2$ and $R_\theta\mathbf{v} \in \mathbf{R}^2$ is exactly θ. The columns of R_θ are orthogonal, as

$$\begin{bmatrix} \cos(\theta) & \sin(\theta) \end{bmatrix} \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix} = -\cos(\theta)\sin(\theta) + \sin(\theta)\cos(\theta) = 0,$$

for any angle θ. The columns are also orthonormal, as

$$\begin{bmatrix} \cos(\theta) & \sin(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} = \cos^2(\theta) + \sin^2(\theta) = 1,$$
$$\begin{bmatrix} -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix} = \sin^2(\theta) + \cos^2(\theta) = 1.$$

Example 7.5. The matrix $A \in \mathcal{M}_{3\times 6}$ does not have all orthogonal rows and columns, as row reduction shows we have only two pivots, meaning the row rank = column rank is 2:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 6 \\ 2 & 4 & 8 & 10 & 12 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad A^T = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 4 \\ 3 & 4 & 8 \\ 4 & 5 & 10 \\ 5 & 6 & 12 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

That is, columns 1 and 3 of A describe the 2-dimensional column space orthogonal to the 1-dimensional left nullspace of row 3. Analogously, columns 2,4,5 of A describe the 3-dimensional nullspace orthogonal to the row space of rows 1 and 2 of A:

$$\text{col}(A) = \text{span}\left\{\begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}3\\4\\8\end{bmatrix}\right\} \quad \text{is orthogonal to} \quad \text{null}(A^T) = \text{span}\left\{\begin{bmatrix}0\\-2\\1\end{bmatrix}\right\},$$

$$\text{null}(A) = \text{span}\left\{\begin{bmatrix}-2\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\-1\\1\\0\end{bmatrix}, \begin{bmatrix}-2\\0\\-1\\0\\1\end{bmatrix}\right\} \quad \text{is orthogonal to} \quad \text{row}(A) = \text{span}\left\{\begin{bmatrix}1\\2\\3\\4\\5\end{bmatrix}, \begin{bmatrix}1\\2\\4\\5\\6\end{bmatrix}\right\}.$$

We are left with a 2×2 *invertible submatrix* of A, hiding in the intersection of the pivot rows and pivot columns. This submatrix is important for finding left and right inverses of non-square matrices, and for *singular value decomposition*, which we will see later in the course.

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 6 \\ 2 & 4 & 8 & 10 & 12 \end{bmatrix}, \qquad A_{inv} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}.$$

7.2 Orthogonal relationships

Definition 7.6. If two subspaces $U, V \subseteq \mathbf{R}^n$ are orthogonal and $\dim(U) + \dim(V) = n$, then each is the *orthogonal complement* of the other in $\mathbf{R}^n$. That is, U is the orthogonal complement of V, written $U = V^\perp$, and V is the orthogonal complement of U, written $V = U^\perp$.

Remark 7.7. Recall the concept of *codimension* from Definition 5.14. The codimension of a space is equal to the dimension of its orthogonal complement. That is, $\text{codim}(U) = \dim(U^\perp)$.

Remark 7.8. It follows that, whenever we have orthogonal complements $U = V^\perp$, with $U, V \subseteq \mathbf{R}^n$ subspaces, then:

- $U + V = \mathbf{R}^n$, or in other words,
- any $\mathbf{x} \in \mathbf{R}^n$ can be expressed as a sum $\mathbf{x} = \mathbf{u} + \mathbf{v}$ of two elements, $\mathbf{u} \in U$ and $\mathbf{v} \in V$.

Theorem 7.9. *Let U, V be subspaces of $\mathbf{R}^n$. Then*

1. $(U^\perp)^\perp = U$
2. $(U + V)^\perp = U^\perp \cap V^\perp$
3. $(U \cap V)^\perp = U^\perp + V^\perp$

Proof. We only prove the second point, you will prove the other points in your homework. Recall that $U + V = \{\mathbf{u} + \mathbf{v} \ : \ \mathbf{u} \in U, \mathbf{v} \in V\}$. Take $\mathbf{u} \in U$, $\mathbf{v} \in V$, and $\mathbf{x} \in (U + V)^\perp$. To see that $(U + V)^\perp \subseteq U^\perp \cap V^\perp$, notice that $\mathbf{u}, \mathbf{v} \in U + V$, hence

$$\mathbf{u} \cdot \mathbf{x} = 0 \implies \mathbf{x} \in U^\perp, \qquad \mathbf{v} \cdot \mathbf{x} = 0 \implies \mathbf{x} \in V^\perp,$$

and so $\mathbf{x} \in U^\perp \cap V^\perp$. Since the vectors were arbitrary, we get that $(U + V)^\perp \subseteq U^\perp + V^\perp$. To see that $U^\perp \cap V^\perp \subseteq (U + V)^\perp$, take $\mathbf{y} \in U^\perp \cap V^\perp$, which means that both $\mathbf{y} \in U^\perp$ and $\mathbf{y} \in V^\perp$. Consider the arbitrary element $\mathbf{u} + \mathbf{v} \in U + V$, for which

$$\mathbf{y} \cdot (\mathbf{u} + \mathbf{v}) = \mathbf{y} \cdot \mathbf{u} + \mathbf{y} \cdot \mathbf{v} = 0 + 0 = 0,$$

meaning that $\mathbf{y} \in (U+V)^\perp$. Again, since the vectors are arbitrary, it follows that $U^\perp \cap V^\perp \subseteq (U+V)^\perp$. Combining these two statements, we get that $(U + V)^\perp = U^\perp \cap V^\perp$. ☐

Example 7.10. Combining Example 7.2 and the rank-nullity theorem from Lecture 6, for $A \in \mathcal{M}_{m\times n}$ we see that

- the nullspace and row space are orthogonal complements: $\text{null}(A) = \text{row}(A)^\perp$
- the left nullspace and column space are orthogonal complements: $\text{null}(A^T) = \text{col}(A)^\perp$

That is, along with Remark 7.8, any $\mathbf{x} \in \mathbf{R}^n$ can be written as a sum $\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$, where $\mathbf{x}_r \in \text{row}(A)$ and $\mathbf{x}_n \in \text{null}(A)$. It follws that no row of A can be in the nullspace of A.

We finish this lecture with an observation about bases of $\mathbf{R}^n$.

Remark 7.11. Recall that to be a basis of $\mathbf{R}^n$, a set of vectors has to be linearly independent and had to span $\mathbf{R}^n$. It follows that:

- If a set of n vectors is linearly independent, it spans $\mathbf{R}^n$.
- If n vectors span $\mathbf{R}^n$, they must be linearly independent.

The second fact comes from considering an $n \times n$ matrix A whose columns span $\mathbf{R}^n$, or equivalently, where for every $\mathbf{b} \in \mathbf{R}^n$ there is a unique solution $\mathbf{x}$ in $A\mathbf{x} = \mathbf{b}$. If we argue that the columns are linearly dependent, then there must be at least one special solution, and so infinitely many solutions to $A\mathbf{x} = \mathbf{b}$, but this contradicts what we originally assumed.

7.3 Exercises

Exercise 7.1. Confirm the observation from Remark 7.3. That is, let $B = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ be a basis for a vector space $U \subseteq \mathbf{R}^n$, and let $B' = \{\mathbf{v}_1, \dots, \mathbf{v}_\ell\}$ be a basis for a vector space $V \subseteq \mathbf{R}^n$. If you know that $\mathbf{u}_i \cdot \mathbf{v}_j = 0$ for all i, j, check that $\mathbf{u} \cdot \mathbf{v} = 0$ for arbitrary elements $\mathbf{u} \in U$ and $\mathbf{v} \in V$.

Exercise 7.2. Check that the claim about the angle between $\mathbf{v}$ and $R_\theta \mathbf{v}$ from Example 7.4 is indeed true.

Exercise 7.3. Let $A \in \mathcal{M}_{m\times n}$. Show that there is a bijective function $f\colon \text{row}(A) \to \text{col}(A)$. Hint: use orthogonality and the decomposition of vectors described in Remark 7.10.

Part II

Operations

Lecture 8: Projections and least squares

Now begins a new part of this course, in which we apply the knowledge about the objects we have learned so far, and put the objects into functions to see their relationships.

8.1 Projecting onto lines and spaces

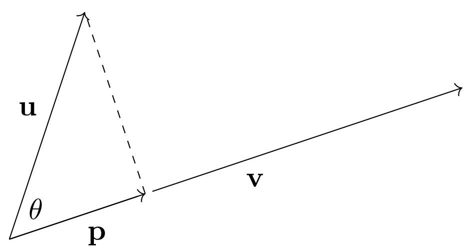
Projecting a vector $\mathbf{v}$ onto some other vector or onto a plane will produce a new vector that points along the other vector (in the first case) or lies in the plane (in the second case). Projections of a vector $\mathbf{v}$ can be understood in (at least) two ways:

1. the projection of $\mathbf{v}$ is the part of $\mathbf{v}$ that lies on the subspace to which you are projecting
2. the projection of $\mathbf{v}$ produces another vector $\mathbf{v}'$, so projecting is simply multiplying by a matrix: $A\mathbf{v} = \mathbf{v}'$

Both of these approaches are correct.

Example 8.1. Projecting $\mathbf{v} \in \mathbf{R}^3$ onto the y-axis is multiplying $\mathbf{v}$ by $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Projecting $\mathbf{v}$ onto the xy-plane is multiplying $\mathbf{v}$ by $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

In general, projecting a vector $\mathbf{u}$ onto a vector $\mathbf{v}$ uses the formula for the angle between them, from Proposition 1.9. Given two such arbitrary vectors, we want to compute the vector $\mathbf{p}$, which goes in the direction of $\mathbf{v}$, and is one side of a right triangle with $\mathbf{u}$ as hypotenuse.



$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{\text{adjacent}}{\text{hypotenuse}}$$

Since the hypotenuse has length $\|\mathbf{u}\|$, the adjacent, which is $\mathbf{p}$, must have length $\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{v}\|}$. The vector $\mathbf{v}$ may not have unit magnitude, but the vector $\frac{\mathbf{v}}{\|\mathbf{v}\|}$ does, and it goes in the same direction as $\mathbf{v}$. Hence $\mathbf{p}$ may be expressed as

$$\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|} \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}\mathbf{v}. \tag{2}$$

Definition 8.2. The *projection* of $\mathbf{u}$ onto $\mathbf{v}$ is the vector $\text{proj}_{\mathbf{v}}(\mathbf{u}) = \frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}$. The difference $\mathbf{u} - \text{proj}_{\mathbf{v}}\mathbf{u}$ is called the *error vector*.

Example 8.3. We note two trivial examples of projections.

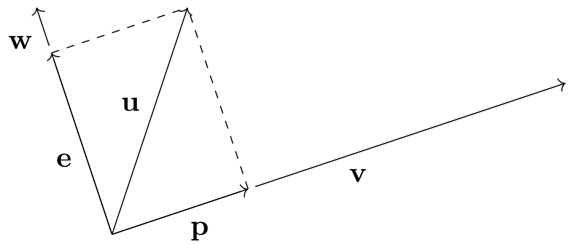
- Projecting $\mathbf{u}$ onto a line which is orthogonal to $\mathbf{u}$ gives the zero vector. This makes sense, because $\mathbf{u} \cdot \mathbf{v} = 0$ for all $\mathbf{v}$ in this line. In this case the error vector is equal to $\mathbf{u}$.
- Projecting $\mathbf{u}$ onto the line on which $\mathbf{u}$ already lies gives back $\mathbf{u}$. This also makes sense, because the line is all vectors $c\mathbf{u}$, for $c \in \mathbf{R}$, and for $\mathbf{v} = c\mathbf{u}$, the expression $\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}$ becomes $\frac{1}{c}$, and $\frac{1}{c}\mathbf{v} = \mathbf{u}$. In this case the error vector is the zero vector.

Considering the dot product as multiplication of matrices, Equation (2) becomes

$$\frac{\mathbf{u}^T\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v} = \frac{\mathbf{v}^T\mathbf{u}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v} = \mathbf{v}\frac{\mathbf{v}^T\mathbf{u}}{\mathbf{v}\cdot\mathbf{v}} = \frac{\mathbf{v}\mathbf{v}^T\mathbf{u}}{\mathbf{v}\cdot\mathbf{v}} = \underbrace{\frac{1}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}\mathbf{v}^T}_{P}\mathbf{u}. \tag{3}$$

The matrix P is the rank one *projection matrix*. The idea for it being rank one is that the projection goes to a 1-dimensional subspace, a line.

Remark 8.4. The error vector $\mathbf{e} = \mathbf{u} - \mathbf{p}$ from Definition 8.2 is also a type of projection, but onto a different vector, one that is orthogonal to $\mathbf{v}$ and $\mathbf{p}$.



To get a matrix for the projection of $\mathbf{u}$ onto $\mathbf{w}$, we want the result to be $\mathbf{e} = \mathbf{u} - \mathbf{p}$. Since $\mathbf{p} = P\mathbf{u}$, we quickly see that $\mathbf{e} = (I - P)\mathbf{u}$. Hence the projection matrix is $I - P$.

Next we consider the more general situation of projection a vector onto a subspace. Since all vector spaces have a spanning set, we consider a subspace to be a span of vectors. Combining these vectors as columns of a matrix, we get the column space.

Definition 8.5. Let $V = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subseteq \mathbf{R}^n$, and let A be the matrix with these vectors as its columns. For any $\mathbf{u} \in \mathbf{R}^n$, the *projection* of $\mathbf{u}$ onto V is the vector

$$\text{proj}_V(\mathbf{u}) = \underbrace{A(A^TA)^{-1}A^T}_{P}\mathbf{u}.$$

We assume the $\mathbf{v}_i$ are linearly independent, as otherwise A^TA does not have an inverse. If the v_i are not independent, remove the vectors that depend on others (this does not change the span).

The motivation for this expression is slightly more tedious, and comes from observing that for $\mathbf{p} = A\mathbf{x}$ the projection (for some appropriate $\mathbf{x}$), the vector $\mathbf{u} - A\mathbf{x}$ is orthogonal to the column space of A.

Remark 8.6. Since $V^\perp$ is the orthogonal complement of V, by Remark 7.8, every $\mathbf{u} \in \mathbf{R}^n$ can be expressed as $\mathbf{u} = \mathbf{v} + \mathbf{v}'$, where $\mathbf{v} \in V$ and $\mathbf{v}' \in V^\perp$. Since matrix multiplication is linear, and using the trivial projections from Example 8.3, it follows that

$$\begin{aligned} \text{proj}_V(\mathbf{u}) &= \text{proj}_V(\mathbf{v} + \mathbf{v}') = \text{proj}_V(\mathbf{v}) + \text{proj}_V(\mathbf{v}') = \mathbf{v} + 0 = \mathbf{v}, \\ \text{proj}_{V^\perp}(\mathbf{u}) &= \text{proj}_{V^\perp}(\mathbf{v} + \mathbf{v}') = \text{proj}_{V^\perp}(\mathbf{v}) + \text{proj}_{V^\perp}(\mathbf{v}') = 0 + \mathbf{v}' = \mathbf{v}', \end{aligned}$$

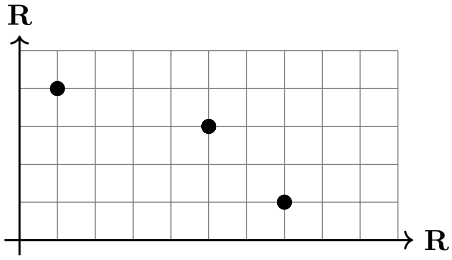
and so we always have $\mathbf{v} = \text{proj}_V(\mathbf{v}) + \text{proj}_{V^\perp}(\mathbf{v})$ for any $\mathbf{v} \in \mathbf{R}^n$. This gives a matrix for projecting onto the orthogonal complement, as

$$\text{proj}_{V^\perp}(\mathbf{u}) = \mathbf{u} - \text{proj}_V(\mathbf{u}) = \mathbf{u} - A(A^TA)^{-1}A^T\mathbf{u} = \underbrace{(I - A(A^TA)^{-1}A^T)}_{P}\mathbf{u}.$$

8.2 Least squares

One of the main applications of projections is finding the the closest solution to a linear system that has no exact solution

Example 8.7. Asking for a line $y = ax + b$ that goes through the three points $(1, 4)$, $(7, 1)$, $(5, 3)$ is impossible, because the points are not colinear. This is equivalent to asking for a solution to three equations, or to a linear system.



$$\begin{aligned} 4 &= a + b \\ 1 &= 7a + b \\ 3 &= 5a + b \end{aligned} \qquad \begin{bmatrix} 1 & 1 \\ 7 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$$

There is no solution to this matrix equation, because [4 1 3] is not in the column space of the matrix on the left. However, we still want to find a line that is "as close as possible", and projections help us do that.

Remark 8.8. Above we had a matrix equation $A\mathbf{x} = \mathbf{b}$ for which $\mathbf{b} \notin \text{col}(A)$. However, we can project $\mathbf{b}$ onto $\text{col}(A)$, which will guarantee a solution. That is, we can always write $\mathbf{b} = \mathbf{p} + \mathbf{e}$, where $\mathbf{p} \in \text{col}(A)$ and $\mathbf{e}$ is orthogonal to $\text{col}(A)$.

Definition 8.9. Let $A \in \mathcal{M}_{m\times n}$ and $\mathbf{b} \in \mathbf{R}^m$, with $\mathbf{b} = \mathbf{p} + \mathbf{e}$ and $\mathbf{p} \in \text{col}(A)$. The *least squares* solution to $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ that, equivalently,

- makes the distance between $A\mathbf{x}$ and $\mathbf{b}$ as small as possible
- makes the number $\|A\mathbf{x} - \mathbf{b}\|$ as small as possible
- makes the number $\|A\mathbf{x} - \mathbf{b}\|^2$ as small as possible

The reason for using the square of the length is to not have square roots, which are hard to deal with. The last two statements are equivalent because $a < b$ iff $a^2 < b^2$ for a, b nonnegative.

The first approach to finding the least squares solution is to use *calculus*, because that is how to find the minimum of a quadratic function.

Example 8.10. Using the equation $A\mathbf{x} = \mathbf{b}$ fom Example 8.7, we have

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \left\| \begin{bmatrix} 1 & 1 \\ 7 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} \right\| = \left\| \begin{bmatrix} a+b \\ 7a+b \\ 5a+b \end{bmatrix} - \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} \right\| = \left\| \begin{bmatrix} a+b-4 \\ 7a+b-1 \\ 5a+b-3 \end{bmatrix} \right\|,$$

which simplifies to

$$M(a, b) = (a + b - 4)^2 + (7a + b - 1)^2 + (5a + b - 3)^2. \tag{4}$$

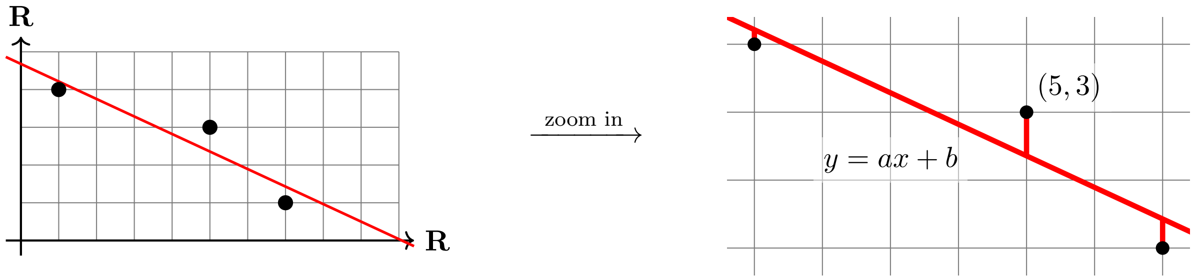
To find its minimum, we take the derivative. Since this is a function in two variables, we have two derivatives to take.

$$\begin{aligned} \frac{\partial M}{\partial a} &= 2(a + b - 4) + 2(7a + b - 1)(7) + 2(5a + b - 3)(5) = 150a + 26b - 52 \\ \frac{\partial M}{\partial b} &= 2(a + b - 4) + 2(7a + b - 1) + 2(5a + b - 3) = 26a + 6b - 16 \end{aligned}$$

Having these derivatives be zero produces a new matrix equation to solve:

$$\begin{bmatrix} 150 & 26 \\ 26 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 52 \\ 16 \end{bmatrix} \quad : \qquad \begin{bmatrix} 150 & 26 & 52 \\ 26 & 6 & 16 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & -\frac{13}{28} \\ 0 & 1 & \frac{131}{28} \end{bmatrix}$$

We now see the line $y = -\frac{13}{28}x + \frac{131}{28}$ is the best approximation:



The vertical distances from the points to the line have been minimized. Indeed, for example with $(5, 3)$, minimizing the vertical distance between it and the line $y = ax + b$ means making the value

$$\|(5, 5a + b) - (5, 3)\|^2 = \|(5 - 5, 5a + b - 3)\|^2 = (5 - 5)^2 + (5a + b - 3)^2 = (5a + b - 3)^2,$$

which is exactly the third term in $M(a, b)$ from Equation (4).

Remark 8.11. The "distance" from a point to the line can be throught of as the shortest length - not always the vertical distance. This is sometimes called the *perpendicular* distance, and will be solved by the method presented later in Lecture 17.

8.3 Exercises

Exercise 8.1. Show that projecting twice onto a line is the same as projecting once. That is, if P is the projection matrix from Equation (3), show that $P^2 = P$.

Exercise 8.2. Let $\mathbf{v} = (1, 1, 1) \in \mathbf{R}^3$.

$\bowtie$ 1. Take random vectors in the unit square in $\mathbf{R}^3$, and plot the average error, up until 1000 vectors, when projecting to $\mathbf{v}$.

2. What does this number converge to?

3. **Bonus:** Prove this limit.

Exercise 8.3. Find the projection of $\mathbf{v} = (-3, -1, 6)$ onto the plane $3x + 4y - 9z = 0$ and its normal vector.

Exercise 8.4. Let $A \in \mathcal{M}_{m \times n}$. Show that A and $A^T A$ have the same nullspace.

Exercise 8.5. Using the setup from Example 8.7, finished in Example 8.10, to come to the same conclusion (that is, the same best fit linear equation), but use the projection matrix instead of partial derivatives.

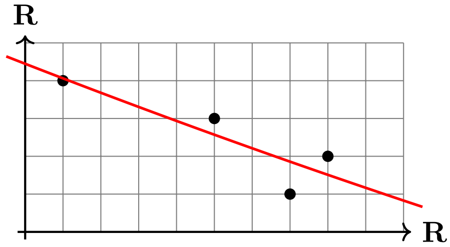
Lecture 9: The Gram–Schmidt process

In this lecture we finish up the idea of least squares from Lecture 8, by improving it with polynomial solutions, instead of just linear solutions.

9.1 Least squares for polynomials

Recall Example 8.7 from Lecture 8, which asked for a line of best fit to three points. The equation of a line is $y = ax + b$, and we found the appropriate a and b. What if we wanted to be more accurate, and find a quadratic function that goes through these three points? Quadratics have the form $y = ax^2 + bx + c$.

Example 9.1. Three points always have a unique quadratic going through them (which can be found by back-substitution), so we add another point $(8, 2)$ for increased difficulty.



$$\begin{aligned} 4 &= a+b+c \\ 1 &= 49a+7b+c \\ 3 &= 25a+5b+c \\ 2 &= 64a+8b+c \end{aligned} \qquad \underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 49 & 7 & 1 \\ 25 & 5 & 1 \\ 64 & 8 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 4 \\ 1 \\ 3 \\ 2 \end{bmatrix}}_{\mathbf{b}}$$

The process then is very similar, except we have three variables:

$$M(a,b,c) = \|A\mathbf{x} - \mathbf{b}\|^2 = (a+b+c-4)^2 + (49a+7b+c-1)^2 + (25a+5b+c-3)^2 + (64a+8b+c-2)^2.$$

Taking the derivative in all three variables gives

$$\begin{aligned} \frac{\partial M}{\partial a} &= 14246a + 1962b + 278c - 512, \\ \frac{\partial M}{\partial b} &= 1962a + 278b + 42c - 84, \\ \frac{\partial M}{\partial c} &= 278a + 42b + 8c - 20, \end{aligned}$$

which, when placed into a system, leads to the solutions $a = \frac{1}{372}$, $b = -\frac{241}{620}$, $c = \frac{2068}{465}$, as shown in the plot above.

Definition 9.2. Let $\mathbf{v}_1 = (x_1, y_1), \dots, \mathbf{v}_n = (x_n, y_n) \in \mathbf{R}^2$. The degree-$k$ polynomial $a_0 + a_1x + a_2x^2 + \cdots + a_kx^k$ that approximates the points $\mathbf{v}_i$ is the least squares solution to the matrix equation

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

The matrix on the left is called the *Vandermonde* matrix. This is the same as we used before, but with rows rearranged (the solution will be the same).

9.2 Orthonormalizing a basis

We previously saw orthogonality and orthonormality in Section 7. We revisit it here from the perspective of bases.

Recall that for a set of vectors $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ to be *orthonormal*, they need to be orthogonal (that is, $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$), and they need to be of unit length (that is $\|\mathbf{v}_i\| = 1$ for all i).

Remark 9.3. Placing orthonormal vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbf{R}^n$ as columns in a matrix Q will always give $Q^TQ = I$.

Example 9.4. We have already seen the *rotation* matrix $R_\theta := \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ from Example 7.4 in Lecture 7 has orthonormal columns:

$$R_\theta^T R_\theta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Every single *permutation* matrix also has orthogonal columns:

$$\underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{P^T} \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Remark 9.5. Whenever $Q \in \mathcal{M}_{m\times n}$ has orthonormal columns, the lengths of $\mathbf{v}$ and $Q\mathbf{v}$ are the same, for any $\mathbf{v} \in \mathbf{R}^n$. This follows directly from Remark 9.3:

$$\|Q\mathbf{v}\|^2 = (Q\mathbf{v}) \cdot (Q\mathbf{v}) = (Q\mathbf{v})^T(Q\mathbf{v}) = (\mathbf{v}^TQ^T)Q\mathbf{v} = \mathbf{v}^T(Q^TQ)\mathbf{v} = \mathbf{v}^TI\mathbf{v} = \mathbf{v}^T\mathbf{v} = \mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2.$$

Remark 9.6. Suppose we have a subspace V with basis vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ as columns in A, but the basis vectors are orthnormal. Using Definiton 8.5, we get that the projection matrix onto V simplifies to $A(A^TA)^{-1}A^T = AI^{-1}A^T = AIA^T = AA^T$. That is, for any $\mathbf{w} \in \mathbf{R}^n$, the projection onto V is

$$\text{proj}_V(\mathbf{w}) = AA^T\mathbf{w} = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_k \\ | & & | \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_1 & - \\ & \vdots & \\ - & \mathbf{v}_k & - \end{bmatrix} \mathbf{w} = (\mathbf{v}_1^T\mathbf{w})\mathbf{v}_1 + (\mathbf{v}_2^T\mathbf{w})\mathbf{v}_2 + \cdots + (\mathbf{v}_k^T\mathbf{w})\mathbf{v}_k.$$

We are considering all the impacts of having an orthonormal basis, because a very helpful simplification to many problems is to have an orthnormal basis. The basis you are given may not be orthonormal, so you have to *orthnormalize* it. This process of making the basis orthonormal is the *Gram–Schmidt process*.

Example 9.7. Consider the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 \in \mathbf{R}^4$, placed as columns in the matrix

$$\begin{bmatrix} 1 & 2 & 0 & 2 \\ 2 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

These vectors form a basis, but the basis is clearly not orthnormal. If it were, the computations below should give values 1 on the diagonal and 0 everywhere else:

$$\begin{array}{cccc} \mathbf{v}_1^T\mathbf{v}_1 = 6 & \mathbf{v}_1^T\mathbf{v}_2 = 2 & \mathbf{v}_1^T\mathbf{v}_3 = 3 & \mathbf{v}_1^T\mathbf{v}_4 = 4 \\ & \mathbf{v}_2^T\mathbf{v}_2 = 8 & \mathbf{v}_2^T\mathbf{v}_3 = 2 & \mathbf{v}_2^T\mathbf{v}_4 = 4 \\ & & \mathbf{v}_3^T\mathbf{v}_3 = 3 & \mathbf{v}_3^T\mathbf{v}_4 = 4 \\ & & & \mathbf{v}_4^T\mathbf{v}_4 = 10 \end{array}$$

The Gram-Schmidt process works by first creating a set $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4$ of orthogonal vectors (making the off-diagonal values 0). Then dividing each vector by its length, called *normalizing*, will give a set of orthnormal vectors $\mathbf{q}_1 = \mathbf{w}_1/\|\mathbf{w}_1\|, \mathbf{q}_2 = \mathbf{w}_2/\|\mathbf{w}_2\|, \mathbf{q}_3 = \mathbf{v}_3/\|\mathbf{w}_3\|, \mathbf{q}_4 = \mathbf{v}_4/\|\mathbf{w}_4\|$. This will make the diagonal values 1 in the dot product computations above.

- **Step 1:** Set $\mathbf{w}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$.

- **Step 2:** Project $\mathbf{v}_2$ onto $\mathbf{w}_1$, and subtract this from $\mathbf{v}_2$ to ensure the new vector will be orthogonal to the previous vector. That is, set $\mathbf{w}_2$ to be the error vector when projecting to $\mathbf{w}_1$. Using the formula from Definition 8.2, we get

$$\mathbf{w}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{w}_1}(\mathbf{v}_2) = \mathbf{v}_2 - \frac{\mathbf{w}_1^T\mathbf{v}_2}{\mathbf{w}_1^T\mathbf{w}_1}\mathbf{w}_1 = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 0 \end{bmatrix} - \frac{2}{6}\begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{5}{3} \\ -\frac{2}{3} \\ 2 \\ -\frac{1}{3} \end{bmatrix}$$

- **Step 3:** Project $\mathbf{v}_3$ onto $\mathbf{w}_1$ and $\mathbf{w}_2$, and subtract these from $\mathbf{v}_3$ to make sure everything is still orthogonal. The formula is

$$\mathbf{w}_3 = \mathbf{v}_3 - \text{proj}_{\mathbf{w}_1}(\mathbf{v}_3) - \text{proj}_{\mathbf{w}_2}(\mathbf{v}_3) = \mathbf{v}_3 - \frac{\mathbf{w}_1^T\mathbf{v}_3}{\mathbf{w}_1^T\mathbf{w}_1}\mathbf{w}_1 - \frac{\mathbf{w}_2^T\mathbf{v}_3}{\mathbf{w}_2^T\mathbf{w}_2}\mathbf{w}_2.$$

- **Step 4:** Repeat the same for $\mathbf{v}_4$ to get

$$\mathbf{w}_4 = \mathbf{v}_4 - \text{proj}_{\mathbf{w}_1}(\mathbf{v}_4) - \text{proj}_{\mathbf{w}_2}(\mathbf{v}_4) - \text{proj}_{\mathbf{w}_3}(\mathbf{v}_4) = \mathbf{v}_4 - \frac{\mathbf{w}_1^T\mathbf{v}_4}{\mathbf{w}_1^T\mathbf{w}_1}\mathbf{w}_1 - \frac{\mathbf{w}_2^T\mathbf{v}_4}{\mathbf{w}_2^T\mathbf{w}_2}\mathbf{w}_2 - \frac{\mathbf{w}_3^T\mathbf{v}_4}{\mathbf{w}_3^T\mathbf{w}_3}\mathbf{w}_3.$$

  We now have an orthogonal basis of vectors $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ for $\mathbf{R}^4$. Note these do not (except for the first one) point in the same directions as the original set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.

Example 9.7 showed how to get from one basis to another. Placing the original vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ as columns in a matrix A, and placing the resulting orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4$ as columns in a matrix A, a natural question arises: How are A and Q related?

Proposition 9.8. There exists a matrix R for which $A = QR$, or $R = Q^T A$, and it is given by

$$R = \begin{bmatrix} - & \mathbf{q}_1 & - \\ - & \mathbf{q}_2 & - \\ - & \mathbf{q}_3 & - \\ - & \mathbf{q}_4 & - \end{bmatrix} \begin{bmatrix} | & | & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1^T\mathbf{v}_1 & \mathbf{q}_1^T\mathbf{v}_2 & \mathbf{q}_1^T\mathbf{v}_3 & \mathbf{q}_1^T\mathbf{v}_4 \\ 0 & \mathbf{q}_2^T\mathbf{v}_2 & \mathbf{q}_2^T\mathbf{v}_3 & \mathbf{q}_2^T\mathbf{v}_4 \\ 0 & 0 & \mathbf{q}_3^T\mathbf{v}_3 & \mathbf{q}_3^T\mathbf{v}_4 \\ 0 & 0 & 0 & \mathbf{q}_4^T\mathbf{v}_4 \end{bmatrix}.$$

The proof of this statement follows immediately by observing that the construction of the $\mathbf{q}_i$ meant that $\mathbf{q}_i \cdot \mathbf{v}_j = 0$ whenever $j < i$. Indeed, we first note that $\mathbf{q}_i \cdot \mathbf{w}_j = 0$ whenever $i \neq j$, since the $\mathbf{q}_i$ point in the same direction as the $\mathbf{w}_i$. So for example,

$$\mathbf{q}_4 \cdot \mathbf{v}_3 = \mathbf{q}_4 \cdot \big(\mathbf{w}_3 + \text{proj}_{\mathbf{w}_1}(\mathbf{v}_3) + \text{proj}_{\mathbf{w}_2}(\mathbf{v}_3)\big) = \underbrace{\mathbf{q}_4 \cdot \mathbf{w}_3}_{0} + \underbrace{\mathbf{q}_4 \cdot \text{proj}_{\mathbf{w}_1}(\mathbf{v}_3)}_{0 \text{ because } \mathbf{q}_4 \cdot \mathbf{w}_1 = 0} + \underbrace{\mathbf{q}_4 \cdot \text{proj}_{\mathbf{w}_2}(\mathbf{v}_3)}_{0 \text{ because } \mathbf{q}_4 \cdot \mathbf{w}_2 = 0} = 0.$$

Remark 9.9. Recall that to find the least squares solution to $A\mathbf{x} = \mathbf{b}$, we projected $\mathbf{b}$ onto col(A) as $\mathbf{p}$. Since

$$A\mathbf{x} = \mathbf{b} = \underbrace{\mathbf{p}}_{\text{in col}(A)} + \underbrace{\mathbf{e}}_{\text{orthogonal to col}(A)}$$

has no solution, but

$$A^T A\mathbf{x} = A^T\mathbf{b} = \underbrace{A^T\mathbf{p}}_{\text{in col}(A^TA)} + \underbrace{A^T\mathbf{e}}_{0}$$

does, least squares was about solving $A^T A\mathbf{x} = A^T\mathbf{b}$. Using the result from Proposition 9.8, this equation becomes

$$\begin{aligned}
A^T A\mathbf{x} &= A^T\mathbf{b} \\
(QR)^T(QR)\mathbf{x} &= (QR)^T\mathbf{b} \\
R^TQ^TQR\mathbf{x} &= R^TQ^T\mathbf{b} \\
R^TR\mathbf{x} &= R^TQ^T\mathbf{b} && \text{(since } Q^TQ = I\text{)} \\
R\mathbf{x} &= Q^T\mathbf{b} && \text{(since } R \text{ and } R^T \text{ have inverses)} \\
\mathbf{x} &= R^{-1}Q^T\mathbf{b} && \text{(since } R \text{ has an inverse)}
\end{aligned}$$

which requires much less multiplications for a computer to do that $\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}$.

9.3 Exercises

Exercise 9.1. Find the least squares degree 1,2,3,4 polynomials that approximate the points

$$(-7, 2),\ (-6, -2),\ (-2, -1),\ (0, 3),\ (3, 0),\ (4, 1).$$

Plot all the functions and points together to confirm that the higher degree polynomials are better approximations to the points.

Exercise 9.2. Check that the columns of the 2×2 rotation matrix (introduced in Lecture 7.1) and of the 3×3 permutation matrices (introduced in Lecture 2.2) are all orthogonal. Are they orthonormal?

Lecture 10: Inner products and distances

We now take a small detour from Strang's *Linear Algebra* and work with the material from Strang's *Learning from Data*, Chapter IV.10. Similar topics are convered in Lay's *Linear Algebra and its applications*, Section 6.7.

10.1 Functions on spaces

Definition 10.1. Let V be a vector space. An *inner product* on V is a function $\langle\ \cdot\ ,\ \cdot\ \rangle\colon V^2 \to \mathbf{R}$ such that for all $\mathbf{v}, \mathbf{u}, \mathbf{w} \in V$ and all $c \in \mathbf{R}$,

- (positive definite) $\langle \mathbf{v}, \mathbf{v}\rangle \geqslant 0$ with $\langle \mathbf{v}, \mathbf{v}\rangle = 0$ if and only if $\mathbf{v} = 0$
- (symmetric) $\langle \mathbf{v}, \mathbf{u}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$
- (multiplicative) $\langle c\mathbf{v}, \mathbf{u}\rangle = c\langle \mathbf{v}, \mathbf{u}\rangle = \langle \mathbf{v}, c\mathbf{u}\rangle$
- (blinear) $\langle \mathbf{v}+\mathbf{u}, \mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{w}\rangle + \langle \mathbf{u}, \mathbf{w}\rangle$

A vector space closed under an inner product is called an *inner product space*.

We have already seen an example of the inner product before in Lecture 1, Definition 1.7, where the *dot product* of two vectors was introduced. Just like there, every inner product has a notion of distance associated to it: the *norm*, or *length*, of $\mathbf{v}$ in an inner product space V is $\sqrt{\langle \mathbf{v}, \mathbf{v}\rangle} = \|\mathbf{v}\|$.

Example 10.2. There are many examples of inner product spaces besides $\mathbf{R}^n$ with the dot product.

- The space $\mathcal{M}_{m\times n}$ of all $m \times n$ matrices over $\mathbf{R}$ is an inner product space when using $\langle A, B\rangle :=$ trace(A^TB). The *trace* is the sum of the entries on the diagonal.
- The space $C[0,1]$ of all continuous functions with domain $[0,1]$ and inner product

  $$\langle f, g\rangle := \int_0^1 f(x)g(x)\ dx$$

  is an inner product space. Adjusting the domain to any interval $[a,b] \subseteq \mathbf{R}$ still makes this an inner product space.

Theorem 10.3. *The inner product $\langle\ \cdot\ ,\ \cdot\ \rangle$ in any inner product space $V \ni \boldsymbol{v}, \boldsymbol{w}$ satisfies:*

- $\langle \boldsymbol{v}, \boldsymbol{w}\rangle \leqslant \|\boldsymbol{v}\|\|\boldsymbol{w}\|$ *with equality iff $\boldsymbol{v}$ and $\boldsymbol{w}$ are linearly dependent*
- $\|\boldsymbol{v}+\boldsymbol{w}\| \leqslant \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$

Example 10.4. Using the first point of Theorem 10.3, we can show that the functions $\sin(x)$ and $\cos(x)$ are linearly independent in $C[0, 2\pi]$. We find that

$$\begin{aligned}
\langle \sin(x), \cos(x)\rangle &= \int_0^{2\pi} \sin(x)\cos(x)\ dx = \int_0^{2\pi} \frac{\sin(2x)}{2}\ dx = \frac{-\cos(4\pi)}{4} - \frac{-\cos(0)}{4} = 0, \\
\|\sin(x)\|^2 &= \int_0^{2\pi} \sin^2(x)\ dx = \int_0^{2\pi} \frac{1-\cos(2x)}{2}\ dx = \pi - \left(\frac{\sin(4\pi)}{4} - \frac{\sin(0)}{4}\right) = \pi, \\
\|\cos(x)\|^2 &= \int_0^{2\pi} \cos^2(x)\ dx = \int_0^{2\pi} \frac{\cos(2x)+1}{2}\ dx = \left(\frac{\sin(4\pi)}{4} - \frac{\sin(0)}{4}\right) + \pi = \pi.
\end{aligned}$$

Since $0 \neq \sqrt{\pi}\sqrt{\pi} = \pi$, these functions are linearly independent. Also note that the positive definite property of the inner product is satisfied.

The notions of angle between vectors, orthogonality, unit length, all apply to inner product spaces in the same way they applied to $\mathbf{R}^n$ with the dot product.

Example 10.5. The angle between the matrices $A = \begin{bmatrix} 4 & 1 \\ -1 & 0 \\ 7 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 2 \\ 3 & -1 \\ 2 & 0 \end{bmatrix}$ is

$$\begin{aligned} \cos^{-1}\left(\frac{\mathrm{trace}(A^TB)}{\mathrm{trace}(A^TA)\mathrm{trace}(B^TB)}\right) &= \cos^{-1}\left(\frac{\mathrm{trace}\left(\begin{bmatrix} 4 & -1 & 7 \\ 1 & 0 & 2 \end{bmatrix}\begin{bmatrix} 0 & 2 \\ 3 & -1 \\ 2 & 0 \end{bmatrix}\right)}{\mathrm{trace}\left(\begin{bmatrix} 4 & -1 & 7 \\ 1 & 0 & 2 \end{bmatrix}\begin{bmatrix} 4 & 1 \\ -1 & 0 \\ 7 & 2 \end{bmatrix}\right)\mathrm{trace}\left(\begin{bmatrix} 0 & 3 & 2 \\ 2 & -1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 2 \\ 3 & -1 \\ 2 & 0 \end{bmatrix}\right)}\right) \\ &= \cos^{-1}\left(\frac{\mathrm{trace}\left(\begin{bmatrix} 9 & 9 \\ 4 & 2 \end{bmatrix}\right)}{\mathrm{trace}\left(\begin{bmatrix} 66 & 18 \\ 18 & 5 \end{bmatrix}\right)\mathrm{trace}\left(\begin{bmatrix} 13 & -3 \\ -3 & 5 \end{bmatrix}\right)}\right) \\ &= \cos^{-1}\left(\frac{11}{1278}\right) \\ &\approx 89.51^\circ \end{aligned}$$

Remark 10.6. The Gram–Schmidt process in Lecture 9 was done on vectors using the usual norm in $\mathbf{R}^n$. By obeserving that the projection operation can be given in terms of inner product, the Gram–Schmidt process can be applied to any inner product space:

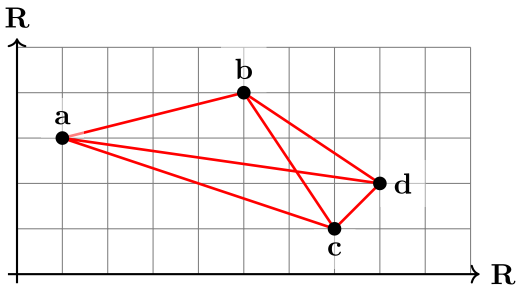
$$\mathrm{proj}_{\mathbf{v}}(\mathbf{u}) = \frac{\mathbf{v}^T\mathbf{u}}{\mathbf{v}^T\mathbf{v}}\mathbf{v} = \frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v} = \frac{\langle \mathbf{v},\mathbf{u}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}.$$

10.2 Distance matrices

Recall the points from Exercise 9.1 in Lecture 9. If the points were located elsewhere but their *relative* position to each other was the same, we could still solve the least squares problem, up to some x-shift and y-shift. This situation has two advantages:

- *only requires relative information*: measurements only need to be made among the data, not between data and something else (like a reference point - the origin)
- *allows for spaces that are not* $\mathbf{R}^n$: on the sphere, on a grid, with barriers, etc

Example 10.7. Consider the distances among the four points, slightly adapted from Exercise 9.1.



$$\begin{aligned} \mathbf{a} &= (1,3) \\ \mathbf{b} &= (5,4) \\ \mathbf{c} &= (7,1) \\ \mathbf{d} &= (8,2) \end{aligned} \qquad X = \begin{bmatrix} 1 & 5 & 7 & 8 \\ 3 & 4 & 1 & 2 \end{bmatrix}$$

The matrix X is called the *position matrix*. We can easily compute the *distance matrix* D among these points. But the relationship among the two is not so clear.

$$D = \begin{bmatrix} \|\mathbf{a}-\mathbf{a}\| & \|\mathbf{a}-\mathbf{b}\| & \|\mathbf{a}-\mathbf{c}\| & \|\mathbf{a}-\mathbf{d}\| \\ & \|\mathbf{b}-\mathbf{b}\| & \|\mathbf{b}-\mathbf{c}\| & \|\mathbf{b}-\mathbf{d}\| \\ & & \|\mathbf{c}-\mathbf{c}\| & \|\mathbf{c}-\mathbf{d}\| \\ & & & \|\mathbf{d}-\mathbf{d}\| \end{bmatrix} = \begin{bmatrix} 0 & \sqrt{17} & \sqrt{40} & \sqrt{50} \\ & 0 & \sqrt{13} & \sqrt{13} \\ & & 0 & \sqrt{2} \\ & & & 0 \end{bmatrix}$$

The key lies in fixing one of the points as a reference point. Without loss of generality, we simply say $\mathbf{a} = 0$. Then the first line of D becomes the lengths $\|\cdot\|$ of all the vectors involved. Notice that the lengths of the vectors are on the diagonal of X^TX, so we can create a column vector from the *diagonal* of the matrix X^TX, which will be $\text{diag}(X^TX) = [0\ \sqrt{17}\ \sqrt{40}\ \sqrt{50}]$. In particular, we have

$$X^TX = \frac{1}{2}\left(\text{diag}(X^TX)\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\text{diag}(X^TX)^T - D\right). \tag{5}$$

Similar to how we have seen matrix decompositions $A = LU$ and $A = QR$, there is a *Cholesky decompositon* $A = U^TU$, which will give X, up to a shift. For this example, we have

$$\begin{aligned}
X^TX &= \frac{1}{2}\left(\begin{bmatrix}0\\\sqrt{17}\\\sqrt{40}\\\sqrt{50}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}0 & \sqrt{17} & \sqrt{40} & \sqrt{40}\end{bmatrix} - \begin{bmatrix}0 & \sqrt{17} & \sqrt{40} & \sqrt{50}\\\sqrt{17} & 0 & \sqrt{13} & \sqrt{13}\\\sqrt{40} & \sqrt{13} & 0 & \sqrt{2}\\\sqrt{50} & \sqrt{13} & \sqrt{2} & 0\end{bmatrix}\right)\\
&= \frac{1}{2}\left(\begin{bmatrix}0 & 0 & 0 & 0\\\sqrt{17} & \sqrt{17} & \sqrt{17} & \sqrt{17}\\\sqrt{40} & \sqrt{40} & \sqrt{40} & \sqrt{40}\\\sqrt{50} & \sqrt{50} & \sqrt{50} & \sqrt{50}\end{bmatrix} + \begin{bmatrix}0 & \sqrt{17} & \sqrt{40} & \sqrt{50}\\0 & \sqrt{17} & \sqrt{40} & \sqrt{50}\\0 & \sqrt{17} & \sqrt{40} & \sqrt{50}\\0 & \sqrt{17} & \sqrt{40} & \sqrt{50}\end{bmatrix} - \begin{bmatrix}0 & \sqrt{17} & \sqrt{40} & \sqrt{50}\\\sqrt{17} & 0 & \sqrt{13} & \sqrt{13}\\\sqrt{40} & \sqrt{13} & 0 & \sqrt{2}\\\sqrt{50} & \sqrt{13} & \sqrt{2} & 0\end{bmatrix}\right).
\end{aligned}$$

We do not finish the computation, but in the end we will recover X as

$$X' = \begin{bmatrix}0 & 4 & 6 & 7\\0 & 1 & -2 & -1\end{bmatrix}.$$

Remark 10.8. If instead we have a set of vectors $\mathbf{v}_i, \dots, \mathbf{v}_k$, then the distance matrix would be defined as $D_{ij} = \|\mathbf{v}_i - \mathbf{v}_j\|$. Note that this means the distance matrix is always symmetric and has a zero diagonal.

Example 10.9. If D is simply symmetric and has a zero diagonal, there is no guarantee that is represents distance among points in a space like $\mathbf{R}^n$, or even any inner product space. Consider the distance matrix

$$D = \begin{bmatrix}0 & 1 & 1 & 1\\1 & 0 & 3 & 3\\1 & 3 & 0 & 3\\1 & 3 & 3 & 0\end{bmatrix},$$

coming from four points $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$. As in the previous example, we let the first point $\mathbf{a} = 0$, so that we get $\|\mathbf{b}\| = \|\mathbf{c}\| = \|\mathbf{d}\| = 1$. We also see that

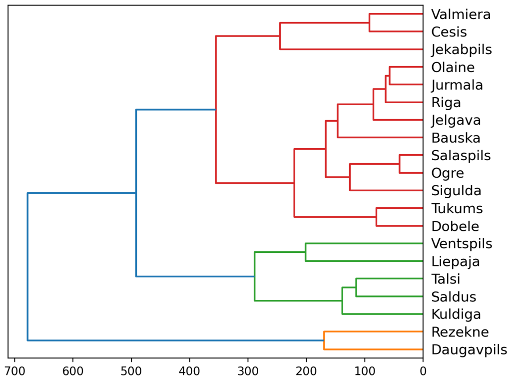
$$\begin{aligned}
3^2 &= \|\mathbf{b} - \mathbf{c}\|^2\\
&= \langle \mathbf{b} - \mathbf{c}, \mathbf{b} - \mathbf{c}\rangle\\
&= \langle \mathbf{b}, \mathbf{b} - \mathbf{c}\rangle - \langle \mathbf{c}, \mathbf{b} - \mathbf{c}\rangle\\
&= \langle \mathbf{b}, \mathbf{b}\rangle - \langle \mathbf{b}, \mathbf{c}\rangle - \langle \mathbf{c}, \mathbf{b}\rangle + \langle \mathbf{c}, \mathbf{c}\rangle\\
&= \|\mathbf{b}\|^2 - 2\langle \mathbf{b}, \mathbf{c}\rangle + \|\mathbf{c}\|^2\\
&= 1 - 2\langle \mathbf{b}, \mathbf{c}\rangle + 1.
\end{aligned}$$

Rearranging, we conclude that $\langle \mathbf{b}, \mathbf{c}\rangle = -7/2$, which contradicts the fact that the inner product must be positive definite. Hence D can not be a distance matrix of points from an inner product space.

Distance matrices can highlight *clustering* among the data. That is, given a distance matrix, we can "connect" points that lie close to each other and so discover which groups of points are close to each other.

Example 10.10. Consider the distances between the 20 largest cities in Latvia, in kilometers. As a distance matrix, it is difficult to get information from it, but we can group cities by distance into clusters. This could be useful, for example, in trying to decide where to build a factory or distribution center.

Bauska	0	119	152	62	103	37	68	149	196	53	46	194	62	54	111	91	136	89	145	191
Cesis	119	0	174	137	96	115	93	200	271	66	97	155	79	69	180	30	159	130	27	220
Daugavpils	152	174	0	214	79	186	203	301	344	156	185	89	187	166	263	170	281	235	194	340
Dobele	62	137	214	0	158	29	47	87	141	81	41	249	59	72	51	108	78	34	158	129
Jekabpils	103	96	79	158	0	129	134	242	297	85	122	91	117	96	208	92	215	171	116	276
Jelgava	37	115	186	29	129	0	38	115	170	54	19	220	40	47	80	86	98	51	139	154
Jurmala	68	93	203	47	134	38	0	111	179	49	22	221	18	39	88	65	80	39	112	142
Kuldiga	149	200	301	87	242	115	111	0	79	159	120	331	129	149	42	175	48	72	214	48
Liepaja	196	271	344	141	297	170	179	79	0	221	181	389	195	212	92	244	127	141	288	98
Ogre	53	66	156	81	85	54	49	159	221	0	40	172	32	11	130	38	129	87	91	191
Olaine	46	97	185	41	122	19	22	120	181	40	0	211	21	31	90	68	96	50	120	156
Rezekne	194	155	89	249	91	220	221	331	389	172	211	0	203	183	299	166	300	259	163	362
Riga	62	79	187	59	117	40	18	129	195	32	21	203	0	21	103	50	98	57	100	159
Salaspils	54	69	166	72	96	47	39	149	212	11	31	183	21	0	120	40	119	77	94	180
Saldus	111	180	263	51	208	80	88	42	92	130	90	299	103	120	0	153	61	52	198	89
Sigulda	91	30	170	108	92	86	65	175	244	38	68	166	50	40	153	0	137	104	54	200
Talsi	136	159	281	78	215	98	80	48	127	129	96	300	98	119	61	137	0	47	170	62
Tukums	89	130	235	34	171	51	39	72	141	87	50	259	57	77	52	104	47	0	147	105
Valmiera	145	27	194	158	116	139	112	214	288	91	120	163	100	94	198	54	170	147	0	229
Ventspils	191	220	340	129	276	154	142	48	98	191	156	362	159	180	89	200	62	105	229	0



To get the *dendrogram* above, each city begins in its own cluster. The two closest cities are connected to create one cluster of 2 cities (Ogre and Salaspils). Create larger clusters by measuring the distance between every pair of clusters c_i and c_j, with distance defined to be

$$\text{(distance between } c_i \text{ and } c_j) = \frac{1}{|c_i||c_j|} \sum_{\mathbf{v}_i \in c_i} \sum_{\mathbf{v}_j \in c_j} \|\mathbf{v}_i - \mathbf{v}_j\|.$$

For clusters of size 1, note that $|c_i| = |c_j| = 1$, and the distance reduces to the usual distance. This is the *average* method of drawing a dendrogram. In the diagram above, the last 3 clusters to be joined are colored differently, but any number can be chosen here.

10.3 Exercises

Exercise 10.1. For each of the following "definitions", show that each cannot be an inner product.

- For $A, B \in \mathcal{M}_{n\times n}$, let $\langle A, B\rangle = \text{trace}(A + B)$
- For $f, g \in C[0, 1]$, let $\langle f, g\rangle = \left|\frac{df}{dx}\frac{dg}{dx}\right|$
- For $a, b \in \mathbf{R}$, let $\langle a, b\rangle = a^2 + b^2$

Exercise 10.2. Check the conditions for the space of $m \times n$ matrices over $\mathbf{R}$ from Example 10.2 being an inner product space. What is the distance between $\left[\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \end{smallmatrix}\right]$?

Exercise 10.3. Consider the following three matrices in $\mathcal{M}_{2\times 2}$:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & -3 \\ 3 & 2 \end{bmatrix}.$$

Using the Gram–Schmidt process to find an orthonormal basis for span$\{A, B, C\}$. Use the inner product on matrices given in Example 10.2.

Exercise 10.4. Given the distance D matrix below, construct the dendrogram using the same average distance method as in Example 10.10. After every step, give the new distance matrix, which measures the distances among the clusters.

$$D = \begin{bmatrix} 0 & 12 & 10 & 13 & 2 & 11 \\ 12 & 0 & 3 & 9 & 13 & 8 \\ 10 & 3 & 0 & 6 & 14 & 5 \\ 13 & 9 & 6 & 0 & 15 & 1 \\ 2 & 13 & 14 & 15 & 0 & 7 \\ 11 & 8 & 5 & 1 & 7 & 0 \end{bmatrix}$$

Lecture 11: Determinants, part 1

This and the following lecture will deal with the *determinant*, which is a number associated to a matrix. This number has deep properties related to all the topics we have already seen.

11.1 The recursive definition

The *determinant* is a function $\det\colon \mathcal{M}_{n\times n} \to \mathbf{R}$. Before we get to definitions and new ideas, we look back at some topics we have seen that are related to the determinant.

Example 11.1. We have already come across the determinant in several disguises.

- The determinant of a 1×1 matrix $[a]$ is a. The matrix is not invertible is $a = 0$.
- The determinant of a 2×2 matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ is $ad - bc$. The matrix is not invertible if $ad - bc = 0$.
- (to be proved later) The determinant is the product of the pivots, up to a sign change.
- (to be proved later) The determinant is zero if and only if the matrix is not invertible.

These are all important facts, especially the last two. The determinant of a matrix is ofteen denoed by vertcal lines on the side. It has geometrical consequences, as the following example shows.

Example 11.2. Recall that a *parallelogram* is the shape made by the vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{a}+\mathbf{b}$. A *parallelotope* is the generalization of this shape to higher dimensions. As vectors, the sides of the parallelotope are $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ for $v_i \in \mathbf{R}^n$. As points, they are the vertices of the parallelotope. The n-dimensional volume of the parallelotope is the absolute value of

$$\begin{vmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{vmatrix}.$$

For example, the volume of the parallelogram made by the vectors $\mathbf{a} = (1,3)$ and $\mathbf{b} = (4,2)$ is the absolute value of

$$\begin{vmatrix} 1 & 4 \\ 3 & 2 \end{vmatrix} = 2 - 12 = -10,$$

so the area is 10.

We now give a precise definition of the determinant so that we can prove the previous and further strong claims.

Definition 11.3. Let $A \in \mathcal{M}_{n\times n}$. The *determinant* $\det(A)$ of A is:

- if $n = 1$, then $\det(A) = A_{11}$
- if $n \geqslant 2$, then $\det(A) = \displaystyle\sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det(A^{ij})$, for any $i \in \{1, \dots, n\}$

The matrix A^{ij} is the $(n-1)\times(n-1)$ *submatrix* of A produced when the ith row and jth column are removed. The number $\det(A^{ij})$ is called the ij-*minor* of A.

The determinant of a matrix is also denoted by vertical lines on the side, instead of the usual square brackets. The word *cofactor* often appears when talking about determinants: this is the ij-minor multiplied by $(-1)^{i+j}$. Then we get the *cofactor matrix* $\operatorname{cofac}(A)$ of A, defined as the $n\times n$ matrix for which $\operatorname{cofac}(A)_{ij} = (-1)^{i+j}\det(A^{ij})$.

Example 11.4. Let $A = \begin{bmatrix} 0 & 3 & 4 \\ 2 & -1 & 2 \\ 1 & 5 & -2 \end{bmatrix}$. Using Definition 11.3 and Example 11.1 with $i = 1$ we see that

$$\begin{aligned} \det(A) &= \begin{vmatrix} 0 & 3 & 4 \\ 2 & -1 & 2 \\ 1 & 5 & -2 \end{vmatrix} \\ &= (-1)^{1+1} 0 \begin{vmatrix} -1 & 2 \\ 5 & -2 \end{vmatrix} + (-1)^{1+2} 3 \begin{vmatrix} 2 & 2 \\ 1 & -2 \end{vmatrix} + (-1)^{1+3} 4 \begin{vmatrix} 2 & -1 \\ 1 & 5 \end{vmatrix} \\ &= 0 - 3(-4-2) + 4(10+1) = 18 + 44 = 62. \end{aligned}$$

We would have gotten the same result with $i = 2$ or $i = 3$.

Now we consider the determinant for some simple matrices.

Proposition 11.5. Let $A \in \mathcal{M}_{n \times n}$.

- If $A = I_n$, then $\det(A) = 1$.
- If A is upper (or lower) triangular, then $\det(A)$ is the product of the diagonal entries.

Proof. The first point follows by induction. For $n = 1$, we clearly have $\det([1]) = 1$. For larger n, use the recursive definition of the determinant, since the only nonzero entries of I are on the diagonal. We see that

$$\det(I_n) = \sum_{j=1}^{n} (-1)^{1+j} (I_n)_{1j} \det(I^{ij}) = (-1)^{1+1} (I_n)_{11} \det(I_{n-1}) = 1 \cdot 1 \cdot 1 = 1.$$

The second point follows similarly, using induction, but instead across the bottom row (top row for lower triangular). □

Example 11.6. Recall that elementary matrices are either *elimination* matrices or *permutation* matrices. The second point from Proposition 11.5 implies the determinant of an elimination matrix is always 1, since it is lower (or upper) triangular with 1's on the diagonal. For a permutation matrix, swapping two rows (one pair) gives a determinant of -1, but swapping three rows (two pairs) gives a determinant of 1:

$$\begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -1, \qquad \begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = 1.$$

Permutation matrices with an *odd* number of row swaps have determinnt -1, and permuattion matrices with an *even* number of row swaps have determinant 1.

Next we describe some more general properties of the determinant.

Proposition 11.7. Let $A \in \mathcal{M}_{n \times n}$. As a function of the rows of A, the determinant is:

- *multilinear*, that is, $\det(\mathbf{r}_1 \ldots, c\mathbf{a} + \mathbf{b}, \ldots, \mathbf{r}_n) = c \det(\mathbf{r}_1, \ldots, \mathbf{a}, \ldots, \mathbf{r}_n) + \det(\mathbf{b}, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_n)$
- *alternating*, that is, $\det(\mathbf{r}_1 \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_j, \ldots, \mathbf{r}_n) = -\det(\mathbf{r}_1 \ldots, \mathbf{r}_j, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_n)$

Proof. The first point follows by induction on n, and by using the recursive definition to expand along row i. The statement is immediately true for a 1×1 matrix. For the inductive step, notice that

$$\begin{aligned} \det A &= \det(\mathbf{r}_1, \ldots, c\mathbf{a} + \mathbf{b}, \ldots, \mathbf{r}_n) \\ &= \sum_{j=1}^{n} (-1)^{i+j} (c\mathbf{a} + \mathbf{b})_j \det(A^{ij}) \\ &= c \left(\sum_{j=1}^{n} (-1)^{i+j} (\mathbf{a})_j \det(A^{ij}) \right) + \left(\sum_{j=1}^{n} (-1)^{i+j} (\mathbf{b})_j \det(A^{ij}) \right), \end{aligned}$$

and A^{ij} is the same in both cases. We do not prove the second point. □

Remark 11.8. There are two immediate consequences of Proposition 11.7:

- a matrix with a zero row has determinant zero:

$$\det(\mathbf{r}_1\ldots,0,\ldots,\mathbf{r}_n)=\det(\mathbf{r}_1\ldots,\mathbf{a}-\mathbf{a},\ldots,\mathbf{r}_n)=\det(\mathbf{r}_1\ldots,\mathbf{a},\ldots,\mathbf{r}_n)-\det(\mathbf{r}_1\ldots,\mathbf{a},\ldots,\mathbf{r}_n)=0$$

- a matrix with two equal rows has determinant zero:

$$\underbrace{\det(\mathbf{r}_1\ldots,\mathbf{a},\ldots,\mathbf{a},\ldots,\mathbf{r}_n)}_{\text{original order}}=-\underbrace{\det(\mathbf{r}_1\ldots,\mathbf{a},\ldots,\mathbf{a},\ldots,\mathbf{r}_n)}_{\text{rows swapped}},$$

  and only 0 is its own negative.

11.2 Properties and applications

We finish this lecture by proving that the determinant is *multiplicative*, that is, that $\det(AB)=\det(A)\det(B)$ for any $n\times n$ matrices A,B.

Proposition 11.9. Let $A,B\in\mathcal{M}_{n\times n}$.

- If E is an elementary matrix and A is invertible, then $\det(EA)=\det(E)\det(A)=\det(AE)$.
- If D is a diagonal matrix and A is invertible, then $\det(DA)=\det(D)\det(A)=\det(AD)$.
- If A and B are invertible, then $\det(AB)=\det(A)\det(B)$.

Proof. The first point follows by multilinearity from Proposition 11.7 and Example 11.6, which gives that $\det(E)=\pm1$. Elimination matrices are row operations, so in terms of the rows $\mathbf{r}_1,\ldots,\mathbf{r}_n$ of A,

$$\begin{aligned}\det(EA)&=\det(\mathbf{r}_1,\ldots,\mathbf{r}_j-\ell_{ij}\mathbf{r}_j,\ldots,\mathbf{r}_n)\\&=\underbrace{\det(\mathbf{r}_1,\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_n)}_{\det(A)}-\ell_{ij}\underbrace{\det(\mathbf{r}_1,\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_n)}_{\text{0 because two rows the same}}\\&=\det(A)\\&=\det(A)\det(E).\end{aligned}$$

The eqution for permutation matrices follows from the alternating property of the determinant.

The second point follows by Propositon 11.5, which says that $\det(D)$ is the product of its diagonal entries, and by the definition of the determinant:

$$\det(DA)=\sum_{j=1}^{n}(-1)^{i+j}(DA)_{ij}\det((DA)^{ij})=D_{ii}\sum_{j=1}^{n}(-1)^{i+j}A_{ij}\det((DA)^{ij}).$$

Using induction we get that $\det(DA)=D_{11}D_{22}\cdots D_{nn}\det(A)=\det(D)\det(A)$. Commutativity of diagonal matrices with others gives us that $\det(DA)=\det(AD)$.

The third point follows by first noticing that the statement is true when $\det(A)=0$. Indeed, it must also be true that $\det(AB)=0$, since if $\det(AB)\neq0$, then AB is invertible, and so has an inverse C. But then $I=(AB)C=A(BC)$, and A has an inverse, which is a contradiciton. So if $\det(A)=0$, then

$$0=\det(A)\det(B)=\det(AB)=0,$$

and the statement holds. If $\det(A)\neq0$, then A is invertible, and may be expressed as

$$\underbrace{F_\ell\cdots F_1}_{\text{upper tri.}}\underbrace{E_k\cdots E_1}_{\text{lower tri.}}\underbrace{P}_{\text{permutation}}A=D,$$

where D is a diagonal matrix. This comes from the process of elimination. Recall that diagonal matrices commute with all other matrices, and we use this fact, and the first result, to get that

$$\begin{aligned}\det(AB) &= \det\left(P^{-1}E_1^{-1}\cdots E_k^{-1}F_1^{-1}\cdots F_\ell^{-1}DB\right)\\ &= \det(P^{-1})\det(E_1^{-1})\cdots\det(E_k^{-1})\det(F_1^{-1})\cdots\det(F_\ell^{-1})\det(DB)\\ &= \det(P^{-1})\det(E_1^{-1})\cdots\det(E_k^{-1})\det(F_1^{-1})\cdots\det(F_\ell^{-1})\det(D)\det(B)\\ &= \det\left(P^{-1}E_1^{-1}\cdots E_k^{-1}F_1^{-1}\cdots F_\ell^{-1}D\right)\det(B)\\ &= \det(A)\det(B).\end{aligned}$$

□

A direct consequence is that the determinant is the product of the pivots, up to a sign change, and that the determinant is zero iff the matrix is not invertible.

11.3 Exercises

Exercise 11.1. Show with a counter example that the set of all invertible $n \times n$ matrices is not a subspace of $\mathcal{M}_{n\times n}$. That is, show it is not a vector space.

Exercise 11.2. Recall the definition of an inverse of a matrix A, which was a matrix B such that $AB = BA = I$. Show that the statement $AB = I$ implies $BA = I$.

Exercise 11.3. How many cofactors, or minors, of the matrix below are nonzero? How many terms in the recursive definition of the determinant are nonzero?

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

Exercise 11.4. Let $A \in \mathcal{M}_{n\times n}$. Show that $\det(A) = 0$ is equivalent to saying that there is a nonzero vector $\mathbf{x}$ for which $A\mathbf{x} = 0$.

Lecture 12: Determinants, part 2

This lecture is the final lecture containing topics for the midterm. We finish off the first part of the semester with more properties of the determinant, some applications, and an alternative definition using combinatorics.

12.1 More properties and applications

First we describe how the determinant works with transposes and inverses.

Proposition 12.1. Let $A \in \mathcal{M}_{m\times n}$.

- $\det(A^T) = \det(A)$
- If A is invertible, then $\det(A^{-1}) = \det(A)^{-1}$

Proof. The first statement follows from $PA = LU$ decomposition and from Proposition 11.9. The second statement follows from Proposition 11.9 and the fact that $AA^{-1} = I$:

$$\begin{aligned} AA^{-1} = I \quad &\implies \quad \det(AA^{-1}) = \det(I) \\ &\implies \quad \det(A)\det(A^{-1}) = 1 \\ &\implies \quad \det(A^{-1}) = \frac{1}{\det(A)} = \det(A)^{-1}. \end{aligned}$$

□

Recall after Definition 11.3 the ij-*minor* of a matrix A was the determinant of the submatrix after the ith row and jth column are removed. The ij-*cofactor* was the ij-minor multiplied by $(-1)^{i+j}$.

Proposition 12.2. Let $A \in \mathcal{M}_{n\times n}$ be invertible, and let $C_{ij} = (-1)^{i+j}\det(A^{ij})$ be the ij-cofactor of A. Then the ij-entry in the inverse is

$$(A^{-1})_{ij} = \frac{C_{ji}}{\det(A)}.$$

In general, for C the cofactor matrix of A, we have $AC^T = \det(A)I$, or $A^{-1} = C^T/\det(A)$.

Proof. This comes from the recursive definition of the determinant, which states that

$$\begin{aligned} \det(A) &= A_{11}C_{11} + A_{12}C_{12} + \cdots + A_{1n}C_{1n} = \mathbf{A}_1^T\mathbf{C}_1, \\ \det(A) &= A_{21}C_{21} + A_{22}C_{22} + \cdots + A_{2n}C_{2n} = \mathbf{A}_2^T\mathbf{C}_2, \end{aligned}$$

and so on, where $\mathbf{A}_i$ is the ith row of A ad $\mathbf{C}_i$ is the ith row of C. Moreover, for $i \neq j$, the sum

$$\det(A') = A_{i1}C_{j1} + A_{i2}C_{j2} + \cdots + A_{in}C_{jn} = \mathbf{A}_i^T\mathbf{C}_j$$

of some new matrix A' must be zero, as this is the determinant for a matrix whose ith and jth rows are the same. That is, A_{j1} does not appear in C_{j1}, so having $A_{j1} = A_{i1}$ is allowed for this determinant. Remark 11.8 told us that a matrix with two equal rows has determinant zero. Hence

$$\begin{bmatrix} - & \mathbf{A}_1^T & - \\ - & \mathbf{A}_2^T & - \\ & \vdots & \\ - & \mathbf{A}_n^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{C}_1 & \mathbf{C}_2 & \cdots & \mathbf{C}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \det(A) & 0 & \cdots & 0 \\ 0 & \det(A) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \det(A) \end{bmatrix},$$

or $A^TC = \det(A)I$. This becomes $AC^T = \det(A)I$ because $C^T/\det(A)$ is the inverse of A, so it can be multiplied on the left or on the right. □

This formula generalizes the formula for the inverse of the 2×2 matrix $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ and $A^{-1} = \frac{1}{ad-bc}\left[\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right]$. The determinant is still in the denominator, but the cofactors come from larger matrices and so the inverse is not just about rearranging elements.

Example 12.3. Consider the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 5 & 0 \\ 0 & 0 & 2 & 8 & 0 \\ 7 & 2 & 9 & 3 & 6 \\ 0 & 0 & 0 & 3 & 1 \end{bmatrix}, \qquad \det(A) = -136.$$

This matrix is invertible, and the $(4,4)$-entry of the inverse will be

$$(A^{-1})_{44} = \frac{(-1)^{4+4}}{-136} \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = \frac{-1}{68}.$$

A final application of the determinant that we will see is in a physical setting. Recall the *standard basis* from Example 5.7 in Lecture 5.

Definition 12.4. Let $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1} \in \mathbf{R}^n$, arranged as columns of $A \in \mathcal{M}_{n\times(n-1)}$, and let $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbf{R}^n$ be the standard basis vectors. The *cross product* of the vectors $\mathbf{v}_i$ is the vector

$$X(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) := \sum_{i=1}^{n} (-1)^{i+n} \det(A^i)\mathbf{e}_i = \begin{vmatrix} | & | & & | & \mathbf{e}_1 \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{n-1} & \vdots \\ | & | & & | & \mathbf{e}_n \end{vmatrix},$$

where $A^i \in \mathcal{M}_{(n-1)\times(n-1)}$ is A with the ith row removed. The expression on the right is a formal determinant, since we can't put in a whole vector $\mathbf{e}_i$ in a single entry.

Example 12.5. What does the cross product represent? In three dimensions, it is the *right-hand rule* of physicists, determining the direction a moving charge from a rotating magnetic field. The vector computed will be perpendicular to the initial vectors:

$$\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = (-1)^{1+3} \begin{vmatrix} 3 & 0 \\ 4 & 1 \end{vmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + (-1)^{2+3} \begin{vmatrix} 2 & 1 \\ 4 & 1 \end{vmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + (-1)^{3+3} \begin{vmatrix} 2 & 1 \\ 3 & 0 \end{vmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix}.$$

Remark 12.6. The cross product has several interesting properties:

- $X(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) = 0$ iff the set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ is linearly dependent
- For $n = 2$, the length of the cross product is $\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\|\|\mathbf{v}\||\sin(\theta)|$
- The cross product is related to the dot product by $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u}$
- The cross product is *anti-commutative*, or *skew-symmetic*: $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$

12.2 A combinatorial definition

Recall the previous discussion that the determinant of a permutation matrix P depended on the *parity* of the permutation. We now define this concept in detail.

Definition 12.7. Let $S = (a_1, \ldots, a_n)$ be an ordered set. A *permutation* of S is either

- a bijective function $\sigma : (1, \ldots, n) \to (1, \ldots n)$, or

- a rearrangement of the elements of S in a different order.

A *transposition* is a permutation in which only two elements are in a different order, that is, for which $\sigma(i) = i$ for all $i = 1, \ldots, n$ except two.

Note that on a set of size n there are $n!$ permutations and $n(n-1)/2$ transpositions. It is a nontrivial fact to show that every permutation is a composition of transpositions, and the *parity* of a permutation is odd or even depeding on if the number of transpositions necessary to represent it is odd or even. The *sign* of a permutation σ is $+1$ if the parity of σ is even, and -1 is the parity of σ is odd. This number is denoted by $\text{sgn}(\sigma)$.

Remark 12.8. Another way to define the determinant of a matrix A is to say

$$\det(A) = \sum_{\text{permutations } \sigma} \text{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)} = \sum_{\text{permutations } \sigma} \text{sgn}(\sigma) \prod_{i=1}^{n} A_{i\sigma(i)}. \tag{6}$$

Example 12.9. There are $3! = 6$ permutations on a set of size 3, so a determinant of a 3×3 matrix is an alternating sum of 6 terms. The permutations are given below.

$$\begin{array}{ccc} & \rho & \\ 1 & \mapsto & 1 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 3 \end{array} \qquad \begin{array}{ccc} & \sigma & \\ 1 & \mapsto & 2 \\ 2 & \mapsto & 1 \\ 3 & \mapsto & 3 \end{array} \qquad \begin{array}{ccc} & \tau & \\ 1 & \mapsto & 3 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 1 \end{array} \qquad \begin{array}{ccc} & \mu & \\ 1 & \mapsto & 1 \\ 2 & \mapsto & 3 \\ 3 & \mapsto & 2 \end{array} \qquad \begin{array}{ccc} & \nu & \\ 1 & \mapsto & 2 \\ 2 & \mapsto & 3 \\ 3 & \mapsto & 1 \end{array} \qquad \begin{array}{ccc} & \lambda & \\ 1 & \mapsto & 3 \\ 2 & \mapsto & 1 \\ 3 & \mapsto & 2 \end{array}$$

The transpositions are σ, τ, μ. Note that $\nu = \tau \circ \sigma$ and $\lambda = \sigma \circ \tau$, which gives us a complete description of the signs of these permutations:

permutation σ	ρ	σ	τ	μ	ν	λ
$\text{sgn}(\sigma)$	1	-1	-1	-1	1	1

So if $A = \begin{bmatrix} 4 & -2 & 1 \\ 7 & 0 & 3 \\ -1 & -3 & 4 \end{bmatrix}$, then the determinant is

$$\begin{aligned} \det(A) &= A_{1\rho(1)} A_{2\rho(2)} A_{3\rho(3)} - A_{1\sigma(1)} A_{2\sigma(2)} A_{3\sigma(3)} + \cdots + A_{1\lambda(1)} A_{2\lambda(2)} A_{3\lambda(3)} \\ &= 4 \cdot 0 \cdot 4 - \cdot(-2) \cdot 7 \cdot 4 + \cdots + 1 \cdot 7 \cdot (-3) \\ &= 77. \end{aligned}$$

However, if we had a different matrix $A = \begin{bmatrix} 4 & 0 & 0 \\ 7 & 0 & 3 \\ 0 & -3 & 4 \end{bmatrix}$, then all permutations except one would have a factor of zero in them. That is, since the product $A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}$ has exactly one element in each row and exactly one elemen in each column, none of the terms in the combinatorial definition of the determinant can have two elements in the same row or in the same column. In other words,

$$\det(A') = \text{sgn}(\mu) \cdot 4 \cdot 3 \cdot (-3) = (-1) \cdot (-36) = 36.$$

Taking 4 in row 1, column 1, we cannot take any other element in column 1, so we must take row 2, column 3, to get a nonzero number. That leaves row 3, column 2 as the final factor (since columns 1 and 3 have already been used). All other terms in the expansion (6) will have at least one factor of 0, so can be safely ignored.

12.3 Exercises

Exercise 12.1. Show that the cross product $X(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1})$ is *skew-symmetric*, in the sense that swapping the order of two entries puts a negative sign in front.

Exercise 12.2. Find the parity of the two permutations below.

$$\begin{array}{ccc} & \sigma & \\ 1 & \mapsto & 1 \\ 2 & \mapsto & 3 \\ 3 & \mapsto & 2 \\ 4 & \mapsto & 4 \end{array} \qquad \begin{array}{ccc} & \rho & \\ 1 & \mapsto & 3 \\ 2 & \mapsto & 1 \\ 3 & \mapsto & 2 \\ 4 & \mapsto & 4 \end{array}$$

Use this to find the determinant of the matrix $A = \begin{bmatrix} 7 & 0 & -1 & 0 \\ 3 & 0 & 2 & 0 \\ 0 & -2 & 6 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Lecture 13: Eigenvalues and eigenvectors

- Fact 1: An $n \times n$ matrix has at most n eigenvalues. Some eigenvalues may be 0, some may be complex.
- Fact 2: The roots of the characteristic polynomial $\det(A - \lambda I)$ are the eigenvalues.

- Skill 1: Find eigenvectors and eigenvalues of a matrix
- Skill 2: Given only eigenvalues and eigenvectors of A, compute $A\mathbf{x}$ for any $\mathbf{x}$
- Skill 3: Given only eigenvalues and eigenvectors, construct a matrix with these eigenvalues and eigenvectors

---

This lecture begins a key topic for understanding the properties of a matrix. Eigenvectors are unique in that their direction does not change when multiplied by a matrix A (though their length may change).

13.1 How to find them

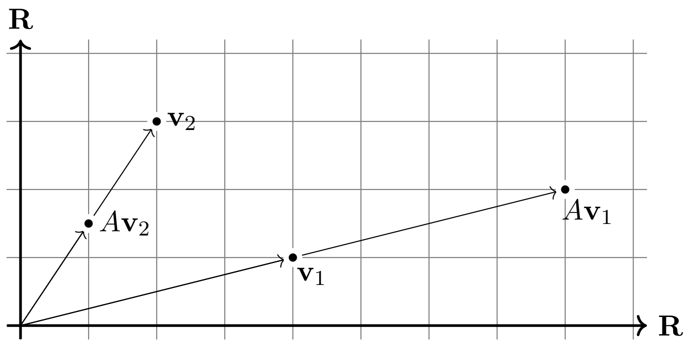
Definition 13.1. Let $A \in \mathcal{M}_{n\times n}$. Whenever there is a vector $\mathbf{v}$ for which $A\mathbf{v} = \lambda\mathbf{v}$, where $\lambda \in \mathbf{R}$, the vector $\mathbf{v}$ is called an *eigenvector* and λ is called its *eigenvalue*.

Example 13.2. Consider the following examples of eigenvectors and eigenvalues.

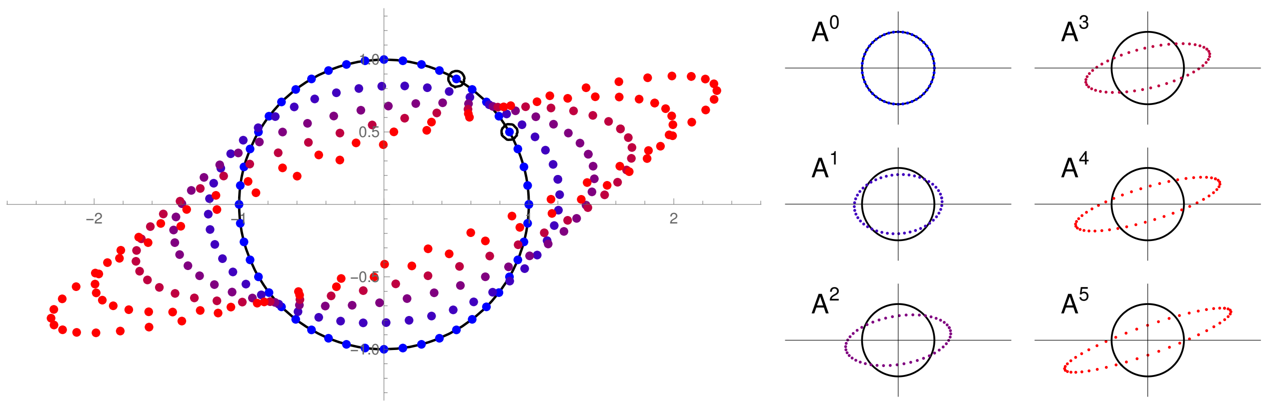
- The matrix $A = \begin{bmatrix} 4 & -1 \\ 0 & 2 \end{bmatrix}$ has eigenvector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ with eigenvalue 2. But A also has eigenvalue $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$ with eigenvalue 2.
- The matrix $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ has no (real) eigenvalues. This is the rotation matrix with $\theta = \frac{\pi}{2}$. In the second part of this lecture we will see how to get an eigenvalue from this matrix.
- The identity matrix has every vector as an eigenvector with eigenvalue 1.
- The projection matrix $P = \text{proj}_U$ (from Lecture 8) has every vector in U as an eigenvector with eigenvalue 1, and has every vector of $U^\perp$ as an eigenvector with eigenvalue 0.

Eigenvectors $\mathbf{v}, \mathbf{w}$ of a matrix A are called *independent* eigenvectors if the set $\{\mathbf{v}, \mathbf{w}\}$ is linearly independent.

Remark 13.3. Eigenvectors describe the *direction* in which a matrix changes $\mathbf{R}^n$, and the eigenvalues desribe the *stretching* that is done in that direction. For example, the matrix $A = \begin{bmatrix} 23/10 & -6/5 \\ 9/20 & 1/5 \end{bmatrix}$ has eigenvector $\mathbf{v}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ with eigenvalue 2, and eigenvector $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ with eigenvalue $\frac{1}{2}$.



The vector $\mathbf{v}_1$ gets longer and $\mathbf{v}_2$ gets shorter as A is applied more times. Adjusting $\mathbf{v}_1$ and $\mathbf{v}_2$ so that they make angles $\frac{\pi}{6}$ and $\frac{\pi}{3}$ with the x-axis, respectively, we can visually see what happens to vectors on the unit circle as A is applied more times.



The unit eigenvectors are marked with black circles around them. They are also distringuished from other vectors because their "trajectory" as A is applied is a striaght line. Below in Remark 13.6 we see what happens to vectors that are not exactly an eigenvector.

To find the eigenvalues λ, the equation $\det(A - \lambda I) = 0$ must be solved for λ. Once the possible solutions λ are found, then $A\mathbf{v} = \lambda v$ can be solved in each coordinate to find the corresponding eigenvector $\mathbf{v}$.

Example 13.4. Consider the matrix $A = \left[\begin{smallmatrix} 2 & 3 \\ -1 & 6 \end{smallmatrix}\right]$. What are its eigenvalues and corresponding eigenvectors? We must solve $\det(A - \lambda I) = 0$:

$$\begin{aligned} 0 &= \det\left(A - \lambda I\right) \\ &= \det\left(\begin{bmatrix} 2 & 3 \\ -1 & 6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\ &= \begin{vmatrix} 2-\lambda & 3 \\ -1 & 6-\lambda \end{vmatrix} \\ &= (2-\lambda)(6-\lambda) + 3 \\ &= 12 - 8\lambda + \lambda^2 + 3 \\ &= \lambda^2 - 8\lambda + 15 \\ &= (\lambda - 5)(\lambda - 3). \end{aligned}$$

Hence the eigenvalues are $\lambda = 5$ and $\lambda = 3$. To find the corrseponding eigenvectors, we solve:

$$A\mathbf{v} = 3\mathbf{v} \quad \iff \quad \begin{bmatrix} 2 & 3 \\ -1 & 6 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 3\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \iff \quad \begin{bmatrix} 2v_1 + 3v_2 \\ -v_1 + 6v_2 \end{bmatrix} = \begin{bmatrix} 3v_1 \\ 3v_2 \end{bmatrix}.$$

This is a linear system of 2 equations, which has solution (by back-substitution) $\mathbf{v} = \left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$, though we can choose any value we want for v_2 (and we choose 1 - to avoid such problems, we often take eigenvectors with unit length). Similarly, $\lambda = 3$ has the eigenvector $\left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$.

Definition 13.5. The function $\det(A - \lambda I)$ is called the *characteristic polynomial* for the matrix A.

Remark 13.6. If $A \in \mathcal{M}_{n\times n}$ has n eigenvectors, then knowing them and their eigenvalues is enough to know the effect of A on any matrix in $\mathbf{R}^n$. In Example 13.4 we found two eigenvalues and two eigenvectors. Then for any other vector we have

$$A\begin{bmatrix} 2 \\ -2 \end{bmatrix} = A\left(2\begin{bmatrix} 3 \\ 1 \end{bmatrix} - 4\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = 2A\begin{bmatrix} 3 \\ 1 \end{bmatrix} - 4A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2\cdot 5\begin{bmatrix} 3 \\ 1 \end{bmatrix} - 4\cdot 3\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 18 \\ -2 \end{bmatrix}.$$

13.2 Properties and applications

Proposition 13.7. Let $A \in \mathcal{M}_{n\times n}$.

- The eigenvalues of A and A^T are the same.
- If A is upper or lower triangular, its eigenvalues are on its diagonal.
- If the rank of A is less than n, A has an eigenvalue 0 for a non-trivial eigenvector.
- If $\mathbf{v}$ is an eigenvector of A with eigenvalue λ, then $\mathbf{v}$ is an eigenvector of A^n with eigenvalue λ^n.

Proof. The first point follows by distributing transposes in a sum (see Remark 2.13) in

$$\det(A-\lambda I) = \det((A-\lambda I)^T) = \det(A^T - (\lambda I)^T) = \det(A^T - \lambda I),$$

so the characteristic polynomial, and hence the eigenvalues, of A and A^T are the same.

The second point follows by using the standard basis of $\mathbf{R}^n$ as eigenvectors.

The third point follows by using a vector in the nullspace.

The fourth point follows from a repeated application of $A\mathbf{v} = \lambda\mathbf{v}$:

$$A^n\mathbf{v} = A^{n-1}(A\mathbf{v}) = A^{n-1}(\lambda\mathbf{v}) = \lambda A^{n-2}(A\mathbf{v}) = \lambda^2 A^{n-3}(A\mathbf{v}) = \cdots = \lambda^n\mathbf{v}.$$

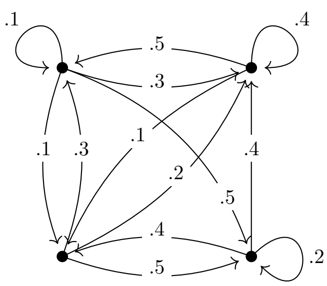
We are allowed to move the λ from the right to the left of A^{n-1} because λ is a number. $\square$

The first point above is similar to the determinant, however: row operations change the eigenvalues (they do not change the determinant). The sum of the diagonal entries in a matrix is called the *trace* of the matrix.

Definition 13.8. A *Markov matrix* is an $n \times n$ matrix with non-negative entries whose columns sum up to 1.

Markov matrices of size $n \times n$ often model a situation with n positions, and probabilities of moving from one position to the other. This is often used to model movement (of people, electricity).

Example 13.9. Consider a small circuit with probabilistic connections, as drawn below.



$$A = \begin{bmatrix} .1 & .5 & .3 & 0 \\ .3 & .4 & .2 & .4 \\ .1 & .1 & 0 & .4 \\ .5 & 0 & .5 & .2 \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} 0.472562 \\ 0.665087 \\ 0.309207 \\ 0.488606 \end{bmatrix}$$

Assuming these probabilities stay the same over time, a common question to ask is: will some flow repeat? Given an initial distribution of resources, how likely is that distribution to repeat after some time? This is called the *steady state* of the system, and corresponds to the eigenvector $\mathbf{v}$ with eigenvalue 1. We revisit this in Example 14.9 in the next lecture.

All Markov matrices have the eigenvalue 1 (they may have more eigenvalues), a claim justified fully in Lecture 23.

Sometime we come across matrices (as in Example 13.2) that do not seem to have eigenvalues, such as $A = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$. Its characteristic polynomial is $\lambda^2 + 1$. This polynomial has no real solutions, but does have *complex* solutions.

Definition 13.10. The *complex numbers* $\mathbf{C}$ are elements the set $\mathbf{R} \times \mathbf{R}$, expressed as $a + bi$, $a, b \in \mathbf{R}$, with a new operation:

$$(0,1) \cdot (0,1) = (-1,0) \quad \Longleftrightarrow \quad i \cdot i = -1.$$

Remark 13.11. Here are some key properties of the complex numbers .

- multiplying a complex number by i is "rotating the vector by 90 degrees"
- skew-symmetric matrices have complex eigenenvalues
- every polynomial with real (or complex) coefficients has roots in the complex numbers

The last statement says that $\mathbf{C}$ is *algebraically closed*.

Proposition 13.12. Let $A, B \in \mathcal{M}_{n \times n}$.

- The eigenvectors of $A + B$ can not be expressed in terms of the eigenvectors of A and B.
- A and B have the same eigenvectors iff A and B commute (that is, $AB = BA$).

13.3 Exercises

Exercise 13.1. Consider the matrix $A = \left[\begin{smallmatrix} 6 & -5 \\ 5 & -2 \end{smallmatrix}\right]$.

1. Find the eigenvalues and eigenvectors of A. Be careful, there may be complex numbers!
2. If $\mathbf{v}_1$ and $\mathbf{v}_2$ are the eigenvectors, compute the dot product $\mathbf{v}_1 \cdot \mathbf{v}_2$. Is it a complex or a real number?

Exercise 13.2. Consider the values $\lambda_1 = -3$, $\lambda_2 = -2$, $\lambda_3 = 5$.

1. Construct two different 3×3 matrices with $\lambda_1, \lambda_2, \lambda_3$ as eigenvalues.
2. What are the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ of the two matrices you created in part (a)?
3. If $\lambda_3 = -2$, explain why every linear combination of $\mathbf{v}_2$ and $\mathbf{v}_3$ is an eigenvector.

Exercise 13.3. You are given that a matrix B has eigenvalues $-1, 2, 5$ and a matrix C has eigenvalues $9, 3, 1$. Find the eigenvalues of the matrix

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & -2 & 0 & 0 \\ -2 & 2 & 0 & 0 & 0 & 7 \\ 8 & 0 & 3 & 0 & 8 & 0 \\ 0 & 0 & 0 & 9 & -9 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -5 & 2 & 1 \end{bmatrix}.$$

Exercise 13.4. Construct a 2×2 matrix with eigenvector $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ having eigenvalue λ, and eigenvector $\left[\begin{smallmatrix} z \\ w \end{smallmatrix}\right]$ having eigenvector μ.

Lecture 14: Diagonalization

- Fact 1: Each eigenvalue has at least one nonzero eigenvector. Eigenvectors cannot be the zero vector.
- Fact 2: An $n \times n$ matrix has exactly n eigenvalues, counting multiplicity. These are the roots of the characteristic polynomial, counting multiplicity.
- Fact 3: It is not always possible to find linearly independent eigenvectors. That is, not every matrix can be diagonalized.

- Skill 1: Diagonalize a matrix with linearly independent eigenvectors.
- Skill 2: Identify matrices that do not have linearly independent eigenvectors.
- Skill 3: Find eigenvalues and eigenvectors of powers of A and similar matrices to A.

The goal of this section is to reveal within each matrix a diagonal matrix. Diagonal matrices are easier to deal with, because they act like numbers rather than matrices. That is, multiplication and all other operations are much easier.

14.1 Diagonalizing matrices

We begin with an example by constructing a matrix from the eigenvectors.

Example 14.1. Let $\mathbf{u} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$, $\mathbf{v} = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, which are two linearly indepedent vectors in $\mathbf{R}^2$. Let A be a matrix with these two as eigenvectors, and corresponding eigenvalues $2, 3$, respectively. Such a matrix $A = \left[\begin{smallmatrix}a & b\\c & d\end{smallmatrix}\right]$ can be constructed by solving the equations

$$\begin{aligned} a+b &= 2 \\ c+d &= 2 \\ b &= 0 \\ d &= 3 \end{aligned} \qquad \iff \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix}.$$

We already have two equations $A\mathbf{u} = 2\mathbf{u}$ and $A\mathbf{v} = 3\mathbf{v}$, which can be combined into a single equation

$$A \underbrace{\begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix}}_{X} = \begin{bmatrix} | & | \\ 2\mathbf{u} & 3\mathbf{v} \\ | & | \end{bmatrix} = \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}}_{\Lambda}.$$

Let X be the matrix with eigenvalues of A as columns, and Λ be the diagonal matrix with eigenvalues of A on the diagonal. Then $AX = X\Lambda$, or $\Lambda = X^{-1}AX$. The inverse of X can be constructed because the rank of X is 2, as the columns are linearly indepedent (that is, the column space is full rank).

Definition 14.2. A matrix $A \in \mathcal{M}_{n\times n}$ can be *diagonalized* if it has n linearly independent eigenvectors. The process of expressing A as the product

$$A = X\Lambda X^{-1} \tag{7}$$

is called the *diagonalization* of A.

The order of the eigenvectors in X matches the order of the eigenvalues on the diagonal of Λ.

Remark 14.3. Note that X is not unique, but is unique up to scaling of columns: this follows from the same observation that eigenvectors are unique only up to scaling. That is, if $A\mathbf{x} = \lambda\mathbf{x}$, then also $A(c\mathbf{x}) = \lambda(c\mathbf{x})$, so $c\mathbf{x}$ is an eigenvector whenever $\mathbf{v}$ is an eigenvector, for any nonzero $c \in \mathbf{R}$. In terms of diagonalization, if $A = X\Lambda X^{-1}$, continuing from Example 14.1, we could have the eigenvectors $5\mathbf{u}$ and $-7\mathbf{v}$ instead of just $\mathbf{u}$ and $\mathbf{v}$. In that case,

$$X = \begin{bmatrix} | & | \\ 5\mathbf{u} & -7\mathbf{v} \\ | & | \end{bmatrix} = \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & -7 \end{bmatrix} \implies X^{-1} = \left(\begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & -7 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & -1/7 \end{bmatrix} \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix}^{-1},$$

and the decomposition in that case is

$$\begin{aligned} A &= \underbrace{\begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & -7 \end{bmatrix}}_{X} \Lambda \underbrace{\begin{bmatrix} 1/5 & 0 \\ 0 & -1/7 \end{bmatrix} \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix}^{-1}}_{X^{-1}} \\ &= \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix} \Lambda \begin{bmatrix} 5 & 0 \\ 0 & -7 \end{bmatrix} \begin{bmatrix} 1/5 & 0 \\ 0 & -1/7 \end{bmatrix} \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix}^{-1} \\ &= \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix} \Lambda \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix}^{-1}, \end{aligned}$$

which is the same decomposition as we had previously, with just $\mathbf{u}$ and $\mathbf{v}$. We used the fact that diagonal matrices commute with each other.

Example 14.4. Consider diagonalization for different types of matrices:

- If $A = I_n$, then we the eigenvectors are the standard basis vectors of $\mathbf{R}^n$, and the only eigenvalue is 1. This eigenvalue has *multiplicity* n, because there are n linearly independent eigenvectors with the same eigenvalue. That is, $A = X = \Lambda = I$.

- If A has all nonzero eigenvalues that are all the same, then A must be a multiple of the identity matrix. Indeed:

$$\Lambda = kI \implies A = X^{-1}(kI)X = kX^{-1}IX = kX^{-1}X = kI.$$

- If $A \in \mathcal{M}_{4\times 4}$ has two nonzero eigenvalues and two zero eigenvalues, then A may be diagonalizable, but not always. For example:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 3 & -1 \end{bmatrix}, \qquad \det(A - \lambda I) = \lambda^2((1-\lambda)(-1-\lambda) - 3) = \lambda^2(-4 + \lambda^2),$$

and the roots of the characteristic polynomial are $\lambda = 0$ and $\lambda = \pm 2$. By solving the appropriate matrix equation, we find the nonzero eigenvector / eigenvalue pairs to be

$$2 \text{ for } \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad -2 \text{ for } \begin{bmatrix} 0 \\ 0 \\ -1 \\ 3 \end{bmatrix}.$$

For the zero eigenvalues, the corresponding eigenvector $\begin{bmatrix} x & y & z & w \end{bmatrix}^T$ will have $z = 0$ and $w = 0$, but there will be no conditions on x, y, so by convention we choose $\mathbf{e}_1$ and $\mathbf{e}_2$ of the standard basis of $\mathbf{R}^4$ to be the eigenvectors. Diagonalization still works:

$$\underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & -\frac{1}{4} & \frac{1}{4} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{X^{-1}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 3 & -1 \end{bmatrix} = A$$

However, this works because we essentially have a diagonal block matrix $\left[\begin{smallmatrix} 0 & 0 \\ 0 & B \end{smallmatrix}\right]$, and the 2×2 matrix B had linearly independent eigenvectors. If we do not have a block matrix form with zero eigenvalues, then we cannot diagonalize. Consider the matrix

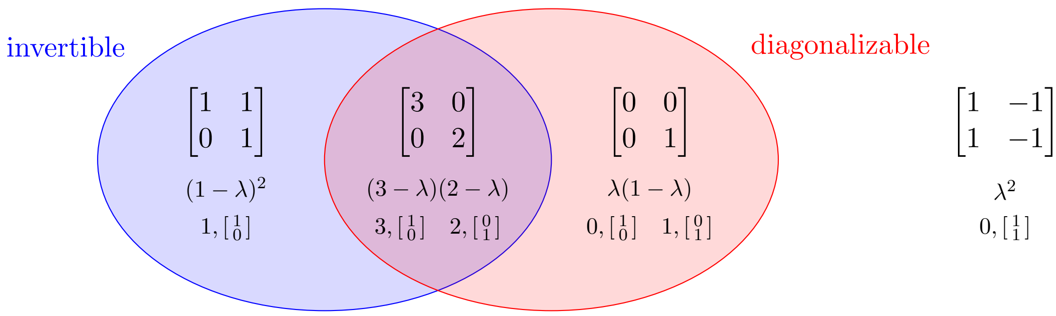
$$C = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \qquad \det(C - \lambda I) = (1-\lambda)(-1-\lambda) + 1 = \lambda^2,$$

and the roots of the characteristic polynomial are only $\lambda = 0$. The matrix equation to solve is

$$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \begin{bmatrix} x \\ y \end{bmatrix} \quad \Longleftrightarrow \quad \begin{array}{rcl} x - y & = & 0, \\ x - y & = & 0. \end{array}$$

It seems like the only eigenvector is $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$, but then $X = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$ does not have full rank and can not be diagonalized.

Remark 14.5. You may be tempted to think that a matrix being *invertible* is the same as being *diagonalizable*, but this is not true. In fact, there is no direct relationship between being invertible and diagonalizable, as the Venn diagram of such matrices below shows.



For eigenvalues λ_i and eigenvectors $\mathbf{v}_i$ of A, invertibility asks whether or not $\lambda_i = 0$. Diagonalizability asks whether or not the $\mathbf{v}_i$ are independent.

Proposition 14.6. If $A \in \mathcal{M}_{n \times n}$ has n different eigenvalues, then A has n linearly independent eigenvectors. That is, A is diagonalizable.

14.2 Consequences of diagonalizability

Diagonalizability allows us to make some nice conclusions.

Remark 14.7. Let $A \in \mathcal{M}_{n \times n}$ be diagonalizable, with eigenvector matrix X and corresponding eigenvalue matrix Λ. Then:

- For any invertible $B \in \mathcal{M}_{n \times n}$, the matrix $C = BAB^{-1}$ has the same eigenvalues as A, and has eigenvector matrix BX. Here C and A are called *similar matrices*.
- For any $k \in \mathbf{N}$, the matrix A^k is diagonalizable with the same eigenvectors as A, and with eigenvalues on the diagonal of Λ^k.
- If $|\lambda_i| = |\Lambda_{ii}| < 1$ for all i, then $\lim\limits_{k \to \infty} A^k \mathbf{x} = 0$ for any $\mathbf{x} \in \mathbf{R}^n$.

All of these facts follow directly from the diagonalizing equation $A = X\Lambda X^{-1}$. In the last point, for complex eigenvalues $\lambda = a + bi$, the absolute value is the product of λ with its *conjugate* $\lambda^* = a - bi$:

$$|\lambda| = |a + bi| = (a + bi)(a - bi) = a^2 - (bi)^2 = a - b^2 i^2 = a^2 + b^2.$$

Example 14.8. Consider the matrix $A = \begin{bmatrix} 1/6 & 1/3 \\ -1/6 & 2/3 \end{bmatrix}$. The roots of its characteristic polynomial are given by

$$0 = \det(A - \lambda I) = \left(\frac{1}{6} - \lambda\right)\left(-\frac{2}{3} - \lambda\right) + \frac{1}{3} \cdot \frac{1}{6} = \lambda^2 - \frac{5}{6}\lambda + \frac{1}{6} \quad \Longleftrightarrow \quad 0 = 6\lambda^2 - 5\lambda + 1,$$

which factors as $0 = (3\lambda - 1)(2\lambda - 1)$, so the eigenvalues are $\lambda_1 = \frac{1}{3}$ and $\lambda_2 = \frac{1}{2}$. By solving the appropriate matrix equations, we get the correspoinding eigenvectors to be $\mathbf{v}_1 = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$ and $\mathbf{v}_2 = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$, so the diagonalization of A is

$$A = \underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}}_{X^{-1}}.$$

The eigenvalues of A^k then are computed by the equation

$$A^k = \left(X\Lambda X^{-1}\right)^k = (X\Lambda X^{-1})(X\Lambda X^{-1}) \cdots (X\Lambda X^{-1}) = X\Lambda(X^{-1}X) \cdots (X^{-1}X)\Lambda X^{-1} = X\Lambda^k X^{-1},$$

and $\Lambda^k = \begin{bmatrix} 1/3^k & 0 \\ 0 & 1/2^k \end{bmatrix}$. Hence the eigenvectors of A^k are the same as those for A, and the eigenvalues are siomply powers of the original eigenvalues. We can even construct the matrix A^k explicitly:

$$\begin{aligned} A^k &= X\Lambda^k X^{-1} \\ &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1/3^k & 0 \\ 0 & 1/2^k \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2/3^k & 1/2^k \\ 1/3^k & 1/2^k \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\ &= \frac{1}{6^k} \begin{bmatrix} 2^{k+1} - 3^k & 2(3^k - 2^k) \\ 2^k - 3^k & 2 \cdot 3^k - 2^k \end{bmatrix} \end{aligned}$$

For example, when $k = 5$, we have

$$A^5 = \frac{1}{7776} \begin{bmatrix} -179 & 422 \\ -211 & 454 \end{bmatrix}.$$

Example 14.9. Recall Markov matrices from Definition 13.8 in Lecture 13. The example we saw has four eigenvalues and four eigenvectors:

$$A = \begin{bmatrix} .1 & .5 & .3 & 0 \\ .3 & .4 & .2 & .4 \\ .1 & .1 & 0 & .4 \\ .5 & 0 & .5 & .2 \end{bmatrix}, \qquad \Lambda \approx \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -.22 + .23i & 0 & 0 \\ 0 & 0 & -.22 - .23i & 0 \\ 0 & 0 & 0 & .13 \end{bmatrix}, \qquad X = \begin{bmatrix} | & | & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \\ | & | & | & | \end{bmatrix}.$$

A quick check shows that all eigenvalues have modulus less than 1 except the first one. The third point from Remark 14.7 is a special case when all eigenvalues have modulus less than 1, and directly implies that $\lambda_i^k \to 0$ for any individual eigenvaluie λ_i, irrespective of the other eigenvalues (since Λ is diagonal). This means that for the Markov matrix A and any $\mathbf{x} \in \mathbf{R}^4$,

$$\begin{aligned}
\lim_{k\to\infty} A^k\mathbf{x} &= \lim_{k\to\infty} A^k \left(\text{proj}_{\mathbf{v}_1}(\mathbf{x}) + \text{proj}_{\text{span}\{\mathbf{v}_1\}^\perp}(\mathbf{x})\right) \\
&= \left(\lim_{k\to\infty} A^k\text{proj}_{\mathbf{v}_1}(\mathbf{x})\right) + \left(\lim_{k\to\infty} A^k\text{proj}_{\text{span}\{\mathbf{v}_1\}^\perp}(\mathbf{x})\right) \\
&= \left(\lim_{k\to\infty} A^k(a\mathbf{v}_1)\right) + \left(\lim_{k\to\infty} A^k(b\mathbf{v}_2 + c\mathbf{v}_3 + d\mathbf{v}_4)\right) \\
&= a\left(\lim_{k\to\infty} A^k\mathbf{v}_1\right) + b\left(\lim_{k\to\infty} A^k\mathbf{v}_2\right) + c\left(\lim_{k\to\infty} A^k\mathbf{v}_3\right) + d\left(\lim_{k\to\infty} A^k\mathbf{v}_4\right) \\
&= a\left(\lim_{k\to\infty} (1)^k\mathbf{v}_1\right) + b\left(\lim_{k\to\infty} (-.22+.23i)^k\mathbf{v}_2\right) + c\left(\lim_{k\to\infty} (-.22-.23i)^k\mathbf{v}_3\right) + d\left(\lim_{k\to\infty} (.13)^k\mathbf{v}_4\right) \\
&= a\mathbf{v}_1 + 0\mathbf{v}_2 + 0\mathbf{v}_3 + 0\mathbf{v}_4 \\
&= \text{proj}_{\mathbf{v}_1}(\mathbf{x}).
\end{aligned}$$

Here we used some real numbers a, b, c, d to ease notation. Interpreting this result in the context of the Markov matrix, we get that any initial distribution of resources will approach the (projection to the) steady state of the system after enough time has passed. This is why steady states are so important to dynamic systems like Markov matrices.

14.3 Exercises

Exercise 14.1. Decopose both matrices below in their $X\Lambda X^{-1}$-decomposition, where Λ is a diagonal matrix with the eigenvalues, and X is the matrix with columns as eigenvectors.

$$A = \begin{bmatrix} 2 & 2 \\ 5 & 5 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

Exercise 14.2. Let $A \in \mathcal{M}_{3\times 3}$ with the eigenvectors $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$ and eigenvalues $-1, 2, -3$, respectively.

1. Construct the eigenvector matrix X and the eigenvalues matrix Λ.
2. Construct A by the diagonalization equation $A = X\Lambda X^{-1}$.

Exercise 14.3. Diagonalize the matrices A, B below and find what A^k and B^k look like, for any $k \in \mathbf{N}$. Your answers should have the value k in them.

$$A = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}, \qquad B = \begin{bmatrix} 5 & 1 \\ 0 & 10 \end{bmatrix}.$$

Part III
Generalizations

Lecture 15: Special matrices

- Fact 1: Symmetric matrices can be decomposed with an orthonormal matrix of eigenvectors.
- Fact 2: Positive definiteness can be expressed in terms of pivots, eigenvalues, determinants, and matix or vector multiplications.

- Skill 1: Apply the results of the spectral theorem
- Skill 2: Express a symmetric matrix as a sum of rank one matrices
- Skill 3: Check if a matrix is positive definite using equivalent properties

One reason we study symmetric and positive definite matrices is that theytie together pivots, eigenvalues, determinants, and other topics we have seen. Another reason is that A^TA and AA^T are both symmetric and positive definite. Understanding A^TA and AA^T instead of A makes life much easier.

15.1 Symmetric matrices

Recall from Definition 2.12 in Lecture 2 that a matrix $A \in \mathcal{M}_{n\times n}$ is *symmetric* if $A_{ij} = A_{ji}$ for all $1 \leqslant i, j \leqslant n$. This property makes many of the previous computations we did before much easier.

Proposition 15.1 (The Spectral Theorem)**.** For $S \in \mathcal{M}_{n\times n}$ symmetric:

- S has real eigenvalues and orthogonal eigenvectors
- S can always be diagonalized

It follows immediately that for S symmetric, its eigenvalue-eigenvector decomposition $S = X\Lambda X^{-1}$ becomes $S = Q\Lambda Q^T$, for Q having orthonormal columns. Recall that then its inverse is equal to its transpose: $Q^{-1} = Q^T$. Although the statement above says "orthogonal", not "orthonormal", we can make the vectors orthonormal by dividing by their length.

Example 15.2. Consider $S = \left[\begin{smallmatrix}1 & 2\\ 2 & 4\end{smallmatrix}\right]$. We can find its eigenvalues by solving

$$0 = \det(S - \lambda I) = (1-\lambda)(4-\lambda) = 4 = -5\lambda + \lambda^2 = \lambda(\lambda - 5),$$

for which $\lambda_1 = 0$ and $\lambda_2 = 5$. We find the eigenvectors by solving

$$\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = 0\begin{bmatrix}0\\ y\end{bmatrix} \iff \begin{array}{rcl} x+2y &=& 0\\ 2x+4y &=& 0\end{array} \implies \begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}1\\ -\frac{1}{2}\end{bmatrix} = \mathbf{v}_1,$$
$$\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = 5\begin{bmatrix}0\\ y\end{bmatrix} \iff \begin{array}{rcl} x+2y &=& 5x\\ 2x+4y &=& 5y\end{array} \implies \begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}\frac{1}{2}\\ 1\end{bmatrix} = \mathbf{v}_2.$$

These vectors are orthogonal as $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. They both have length $\sqrt{5}/2$, so the normalized vectors are

$$\mathbf{q}_1 = \begin{bmatrix}2/\sqrt{5}\\ -1/\sqrt{5}\end{bmatrix}, \qquad \mathbf{q}_2 = \begin{bmatrix}1/\sqrt{5}\\ 2/\sqrt{5}\end{bmatrix}.$$

This gives us the diagonalization as

$$\underbrace{\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix}}_{S} = \underbrace{\begin{bmatrix}2/\sqrt{5} & 1/\sqrt{5}\\ -1/\sqrt{5} & 2/\sqrt{5}\end{bmatrix}}_{Q}\underbrace{\begin{bmatrix}0 & 0\\ 0 & 5\end{bmatrix}}_{\Lambda}\underbrace{\begin{bmatrix}2/\sqrt{5} & -1/\sqrt{5}\\ 1/\sqrt{5} & 2/\sqrt{5}\end{bmatrix}}_{Q^T}.$$

Remark 15.3. The fact that $S = Q\Lambda Q^T$, where Q has orthonormal columns, allows us to write S in another way. If $S \in \mathcal{M}_{3\times 3}$, then

$$\begin{aligned} S &= \underbrace{\begin{bmatrix} | & | & | \\ \mathbf{u} & \mathbf{v} & \mathbf{w} \\ | & | & | \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} - & \mathbf{u}^T & - \\ - & \mathbf{v}^T & - \\ - & \mathbf{w}^T & - \end{bmatrix}}_{Q^T} \\ &= \underbrace{\begin{bmatrix} | & | & | \\ \lambda_1\mathbf{u} & \lambda_2\mathbf{v} & \lambda_3\mathbf{w} \\ | & | & | \end{bmatrix}}_{Q\Lambda} \begin{bmatrix} - & \mathbf{u}^T & - \\ - & \mathbf{v}^T & - \\ - & \mathbf{w}^T & - \end{bmatrix} \\ &= \lambda_1\mathbf{u}\mathbf{u}^T + \lambda_2\mathbf{v}\mathbf{v}^T + \lambda_3\mathbf{w}\mathbf{w}^T, \end{aligned}$$

which is a sum of 3×3 rank one matrices. This description will be important for the next lecture.

We finish off the first part of this lecture with another comment about the relationship between pivots and eigenvalues.

Remark 15.4. Let $A \in \mathcal{M}_{n\times n}$. Below are the main facts about pivots and eigenvalues summarized, along with a new one:

- $\det(A)$ = (product of pivots) = (product of eigenvalues)
- $\text{trace}(A)$ = (sum of eigenvalues)
- (number of pivots > 0) = (number of eigvals > 0) whenever A is symmetric

This last fact is counting multiplicity. It follows from the LDU-decomposition of a symmetric matrix, which turns into LDL^T.

15.2 Positive definite matrices

The second part of this lecture focuses on special types of symmetric matrices.

Definition 15.5. A symmetric matrix with all positive eigenvalues is called *positive definite*. A symmetric matrix with some positive and some zero eigenvalues is called (*positive*) *semidefinite*.

Finding eigenvalues is computationally intensive for large matrices, so we use the relationship with pivots from Remark 15.4 to determine when eigenvalues are positive. This gives several quick ways to determine when a matrix is positive definite.

Example 15.6. The 2×2 symmetric matrix $\left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]$ has pivots $a, c - \frac{b^2}{a}$, so the pivots are positive iff $a > 0$ and $ac - b^2 > 0$. For example, all the symmetric matrices

$$\begin{bmatrix} 1 & 10 \\ 10 & 200 \end{bmatrix}, \qquad \begin{bmatrix} 22 & -3 \\ -3 & 2 \end{bmatrix}, \qquad \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

are positive definite because they have positive eigenvalues.

Remark 15.7. The $n \times n$ positive definite matrix S with eigenvector $\mathbf{v}$ and positive eigenvalue λ has

$$S\mathbf{v} = \lambda\mathbf{v} \quad \Longrightarrow \quad \mathbf{v}^T S\mathbf{v} = \lambda\mathbf{v}^T\mathbf{v} = \lambda(v_1^2 + \cdots v_n^2) > 0.$$

Even more, for any $\mathbf{x} \in \mathbf{R}^n$, we can express it as a linear combination $a_1\mathbf{v}_1 + \cdots a_n\mathbf{v}_n$ of the orthonormal eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ of S, and by orthonormality of these vectors,

$$\mathbf{x}^T S\mathbf{x} = (a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n)^T(a_1\lambda_1\mathbf{v}_1 + \cdots + a_n\lambda_n\mathbf{v}_n) = a_1^2\lambda_1\|\mathbf{v}_1\|^2 + \cdots + a_n^2\lambda_n\|\mathbf{v}_n\|^2 > 0.$$

In other words, saying S is positive definite is equivalent to saying that $\mathbf{x}^T S\mathbf{x} > 0$ for any vector $\mathbf{x} \in \mathbf{R}^n$ (except of course the zero vector).

Proposition 15.8. The previous remark has some nice consequences:

- If $S, T \in \mathcal{M}_{n\times n}$ are positive definite, then $S+T$ is positive definite.
- If $A \in \mathcal{M}_{m\times n}$ has independent columns, then A^TA is positive definite.

Proof. The first point follows from distributing

$$\mathbf{x}^T(S+T)\mathbf{x} = \mathbf{x}^TS\mathbf{x} + \mathbf{X}^TT.$$

The second point comes from rewriting

$$\mathbf{x}^T(A^TA)\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = \|A\mathbf{x}\|^2 > 0.$$

□

The proof of the second claim implies that A^TA (and also AA^T) is always positive semidefinite. We have now arrived at a nice summary of positive definite matrices.

Proposition 15.9. The following statements about $S \in \mathcal{M}_{n\times n}$ symmetric equivalent.

- S is positive definite
- S has all positive pivots
- S has all positive eigenvalues
- Every top-left submatrix of S has positive determinant
- $\mathbf{x}^TS\mathbf{x} > 0$ for any nonzero $\mathbf{x} \in \mathbf{R}^n$
- There exists $A \in \mathcal{M}_{m\times n}$ with independent columns and $S = A^TA$

Example 15.10. Let's check all the claims above on a simple matrix $S = \left[\begin{smallmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{smallmatrix}\right]$. For the pivots, we quickly row reduce:

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \to \begin{bmatrix} 2 & -1 & 0 \\ 0 & 3/2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \to \begin{bmatrix} 2 & -1 & 0 \\ 0 & 3/2 & -1 \\ 0 & 0 & 4/3 \end{bmatrix}$$

The pivots are $2, 3/2, 4/3$, which are all positive. The eigenvalues are the roots of

$$\det(S - \lambda I) =$$

15.3 Exercises

Exercise 15.1. Let $a \in \mathbf{R}$ be nonzero.

1. Find the eigenvalues of $\left[\begin{smallmatrix} 0 & a \\ -a & 0 \end{smallmatrix}\right]$.
2. Find the eigenvalues of $\left[\begin{smallmatrix} 0 & 0 & a \\ 0 & ia & 0 \\ -a & 0 & 0 \end{smallmatrix}\right]$.
3. Using a, construct a 4×4 skew-symmetric matrix that has all imaginary eigenvalues.
4. Construct a 3×3 symmetric matrix that has three pivots a and no zero entries.

Exercise 15.2. Let $A \in \mathcal{M}_{m\times n}$. Show that AA^T and A^TA are both symmetric matrices.

Exercise 15.3. The numbers a, b, c are chosen randomly from the set of integers $\{-3, -2, \ldots, 2, 3\}$, with replacement, to create a matrix $A = \left[\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right]$.

1. What is the probability that A is symmetric?

2. What is the probability that A is positive definite?

Exercise 15.4. Consider the two symmetric matrices below, for $a, b \in \mathbf{R}$:

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & a & 2 \\ 2 & 2 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} b & 2 & 0 \\ 2 & b & 3 \\ 0 & 3 & b \end{bmatrix}.$$

1. Find the pivots for both matrices. For what values of a, b will the pivots be positive?
2. Find the eigenvalues for both matrices. For what values of a, b will the eigenvalues be positive?
3. Find the upper left determinants for both matrices. For what values of a, b will the determinants be positive?
4. Choose some b so that pivots, eigenvalues, determinants are positive. Find the $Q\Lambda Q^T$-decomposition for B.

Lecture 16: Singular value decomposition

- Fact 1: No matter what size A has, AA^T and A^TA have the same nonzero eigenvalues
- Fact 2: The eigenvectors of AA^T and A^TA are related by A and the singular values σ_i
- Fact 3: The SVD contains orthonormal bases of the four fundamental subspaces

---

- Skill 1: Compute the rank r approximation to A
- Skill 2: Decompose a non-square matrix A by the SVD

This lecture has a very practical application - image compression. An image can be considered as a matrix with entries in the range of colors. Colors are on a specturm, so there is a minimum and maximum color number.

16.1 Eigenvalues of symmetric matrices

The word *singular* so far has been used when talking about matrices. A square matrix is singular if its determinant is zero, and non-singular otherwise.

Definition 16.1. Let $A \in \mathcal{M}_{m\times n}$. The *singular values* of A are

- the square roots of the positive eigenvalues of AA^T or A^TA, if A is not symmetric;
- the positive eigenvalues of A, if A is symmetric.

The first definition is the same as the second when A is symmetric.

Very often A is not symmetric, but always both AA^T and A^TA are symmetric. That is they have real eigenvalues - but they also have the same eignvalues. To see this, note that for $\lambda \neq 0$,

$$\begin{aligned} A^TA\mathbf{x} = \lambda\mathbf{x} \quad &\implies \quad AA^T(A\mathbf{x}) = \lambda(A\mathbf{x}), \\ AA^T\mathbf{y} = \lambda\mathbf{y} \quad &\implies \quad A^TA(A^T\mathbf{y}) = \lambda(A^T\mathbf{y}). \end{aligned}$$

That is, λ is an eigenvalue of AA^T whenever λ is an eigenvalue of A^TA, and λ is an eigenvalue of A^TA whenever λ is an eigenvalue of AA^T. This implies that:

- AA^T and A^TA have the same number k of nonzero eigenvalues, and they are equal
- if AA^T has more eigenvalues (when $m > n$) than A^TA, then the extra ones are zero

However, this does not imply that AA^T and A^TA have the same number of independent eigenvectors!

Example 16.2. Let's do a quick example before making a general observation about the relationship of the eigenvectors of AA^T and A^TA. Consider the matrix $A = \left[\begin{smallmatrix} 1&1&1&2&3 \\ 2&2&2&2&3 \\ 3&3&3&3&3 \end{smallmatrix}\right] \in \mathcal{M}_{3\times 5}$, for which

$$AA^T = \begin{bmatrix} 16 & 19 & 24 \\ 19 & 25 & 33 \\ 24 & 33 & 45 \end{bmatrix} \in \mathcal{M}_{3\times 3}, \qquad A^TA = \begin{bmatrix} 14 & 14 & 14 & 15 & 18 \\ 14 & 14 & 14 & 15 & 18 \\ 14 & 14 & 14 & 15 & 18 \\ 15 & 15 & 15 & 17 & 21 \\ 18 & 18 & 18 & 21 & 27 \end{bmatrix} \in \mathcal{M}_{5\times 5}.$$

We can find the eigenvalues and eigenvectors of both AA^T and A^TA, as they are square:

$$AA^T : (\lambda_1, \lambda_2, \lambda_3) \approx (83.38, 2.49, 0.13)$$
$$A^TA : (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (83.38, 2.49, 0.13, 0, 0)$$

Applying the decomposition from Remark 15.3, we can write them as sums of rank one matrices. We do this only for AA^T:

$$AA^T \approx \underbrace{83.38}_{\lambda_1} \underbrace{\begin{bmatrix} 0.17 & 0.23 & 0.3 \\ 0.23 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.53 \end{bmatrix}}_{\mathbf{u}_1\mathbf{u}_1^T} + \underbrace{2.49}_{\lambda_2} \underbrace{\begin{bmatrix} 0.69 & 0.08 & -0.45 \\ 0.08 & 0.01 & -0.05 \\ -0.45 & -0.05 & 0.3 \end{bmatrix}}_{\mathbf{u}_2\mathbf{u}_2^T} + \underbrace{0.13}_{\lambda_3} \underbrace{\begin{bmatrix} 0.14 & -0.31 & 0.15 \\ -0.31 & 0.69 & -0.34 \\ 0.15 & -0.34 & 0.17 \end{bmatrix}}_{\mathbf{u}_3\mathbf{u}_3^T}.$$

Notice the very large eigenvalue and the two smaller ones. This decomposition will be useful when we ignore the smaller eigenvalues.

Remark 16.3. Let $A \in \mathcal{M}_{m\times n}$. Let $\mathbf{u}_1, \dots, \mathbf{u}_m \in \mathbf{R}^m$ be the eigenvectors of AA^T and $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbf{R}^n$ be the eigenvectors of A^TA. Without loss of generality, we assume that $n \geqslant m$, so $\mathbf{v}_{m+1}, \dots, \mathbf{v}_n$ are all eigenvectors for the zero eigenvalue. Let $\sigma_1, \dots, \sigma_m \in \mathbf{R}$ be such that

$$AA^T\mathbf{u}_i = \sigma_i^2\mathbf{u}_i \qquad \text{and} \qquad A^TA\mathbf{v}_i = \sigma_i^2\mathbf{v}_i,$$

for all $i = 1, \dots, m$. We may do this because AA^T and A^TA are both positive semidefinite (so we can take square roots of the eigenvalues). We use σ instead of λ because these are the *singular values* - the letter σ is the letter "s" in English. The relationship among the $\mathbf{u}_i$, $\mathbf{v}_i$, σ_i and the original matrix A is then given by

$$A^T\mathbf{u}_i = \sigma_i\mathbf{v}_i \qquad \text{and} \qquad A\mathbf{v}_i = \sigma_i\mathbf{u}_i,$$

as multiplying the left equation by A on the left means the equation on the right must be true (for the previous equation to hold). We now get a decomposition

$$A\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \cdots & \sigma_m\mathbf{u}_m \\ | & | & & | \end{bmatrix},$$

which, after using the orthonormality of the eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_m$ and decomposing the right side, becomes

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_1^T & - \\ - & \mathbf{v}_2^T & - \\ & \vdots & \\ - & \mathbf{v}_m^T & - \end{bmatrix} = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_m\mathbf{u}_m\mathbf{v}_m^T.$$

We now consider an examle of how decomposition lets us reduce the size of information.

Example 16.4. Suppose we want to transmit via some medium the flag of Latvia, as a matrix:

$$L = \begin{bmatrix} r & r & r & r & r & r & r & r & r & r \\ r & r & r & r & r & r & r & r & r & r \\ w & w & w & w & w & w & w & w & w & w \\ r & r & r & r & r & r & r & r & r & r \\ r & r & r & r & r & r & r & r & r & r \end{bmatrix}$$

There are 50 pieces of information, but we can easily see this is a rank one product of two vectors:

$$L = \mathbf{u}\mathbf{v}^T = \begin{bmatrix} r \\ r \\ w \\ r \\ r \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

So now instead of sending 5×10 pieces of data, we can just send $5 + 10$, a 70% reduction in size.

What if the flag is more complex, such as the flag of navy ships of Latvia:

$$W = \begin{bmatrix} w & w & w & r & w & r & w & w & w \\ w & w & w & r & w & r & w & w & w \\ r & r & r & r & w & r & r & r & r \\ w & w & w & w & w & w & w & w & w \\ r & r & r & r & w & r & r & r & r \\ w & w & w & r & w & r & w & w & w \\ w & w & w & r & w & r & w & w & w \end{bmatrix}.$$

We can reduce the 7×9 pieces of data by singular value decomposition. For ease of notation, change r to 1 and w to 0. We then simply compute the eigenvalues and eigenvectors of WW^T and W^TW. We are lucky and see there are only 2 nonzero eigenvalues:

$$(\sigma_1^2, \sigma_2^2) \approx (18.93, 5.07), \qquad \mathbf{u}_1 \approx \begin{bmatrix} -0.23 \\ -0.23 \\ -0.63 \\ 0 \\ -0.63 \\ -0.23 \\ -0.23 \end{bmatrix}, \mathbf{u}_2 \approx \begin{bmatrix} -0.44 \\ -0.44 \\ 0.33 \\ 0 \\ 0.33 \\ -0.44 \\ -0.44 \end{bmatrix}, \qquad \mathbf{v}_1 \approx \begin{bmatrix} -0.29 \\ -0.29 \\ -0.29 \\ -0.5 \\ 0 \\ -0.5 \\ -0.29 \\ -0.29 \\ -0.29 \end{bmatrix}, \mathbf{v}_2 \approx \begin{bmatrix} 0.29 \\ 0.29 \\ 0.29 \\ -0.5 \\ 0 \\ -0.5 \\ 0.29 \\ 0.29 \\ 0.29 \end{bmatrix}.$$

Reducing from $7 \times 9 = 63$ to $2 + 2 \times 7 + 2 \times 9 = 34$ is done by the decomposition

$$W = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T.$$

Definition 16.5. Let $A \in \mathcal{M}_{m \times n}$, and let $\sigma_1, \sigma_2, \ldots$ be the eigenvalues of AA^T (or A^TA) in decreasing order. The *rank r approximation* of A is the sum

$$\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T \in \mathcal{M}_{m \times n},$$

for every $1 \leqslant r \leqslant \operatorname{rank}(A)$.

16.2 Bases in the decomposition

The rank r approximation to $A \in \mathcal{M}_{m \times n}$, for $r \leqslant \operatorname{rank}(A) \leqslant \min\{m, n\}$, gives a decomposition of A into three matrices, using eigenvalues and eigenvectors of AA^T and A^TA:

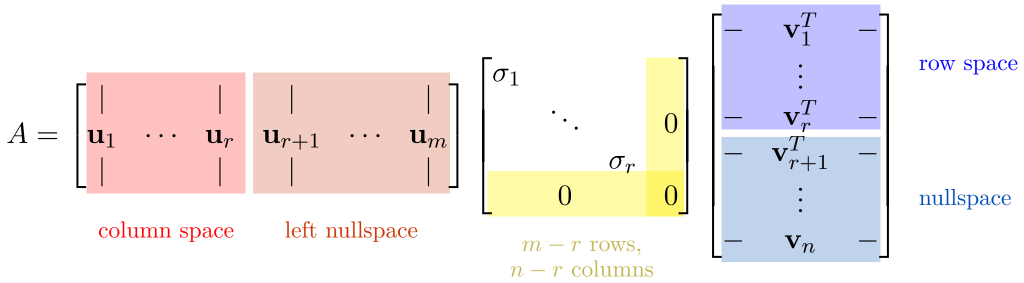
$$A = \underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \\ | & | & & | \end{bmatrix}}_{\text{eigenvectors of } AA^T} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}}_{\text{eigenvalues}} \underbrace{\begin{bmatrix} - & \mathbf{v}_1^T & - \\ - & \mathbf{v}_2^T & - \\ & \vdots & \\ - & \mathbf{v}_r^T & - \end{bmatrix}}_{\text{eigenvectors of } A^TA}. \tag{8}$$

This equation can be generalized, and it reveals bases for the four fundamental subspaces that we have already seen in Lecture 6.

Definition 16.6. The *singular value decomposition* of $A \in \mathcal{M}_{m \times n}$ is $A = U\Sigma V^T$, where

- $U \in \mathcal{M}_{m \times m}$ has the eigenvectors of AA^T as columns,
- $V \in \mathcal{M}_{n \times n}$ has the eigenvectors of A^TA as columns,
- $\Sigma \in \mathcal{M}_{m \times n}$ has the eigenvalues of AA^T (or A^TA) on the diagonal of its upper left $\operatorname{rank}(A) \times \operatorname{rank}(A)$ submatrix, in decreasing order from the largest in Σ_{11}.

The order of the eigenvectors in U and V corresponds to the order of the eigenvalues in Σ. Moreover, this decomposition contains orthonormal basis vectors of other subspaces:



The vectors $\mathbf{u}_i$ are called the *left singular vectors* and the $\mathbf{v}_i$ are called the *right singular vectors* of A.

Example 16.7. Let's compute the full SVD for a matrix, and get the appropriate bases. Consider

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ -1 & -1 \end{bmatrix}, \qquad AA^T = \begin{bmatrix} 2 & 4 & -2 \\ 4 & 8 & -4 \\ -2 & -4 & 2 \end{bmatrix}, \qquad A^TA = \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}.$$

It is immediate that A has rank 1, as the rows are all multiples of the first row. We already know both A^TA and AA^T have the same eigenvalues, so we just find them for the easier of the two, A^TA. The roots of the characteristic polynomial are found by

$$0 = \det(A^TA - \lambda I) = (6-\lambda)^2 - 36 = 36 - 12\lambda + \lambda^2 - 36 = \lambda^2 - 12\lambda = (\lambda - 12)\lambda,$$

so the eigenvalues are 12 and 0. Hence the only singular value is $\sigma_1 = 2\sqrt{3}$. To find the eigenvectors, we row reduce the appropriate augmented matrices, remembering to normalize the eigenvectors.

$$12 \text{ for } AA^T: \begin{bmatrix} -10 & 4 & -2 & 0 \\ 4 & -4 & -4 & 0 \\ -2 & -4 & -10 & 0 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \mathbf{u}_1 = \begin{bmatrix} -1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$$

$$0 \text{ for } AA^T: \begin{bmatrix} 2 & 4 & -2 & 0 \\ 4 & 8 & -4 & -0 \\ -2 & -4 & 2 & 0 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \mathbf{u}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}$$

$$12 \text{ for } A^TA: \begin{bmatrix} -6 & 6 & 0 \\ 6 & -6 & 0 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies \mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$0 \text{ for } A^TA: \begin{bmatrix} 6 & 6 & 0 \\ 6 & 6 & 0 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

We could have also found $\mathbf{v}_1$ by $A^T\mathbf{u}_1 = 2\sqrt{3}\mathbf{v}_1$. This gives us the complete decomposition

$$A = \underbrace{\begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{2} & -2/\sqrt{5} \\ -2/\sqrt{6} & 0 & 1/\sqrt{5} \\ 1/\sqrt{6} & 1/\sqrt{2} & 0 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}}_{V^T},$$

as well as bases

$$\operatorname{col}(A) = \operatorname{span}\left\{ \begin{bmatrix} -1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix} \right\}, \qquad \operatorname{null}(A^T) = \operatorname{span}\left\{ \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix} \right\},$$

$$\operatorname{row}(A) = \operatorname{span}\left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}, \qquad \operatorname{null}(A) = \operatorname{span}\left\{ \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}.$$

Remark 16.8. If A is symmetric, then the SVD is the same as the $Q\Lambda Q^T$-decomposition. In this way, the SVD is a more general decomposition that captures the nice properties of the $Q\Lambda Q^T$-decomposition.

16.3 Exercises

Exercise 16.1. This question uses Python. You may use the folowing resources:

- Sample code: `jlazovskis.com/teaching/linearalgebra`
- Sample images: `links.uwaterloo.ca/Repository.html`

Find a grayscale image online at least 100×100 pixels in size. It does not have to be square.

⋈ 1. Find the singular values of the image. How many of them are less than 1/100 of the largest singular value?

⋈ 2. Compute the rank r approximation to the image for $r = 1, 2, 3, 5, 10$.

3. If the image had size $m \times n$, what is the percent reduction in size for the rank r approxmation?

Exercise 16.2. Let $a \in \mathbf{R}_{\neq 0}$, and consider the matrix

$$A = \begin{bmatrix} a & 0 & a & 0 \\ 0 & 0 & 0 & 2a \end{bmatrix}.$$

1. Compute the SVD of A by finding the eigenvalue / eigenvector pairs for AA^T and A^TA.
2. What are the dimensions of the four fundamental subspaces of A?

Exercise 16.3. 1. Construct a 3×4 matrix with singular values $1, 2, 3$.

2. Construct a 2×2 rank 1 matrix with right singular vectors $\begin{bmatrix} 1/2 \\ \sqrt{3}/2 \end{bmatrix}, \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix}$.

3. Find the rank 1 and rank 2 approximations for

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

Hint: Since two eigenvalues are the same, there are two rank 2 approximations!

Lecture 17: Principal component analysis

Chapter 7.3 in Strang

- Fact 1: The first principal component solves the perpendicular least squares problem
- Fact 2: The first two principal components give a reasonable way to plot high-dimensional data

- Skill 1: Solve the perpendicular least squares problem using SVD
- Skill 2: Identify the principal components of $A \in \mathcal{M}_{m\times n}$, in terms of the total covariance of A
- Skill 3: Normalize and center data on its mean

In the previous lecture, we saw how to simplify fully known images (thought of as a matrix A), like a flag, for compressed communication, using the singular value decomposition of the symmetric matrices AA^T and A^TA.

17.1 The first significant direction

All data used in this lecture is available on the course website `jlazovskis.com/teaching/linearalgebra`.

Example 17.1. Consider the following data set, representing the number of instructors (x-value) and the number of students (y-value) at 32 different post-secondary institutions in Latvia.

x-value	1531	904	509	305	182	142	120	101	88	75	71	65	58	54	47	45
y-value	15260	14006	9541	3891	2068	563	876	1218	1650	662	769	695	1406	441	557	1079

x-value	44	41	37	36	30	25	22	22	20	18	15	15	13	12	10	2
y-value	670	667	1076	593	393	354	261	50	567	261	211	155	111	153	262	33

Each column (an (x, y)-pair) is a *sample*, so we can costruct a *sample matrix* $A \in \mathcal{M}_{2\times 32}$.

Definition 17.2. Let $A \in \mathcal{M}_{m\times n}$ and consider each of the n columns of A as a sample. The *mean-centered* matrix of A is $M \in \mathcal{M}_{m\times n}$, with

$$M_{ij} = A_{ij} - \underbrace{\frac{1}{n}\sum_{k=1}^{n} A_{ik}}_{\text{mean of row } i} .$$

The *sample covariance* matrix of A is $S = \frac{AA^T}{n-1} \in \mathcal{M}_{m\times m}$.

We then create the mean-centered and sample covariance matrices for the data above:

$$M = \begin{bmatrix} 1385.41 & 758.41 & 363.41 & \cdots & -133.59 & -135.59 & -143.59 \\ 13369.41 & 12115.41 & 7650.41 & \cdots & -1737.59 & -1628.59 & -1857.59 \end{bmatrix},$$

$$S = \begin{bmatrix} 73909.14 & 864786.84 \\ 864786.84 & 10971745.39 \end{bmatrix}.$$

Next, we compute the SVD of the sample covariance matrix S. Since S is already symmetric, the matrices U and V are the same. The eigenvector with the largest eigenvalue identifies the *principal component* of $A \in \mathcal{M}_{m\times n}$, as a 1-dimensional subspace of $\mathbf{R}^m$ that does the best job (that a 1-dimensional subspace could do) of approximating all the data:

$$S = \underbrace{\begin{bmatrix} -0.0786 & -0.9969 \\ -0.9969 & 0.0786 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 11039942.91 & 0 \\ 0 & 5711.62 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} -0.0786 & -0.9969 \\ -0.9969 & 0.0786 \end{bmatrix}}_{V^T}. \tag{9}$$

Here the first eigenvalue dominates the second one, indicating the data is very close to a straight line. The straight line is given by the eigenvector corresponding to the large eigenvalue.

Definition 17.3. Let $A \in \mathcal{M}_{m\times n}$.

- For each $1 \leqslant i \leqslant m$, the *variance* of coordinate i is S_{ii}.

A large variance means coordinate i is spread out, and a small variance means coordinate i is densely packed.

- For each $1 \leqslant i, j \leqslant m$, the *covariance* of coordinate i with coordinate j is $S_{ij} = S_{ji}$.

A large positive covariance means coordinate i increases when coordinate j increases, and a large negative covariance means coordinate i decreases when coordinate j increases.

- The *total variance* of A is $\text{trace}(S)$.

Remark 17.4. The variance of our data, as seen in (9), is $\text{trace}(S) = \lambda_1 + \lambda_2 \approx 1.4 \times 10^7$, and the first principal component accounts for $\lambda_1/\text{trace}(S) \approx 0.99$, or about 99% of the total covariance. In general, it may take more than the first principal component to accont for so much of the covariance - your choice of when to stop determines the *princial components* of the data.

The line best approximating the data from Example 17.1 is:

$$\begin{aligned} \text{for mean-centered data: } y &\approx \frac{-0.9969}{-0.0786}x \approx 12.68x \\ \text{for original data: } y &\approx 12.68(x - (\text{mean of } x\text{-values})) + (\text{mean of } y\text{-values}) \\ &\approx 12.68x + 44.37 \end{aligned}$$

Remark 17.5. The first principal component of A solves the *perpendicular least squares* problem. That is, the first eigenvector minimizes the square of the distance from its line to the data. This is alternative to the least squares solution we saw in Lecture 8, which minimized the the vertical distance.

17.2 PCA for higher dimensions

In the first part of this lecture we used only a 2-dimensional data set, and considered only one principal component. More often than not the data we see is many-dimensional, and has more than one important component.

Example 17.6. Consider data given by $A \in \mathcal{M}_{10\times 41}$, representing individual results from decathlon completitions. Since 10 dimensions are very difficult to visualize, we will project the data to the two largest principal components:

$$A = \begin{bmatrix} 11.04 & 10.76 & \cdots & 11.23 & 11.36 \\ 7.58 & 7.40 & \cdots & 6.99 & 6.68 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 291.70 & 301.50 & \cdots & 281.70 & 296.12 \end{bmatrix}$$

However, the each coordinate has different ranges: comparing the data this way gives more wieght to coordinates with a larger range, even if the range has nothing to do with its importance. To compare everything with the same weight, we need to *normalize*.

Definition 17.7. Let $\mathbf{x} \in \mathbf{R}^n$. The *normalization* $\hat{\mathbf{x}}$ of the vector $\mathbf{x}$ means one of two things:

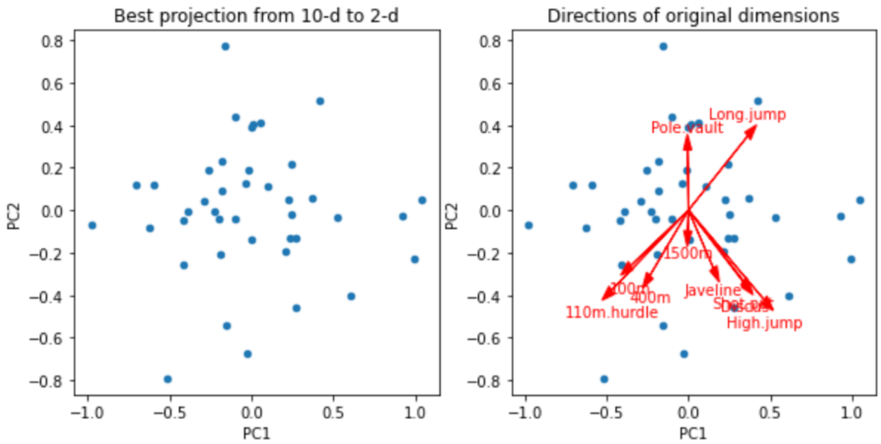
- scale the vector so that it has unit length: $\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$
- shift and scale the vector so that it lies in $[0,1]^n$: $\hat{\mathbf{x}} = \frac{\mathbf{x}-\mathbf{m}}{M-m}$, where $m = \min_i x_i$, $M = \max_i x_i$, and $\mathbf{m} = [m\ m\ \cdots\ m]^T$.

The second case is also called *min-max normalization*, and is used when $\mathbf{x}$ represents parts of many samples, as in our current example.

We normalize each row of A, then center it at its mean. From this matrix we construct the sample covariance matrix S and get its two principal components.

$$A_{norm,\ cent} \approx \begin{bmatrix} 0.03 & -0.2 & \cdots & 0.19 & 0.3 \\ 0.24 & 0.1 & \cdots & -0.2 & -0.43 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0..23 & 0.41 & \cdots & 0.05 & 0.31 \end{bmatrix} \quad \begin{matrix} \lambda_1 = 0.192 \\ \lambda_2 = 0.095 \end{matrix} \quad \mathbf{u}_1 = \begin{bmatrix} -0.35 \\ 0.36 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -0.26 \\ 0.35 \\ \vdots \\ -0.1 \end{bmatrix}.$$

In this case, the two principal components account for only $(\lambda_1 + \lambda_2)/\text{trace}(S) \approx 0.51$ of the total variance. This means that no pairs of coordinates had high covariance.



Projecting each sample to the subspace spanned by two principal components gives a nice spread-out picture of the data. We can also overlay the 10 individual disciplines in the decathlon to see a shadow of the original 10-dimensional space.

17.3 Exercises

Exercise 17.1. Find samples of high-dimensional (at least 4) data online.

1. Construct the sample covariance matrix S and find the two largest eigenvalue / eigenvector pairs from its SVD.
2. What percentage of the total covariance do the first two principal components cover?
3. Plot the data on the axes of the two principal components.
4. Create two plots of the data having for axes:
   (a) the first principal component against the coordinate with the highest (in magnitude) association
   (b) the second principal component against the coordinate with the highest (in magnitude) association

Exercise 17.2. Create a matrix of 2-dimensional data for which the first principal component of the data is a multiple of the eigenvector $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$, for $a, b \in \mathbf{R}_{\neq 0}$. Make sure that:

- the matrix has at least 3 columns (samples),
- no 3 samples are colinear.

Exercise 17.3. 1. Create a matrix of 3-dimensional data for which first two principal components are the vectors $[1\ 0\ 0]^T$ and $[0\ 1\ 0]^T$. Make sure that:

- the data is centered at 0,
- the matrix has at least 4 columns (samples),
- no 3 samples are colinear.

2. Do the same as in part (a), but change the last condition to "no 4 samples lie on a plane."

Lecture 18: Linear transformations

Chapters 8.1 and 8.2 in Strang

- Fact 1: A linear transformation is the same thing as a matrix.
- Fact 2: A linear transformtion is injective iff it is surjective.

- Skill 1: Determine whether or not a function is a linear transformation.
- Skill 2: Construct a matrix for a linear transformation, given what it does to a basis.
- Skill 3: Construct the image and kernel of a linear transformation

---

In this lecture, we will make the connection between $m \times n$ matrices and functions $\mathbf{R}^n \to \mathbf{R}^m$. We have already seen the interpretation of a matrix as a function with the rotation matrix R_θ in Lecture 7. By the end of this lecture, we will see that every such function comes from a matrix.

18.1 Types of linear transformations

Definition 18.1. Let V, W be vector spaces. A *linear transformation*, or *linear map*, is a function $f\colon V \to W$ that satisfies

$$f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) \qquad \text{and} \qquad f(c\mathbf{x}) = cf(\mathbf{x}) \tag{10}$$

for every $\mathbf{x}, \mathbf{y} \in V$ and every $c \in \mathbf{R}$. These are conditions for *linearity*.

Example 18.2. We have already seen examples (and non-examples) of linear transformations:

- Every $m \times n$ matrix is a linear transformation $\mathbf{R}^n \to \mathbf{R}^m$, because $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y}$.
- The dot product of $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ with anything in $\mathbf{R}^2$ is a linear transformation $\mathbf{R}^2 \to \mathbf{R}$, because

$$\begin{bmatrix} a \\ b \end{bmatrix} \cdot (\mathbf{x}+\mathbf{y}) = \begin{bmatrix} a \\ b \end{bmatrix} \cdot \mathbf{x} + \begin{bmatrix} a \\ b \end{bmatrix} \cdot \mathbf{y}.$$

- Differentiation and integration on $C[\mathbf{R}]$ is linear.
- The shift function $\mathbf{x} \mapsto \mathbf{x}+\mathbf{y}$ for nonzero $\mathbf{y}$ is not linear, because splitting up the function on two vectors adds $2\mathbf{y}$ instead of just $\mathbf{y}$.
- The length function is not a linear transformation $\mathbf{R}^n \to \mathbf{R}$, because

$$\left\| \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\| = \left\| \begin{bmatrix} -1 \\ -2 \end{bmatrix} \right\| = \sqrt{3}, \qquad \text{but} \qquad \left\| \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ -2 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\| = 0 \neq 2\sqrt{3}.$$

Proposition 18.3. Any linear map $V \to W$ is completely determined by what it does to the basis of V.

This follows immediately by linearity. Another way to say the above proposition is that choosing a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V and taking any (not necessarily linearly independent!) vectors $\mathbf{w}_1, \ldots, \mathbf{w}_n \in W$, there is only one unique linear map $f\colon V \to W$ for which $f(\mathbf{v}_i) = \mathbf{w}_i$, for all i.

Definition 18.4. Let $f\colon V \to W$ be a linear transformation.

- The *kernel* of f is $\ker(f) = \{\mathbf{x} \in V \ : \ f(\mathbf{x}) = 0\} \subseteq V$
- The *image*, or *range* of f is $\operatorname{im}(f) = \{f(\mathbf{x}) \in W \ : \ \mathbf{x} \in\} \subseteq W$

Note that $\ker(f) \subseteq V$ is a subspace of V, and $\mathrm{im}(f) \subseteq W$ is a subspace of W.

Example 18.5. For $f(\mathbf{x}) = A\mathbf{x}$, multiplication by a matrix, the kernel is the nullspace and the image is the column space. That is,

$$\ker(f) = \mathrm{null}(A), \qquad \mathrm{im}(f) = \mathrm{col}(A).$$

Recall that a function $f : X \to Y$ is *injective*, or *one-to-one*, if $f(a) = f(b)$ implies $a = b$. Further, the function f is *surjective*, or *onto*, if for every $y \in Y$ there exists $x \in X$ such that $f(x) = y$. We will apply these concepts to linear transformations.

Proposition 18.6. Let $f : V \to W$ be linear.

- f is injective iff $\ker(f) = \{0\}$
- if $\dim(W) = \dim(\mathrm{im}(f))$, then f is surjective.

Proof. For the first point, if $\ker(f) = \{0\}$, then we immediately get injectivity. For any $\mathbf{x}, \mathbf{y} \in V$,

$$f(\mathbf{x}) = f(\mathbf{y}) \iff f(\mathbf{x}) - f(\mathbf{y}) = 0 \iff f(\mathbf{x} - \mathbf{y}) = 0 \implies \mathbf{x} - \mathbf{y} = 0 \iff \mathbf{x} = \mathbf{y}.$$

Conversely, suppose that f is injective. If there is some nonzero $\mathbf{z} \in \ker(f)$, then $f(\mathbf{z}) = 0$. But we already know that $f(0) = 0$, so $f(0) = f(\mathbf{z})$, but $0 \neq \mathbf{z}$, violating injectivity.

The second point follows immediately by the fact that $\mathrm{im}(f) \subseteq W$ and by Remark 5.13. □

Definition 18.7. A linear transformation $f : V \to W$ that is both injective and surjective is an *isomorphism*.

You may have seen the word *bijective* be used for functions that are both injective and surjective, but for linear maps we use this special word. Isomorphisms are important because they preserve the fundamental structure of the vector space V.

Example 18.8. We have already seen examples of isomorphisms:

- The map $f : \mathbf{R}^n \to \mathbf{R}^n$ with $f(\mathbf{x}) = 2\mathbf{x}$ is an isomoprhism.
- The change of basis matrix from Lecture 5 is an isomorphism
- The dot product of any vector in $\mathbf{R}^2$ with $(-1, 2)$ is not an isomoprhism, as it fails injectivity: $(3, 4) \cdot (-1, 2) = (-5, 0) \cdot (-1, 2)$.

18.2 The matrix of a linear transformation

Theorem 18.9. *Let $f : \mathbf{R}^n \to \mathbf{R}^m$ be linear. Then there is a unique matrix A for which $A\boldsymbol{x} = f(\boldsymbol{x})$ for all $x \in \mathbf{R}^n$.*

Proof. First we do this proof in a special case, using the standard bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ for $\mathbf{R}^n$ and $\mathbf{e}_1, \ldots, \mathbf{e}_m$ for $\mathbf{R}^m$. By Proposition 18.3, f is completely determined by what it does on the $\mathbf{e}_i$. Suppose that

$$\begin{aligned} f(\mathbf{e}_1) &= a_{11}\mathbf{e}_1 + \cdots + a_{m1}\mathbf{e}_m, \\ f(\mathbf{e}_2) &= a_{12}\mathbf{e}_1 + \cdots + a_{m2}\mathbf{e}_m, \\ &\vdots \\ f(\mathbf{e}_n) &= a_{1n}\mathbf{e}_1 + \cdots + a_{mn}\mathbf{e}_m, \end{aligned}$$

for some $a_{ij} \in \mathbf{R}$. Then on an arbitrary $\mathbf{x} = b_1\mathbf{e}_1 + \cdots + b_n\mathbf{e}_n \in \mathbf{R}^n$, the linear map f takes it to

$$\begin{aligned} f(\mathbf{x}) &= f(b_1\mathbf{e}_1 + \cdots + b_n\mathbf{e}_n) \\ &= b_1 f(\mathbf{e}_1) + \cdots + b_n f(\mathbf{e}_n) \\ &= b_1(a_{11}\mathbf{e}_1 + \cdots + a_{m1}\mathbf{e}_m) + \cdots + b_n(a_{1n}\mathbf{e}_1 + \cdots + a_{mn}\mathbf{e}_m) \\ &= (b_1 a_{11} + \cdots + b_n a_{1n})\mathbf{e}_1 + \cdots + (b_1 a_{m1} + \cdots b_n a_{mn})\mathbf{e}_m. \end{aligned}$$

Since $\mathbf{e}_i$ is all zeros except a 1 on line i, the last line above can be rewritten as

$$\begin{bmatrix} b_1a_{11} + \cdots + b_na_{1n} \\ b_2a_{21} + \cdots + b_na_{2n} \\ \vdots \\ b_1a_{m1} + \cdots b_na_{mn} \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{A} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = A(b_1\mathbf{e}_1 + \cdots + b_n\mathbf{e}_n) = A\mathbf{x}.$$

So in this case, f is exactly A.

In the general case, where $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is some basis for $\mathbf{R}^n$ and $\mathbf{w}_1, \ldots, \mathbf{w}_m$ is some basis for $\mathbf{R}^m$, construct the change of basis matrices C_V, that takes the v_i to the $\mathbf{e}_i$, and C_W, that takes the $\mathbf{w}_i$ to the e_i. Then the matrix of the function f is $C_W^{-1}AC_V$. □

This statement has several implications. Combining the rank-nullity theorem from Lecture 6 along with observations above, we immediately get the following.

Corollary 18.10. Let $f\colon V \to W$ be linear, with $\dim(V) = \dim(W)$.

- [Dimension Theorem] $\dim(V) = \dim(\ker(f)) + \dim(\operatorname{im}(f))$
- The map f is surjective iff it is injective

Proof. The first point follows by the rank-nullity theorem and applying Theorem 18.9 in Example 18.5 to describe every linear map as a matrix.

The second point follows immediately from the first point and Proposition 18.6. □

Remark 18.11. We also get a nice result for compositions of linear maps. Given two linear maps $f\colon V \to W$ and $g\colon W \to Z$, their *composition* is a linear map $(g \circ f)\colon V \to Z$ (you will check this in an exercise). If f, g have associated matrices A, B, respectively, then the composition $g \circ f$ has associated matrix BA. This follows by using the equations $f(\mathbf{x}) = A\mathbf{x}$ and $g(\mathbf{y}) = B(\mathbf{y})$ in simplifying

$$(g \circ f)(\mathbf{x}) = g(f(\mathbf{x})) = g(A\mathbf{x}) = B(A\mathbf{x}) = (BA)\mathbf{x}.$$

18.3 Exercises

Exercise 18.1. Consider the following transformations T_i:

$$T_1 \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} w \\ y \\ z \\ x \end{bmatrix} \qquad T_2 \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2e^y \\ x \end{bmatrix} \qquad T_3 \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x^2 \\ y^2 \end{bmatrix} \qquad T_4 \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sin(x^2 + y^2) \\ \cos(x^2 + y^2) \end{bmatrix}$$

$$T_5 \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3y + x \\ 0 \\ x^2 - y \end{bmatrix} \qquad T_6 \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad T_7 \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -3x \\ z + y \end{bmatrix} \qquad T_8 \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2x + 2y \\ y + z \\ 0 \end{bmatrix}$$

1. Which of the T_i are linear? For those that are not, give a counterexample in which one of the linearity conditions fail. For those that are, give the associated matrix.
2. Let $S\colon \mathbf{R}^3 \to \mathbf{R}^3$ be the linear transformation for which

$$ST_5 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad ST_8 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \qquad ST_8 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Construct the 3×3 matrix of S.

Exercise 18.2. Prove the claim from Definition 18.4 that the kernel and image of $f\colon V \to W$ are subspaces of V and W, respectively. Use linearity to check the vector space conditions.

Exercise 18.3. Let $f\colon V \to W$ be a linear transformation, and let $v_1, \dots, v_n$ be a basis of V. Show that f is injective iff the set of vectors $f(\mathbf{v}_1), \dots, f(\mathbf{v}_n) \subseteq W$ is linearly independent.

Exercise 18.4. Consider the three orthogonal vectors

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \qquad \mathbf{z} = \begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix}.$$

1. Find the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$.

2. Construct a symmetric matrix A of full rank for which the unit vectors from part (a) are eigenvectors.

3. Let $f\colon \mathbf{R}^3 \to \mathbf{R}^3$ be the linear transformation for which

$$f(\mathbf{x}) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad f(\mathbf{y}) = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}, \qquad f(\mathbf{z}) = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

   Construct the 3×3 matrix for f.

Exercise 18.5. Let V be the vector space of polynomials in two variables x and y of degree at most 2. This space has dimension 6, and has basis with basis $1, x, y, x^2, y^2, xy$. Let $L\colon V \to V$ be the linear operator defined by $L(f(x, y)) = f(x - y, y - x)$.

1. Find the matrix of L using the basis specified.

2. Find a basis for the image and kernel of L.

Exercise 18.6. Prove the claim from Remark 18.11 that the composition of two linear maps is linear.

Lecture 19: Jordan form

Chapter 8.3 in Strang

- Fact 1: Every square matrix is similar to a square matrix in Jordan normal form.
- Fact 2: Jordan normal form generalizes diagonalization by introducing generalized eigenvectors.

- Skill 1: Construct the Jordan normal form of a square matrix.
- Skill 2: Find the higher rank generalized eigenvectors when geometric multiplicity exceeds algebraic multiplicity.

In this section we will see yet another decomposition, the Jordan normal form. This will be a more general decomposition for square matrices, giving us a nice result even when the matrix is not full rank or has repeated eigenvalues.

19.1 Jordan blocks and generalized eigenvectors

Previously in Example 14.4 in Lecture 14 we saw the idea of *multiplicity*. We now revisit it and give it two similar but related meanings.

Definition 19.1. Let $A \in \mathcal{M}_{m\times n}$ have characteristic polynomial $\chi(\lambda)$. For λ_i an eigenvalue of A:

- the exponent m of the factor $(\lambda - \lambda_i)^m$ of χ is the *algebraic multiplicity* of λ_i
- the number of linearly independent eigenvectors of A having λ_i as an eigenvalue is the *geometric multiplicity* of λ_i.

Example 19.2. Here we show examples of matrices with different algebraic and geometric multiplicities. Consider the matrices

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 2 \\ 0 & 2 \end{bmatrix}.$$

Both matrices have eigenvalue $\lambda = 2$. Here A has algebraic multiplicity and geometric multiplicity 2 of λ. The matrix B has algebraic multiplicity 2 but geometric multiplicity 1 of λ. However, not all combinations of these two numbers are possible: whenever λ is an eigenvalue, we have

$$1 \leqslant (\text{geometric multiplicity of } \lambda) \leqslant (\text{algebraic multiplicity of } \lambda) \leqslant \text{rank}(A).$$

The inequality on the left follows as having an eigenvalue λ means $\det(A-\lambda I) = 0$, which, by Exercise 11.4 is equivalent to saying $(A - \lambda I)\mathbf{x} = 0$ for some $\mathbf{x}$. Hence we always have at least one eigenvector for every eigenvalue.

Definition 19.3. Let $A \in \mathcal{M}_{n\times n}$. The *Jordan normal form* of A is the matrix

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix},$$

where every $J_1, \ldots, J_s$ is a *Jordan block*. A Jordan block is a matrix

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix},$$

where λ_i is an eigenvalue of A. For every i, the number of Jordan blocks with eigenvalue λ is the geometric multiplicity of λ. Jordan normal form is also known under the names *normal form* or *Jordan canonical form*.

To get the size of each Jordan block, we need to do some more work.

Remark 19.4. Note that the eigenvalue / eigenvector pairs of A are the same as those of J. Indeed, consider the matrix in Jordan normal form

$$J = \begin{bmatrix} 2 & 1 & & \\ & 2 & & \\ & & 1 & \\ & & & 0 \end{bmatrix}, \qquad \det(J - \lambda I) = (2-\lambda)^2(1-\lambda)\lambda.$$

Both the matrix $A \in \mathcal{M}_{4\times 4}$ that J represents and J itself have three eigenvalues, $\lambda_1 = 2$ (with multiplicity 2), $\lambda_2 = 1$, $\lambda_3 = 0$. The eigenvectors of J are

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

The vector that is naturally missing from this list is $\mathbf{u}_1^* = [0\ 1\ 0\ 0]^T$, which seems like it should be also associated to λ_1. We call this the *generalized eigenvector*. It is clear that $J\mathbf{u}_1^* \neq 0$, but we do have

$$(J - 2I)\mathbf{u}_1^* = \mathbf{u}_1 \qquad \text{or} \qquad (J - 2I)^2\mathbf{u}_1^* = 0.$$

Definition 19.5. Let $A \in \mathcal{M}_{n\times n}$ with Jordan normal form J and eigenvalue λ. A vector $\mathbf{x} \in \mathbf{R}^n$ is a *generalized eigenvector* associated to λ if $(J - \lambda I)^k\mathbf{x} = 0$ for some $k > 1$. If additionally we have $(J - \lambda I)^{k-1}\mathbf{x} \neq 0$, then k is the *rank* of the generalized eigenvector $\mathbf{x}$.

Remark 19.6. Every Jordan block has one (rank 1) generalized eigenvector, and the rest have higher rank. If a Jordan block of the eigenvalue λ has size m, then there is a *cycle* of generalized eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$, where u_i has rank i and

$$(J - \lambda I)\mathbf{u}_{i+1} = \mathbf{u}_i \qquad \text{or} \qquad (J - \lambda I)^{m-i}\mathbf{u}_m = \mathbf{u}_i \tag{11}$$

for all $i = 1, \ldots, m$. This implies that $(J - \lambda I)^i\mathbf{u}_i = 0$ and $(J - \lambda I)^{i-1}\mathbf{u}_i \neq 0$.

Example 19.7. If we have the eigenvector with the highest rank in the cycle, we can generate the others. Consider

$$J = \begin{bmatrix} 2 & 1 & & & & & \\ & 2 & & & & & \\ & & 2 & 1 & & & \\ & & & 2 & 1 & & \\ & & & & 2 & 1 & \\ & & & & & 2 & 1 \\ & & & & & & 2 \end{bmatrix},$$

which has only one eigenvalue $\lambda = 2$, with algebraic multiplicity 7, geometric multiplicity 2 and two Jordan blocks associated to it. The rank 1 eigenvectors are

$$\mathbf{u}_1 = [1\ 0\ 0\ 0\ 0\ 0\ 0]^T, \qquad \mathbf{v}_1 = [0\ 0\ 1\ 0\ 0\ 0\ 0]^T.$$

To get the higher rank generalized eigenvectors, we check the positions of the 1's above the diagonal. It is immediate that

$$\mathbf{u}_2 = [0\ 1\ 0\ 0\ 0\ 0\ 0]^T, \qquad \mathbf{v}_2 = [0\ 0\ 0\ 1\ 0\ 0\ 0]^T, \quad \cdots \quad \mathbf{v}_5 = [0\ 0\ 0\ 0\ 0\ 0\ 1]^T.$$

The relationship is also $(J - 2I)\mathbf{v}_5 = \mathbf{v}_4$, $(J - 2I)^2\mathbf{v}_5 = \mathbf{v}_3$, and so on.

19.2 Similar matrices and Jordan's theorem

We have already seen *similar matrices* in Remark 14.7 in Lecture 14. A matrix B is similar to a matrix A if there exists an invertible matrix C with $B = C^{-1}AC$. Here we revisit the idea, making precise the relationship between the matrices A and J.

Remark 19.8. Similar matrices do not have the same eigenvectors, but they do have the same eigenvalues. The eigenvectors of similar matrices are related: If $B = C^{-1}AC$ has eigenvector $\mathbf{x}$ with eigenvalue λ, then

$$B\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad C^{-1}AC\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad A(C\mathbf{x}) = \lambda(C\mathbf{x}).$$

That is, $C\mathbf{x}$ is an eigenvector of A with eigenvalue λ.

Now we combine similr matrices with generalized eigenvectors. Fortunately, generalized eigenvectors apply to any matrix, not just matrices in Jordan form.

Example 19.9. Consider the following matrix A, which has a single eigenvalue $\lambda = 6$ with algebraic multiplicity 4 and geometric multiplicity 1:

$$A = \begin{bmatrix} 9 & -1 & -1 & -3 \\ -3 & 5 & 1 & 1 \\ 5 & -5 & 5 & -9 \\ 3 & 1 & -1 & 5 \end{bmatrix}, \qquad \mathbf{u}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

How do we find its generalized eigenvectors? We apply equation (11) from Remark 19.6 above:

$$(A - 6I)\mathbf{u}_2 = \mathbf{u}_1 \quad \Longleftrightarrow \quad \begin{bmatrix} 3 & -1 & -1 & -3 \\ -3 & -1 & 1 & 1 \\ 5 & -5 & -1 & -9 \\ 3 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$$

$$\Longleftrightarrow \quad \begin{bmatrix} 3 & -1 & -1 & -3 & 1 \\ -3 & -1 & 1 & 1 & -1 \\ 5 & -5 & -1 & -9 & 1 \\ 3 & 1 & -1 & -1 & 1 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and so $\mathbf{u}_2 = \begin{bmatrix} 1 & -1 & 0 & 1 \end{bmatrix}^T$. We similarly solve $(A - 6I)\mathbf{u}_3 = \mathbf{u}_2$ and $(A - 6I)\mathbf{u}_4 = \mathbf{u}_3$ to get the matrix $B \in \mathcal{M}_{4\times 4}$, which has the generalized eigenvectors as its columns. Moreover, we notice that

$$B^{-1}AB = \underbrace{\begin{bmatrix} -3 & -1 & 1 & 2 \\ 5 & 1 & -1 & -3 \\ -2 & -4 & 0 & -2 \\ 0 & 4 & 0 & 4 \end{bmatrix}}_{B^{-1}} \underbrace{\begin{bmatrix} 9 & -1 & -1 & -3 \\ -3 & 5 & 1 & 1 \\ 5 & -5 & 5 & -9 \\ 3 & 1 & -1 & 5 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1 & 1 & \frac{1}{2} & \frac{1}{2} \\ -1 & -1 & -1 & -\frac{3}{4} \\ 1 & 0 & -\frac{3}{2} & -\frac{5}{4} \\ 1 & 1 & 1 & 1 \end{bmatrix}}_{B} = \begin{bmatrix} 6 & 1 & 0 & 0 \\ 0 & 6 & 1 & 0 \\ 0 & 0 & 6 & 1 \\ 0 & 0 & 0 & 6 \end{bmatrix},$$

which is the Jordan form of A. This leads us to this lecture's big theorem.

Theorem 19.10 (Jordan)**.** *For every $A \in \mathcal{M}_{n\times n}$, there exists an invertible $B \in \mathcal{M}_{n\times n}$ such that $J = B^{-1}AB$ is in Jordan normal form. The matrix B has the generalized eigenvectors of A as columns.*

Remark 19.11. Let $A \in \mathcal{M}_{n\times n}$ with $J = B^{-1}AB$ in Jordan normal form, and let C be similar to A. That is, there exists some $D \in \mathcal{M}_{n\times n}$ with $C = DAD^{-1}$. It follows that

$$J = B^{-1}AB = B^{-1}(D^{-1}CD)B = (DB)^{-1}C(DB),$$

so C has the same Jordan normal form as A.

Remark 19.12. If the eigenvectors of A are all linearly independent, then the Jordan normal form $J = \Lambda$ of A will be diagonal with eigenvalues on its diagonal, and the generalized eigenvectors in $B = X$ will all be of rank 1. Then the Jordan normal form decomposition becomes the $X\Lambda X^{-1}$ decomposition from Lecture 14.

19.3 Exercises

Exercise 19.1. How many different matrices in $\mathcal{M}_{7\times 7}$, up to similarity, are there with one eigenvalue $\lambda = 2$ that has algebraic multiplicity 2 and

1. geometric multiplicity 2?
2. geometric multiplicity 3?
3. any geometric multiplicity?

Exercise 19.2. Let $J \in \mathcal{M}_{6\times 6}$ be a matrix in Jordan form with two eigenvalues 3 (having algebraic multiplicity 4 and geometric multiplicity 2) and -3 (having algebraic multiplicity 2 and geometric multiplicity 1).

1. How many Jordan blocks will J have? Give the two possibilities for their sizes.
2. Suppose that the Jordan blocks of J all have the same size. Find a matrix B that is similar to J and has no zero entries.
3. For the matrix B from part (b), find all its generalized eigenvectors.

Lecture 20: Complex numbers and complex matrices

Chapters 9.1 and 9.2 in Strang

- Fact 1: All that we have done so far can be considered over $\mathbf{C}$ instead of $\mathbf{R}$
- Fact 2: Complex number adition and multiplication have geomtric meaning

---

- Skill 1: Express a complex number in one of four different ways
- Skill 2: Apply the new results for Hermitian vectors and matrices

---

In this lecture we will take some time to introduce fully the topic of coplex numbers. The goal is to get a better feel for them - almost all the results we have seen so far apply to them as well! And there are more nicer consequences.

20.1 The space of complex numbers

Definition 20.1. The *complex numbers* are elements of the set $\mathbf{C} = \{x+yi \ : \ x, y \in \mathbf{R}\}$. The symbol i is the *imaginary number*, having the property that $i^2 = -1$.

Let $z = x + yi$ and $w = a + bi$ be complex numbers. Addition and multiplication are defined in the following way:

$$z + w = (a + x) + (y + b)i$$
$$zw = xa + xbi + yai + ybi^2 = (xa - yb) + (xb + ya)i$$

A complex number z written as $x + yi$ is in *standard form*, and written as (x, y) is in *Cartesian coordinates*. The *real part* of z is x and the *imaginary part* of z is y. If $x = 0$, then z is a *purely imaginary number*.

Example 20.2. What does the complex number $(1 + i)^{-2}$ look like in standard form? Observe that

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

Definition 20.3. Let $z = x + yi \in \mathbf{C}$. The (*complex*) *conjugate* of z is $\overline{z} = z^* = x - yi$. The *absolute value*, or *modulus* of z is

$$|z| = \sqrt{z\overline{z}} = \sqrt{(x+yi)(x-yi)} = \sqrt{x^2+y^2}.$$

Proposition 20.4. Let $z = x + yi, w = a + bi \in \mathbf{C}$. Then the conjugate satisfies:

1. $\overline{z+w} = \overline{z} + \overline{w}$
2. $\overline{zw} = \overline{z}\,\overline{w}$
3. $\overline{\overline{z}} = z$
4. $z + \overline{z} = 2x$
5. $z - \overline{z} = 2yi$
6. $z^{-1} = \overline{z}/|z|^2$ for $z \neq 0$

And the absolute value satisfies:

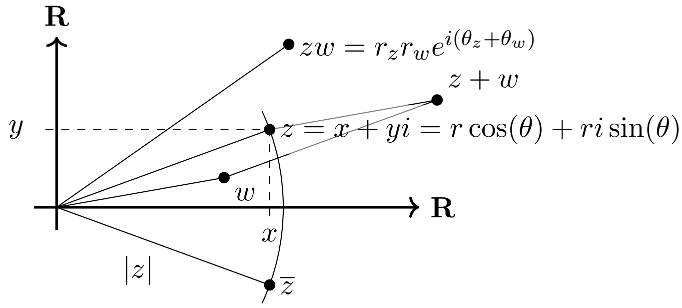
1. $|z| = 0$ iff $z = 0$
2. $|\overline{z}| = |z|$
3. $|zw| = |z||w|$
4. $|z + w| \leqslant |z| + |w|$

Definition 20.5. The third way to express $z = x + yi \in \mathbf{C}$ is with *polar coordinates* (r, θ), where $r = |z|$ and θ is the angle from the positive x axis to the vector (x, y). Note that

$$x + yi = r\cos(\theta) + ri\sin(\theta) = re^{i\theta},$$

where the second equality is known as *Euler's formula*. This last expression is in *exponential form*.

Remark 20.6. All that we have seen so far about the complex numbers, and a new observation about multiplying complex numbers, can be drawn together in a picture.



Example 20.7. Let's compute complex numbers in different forms.

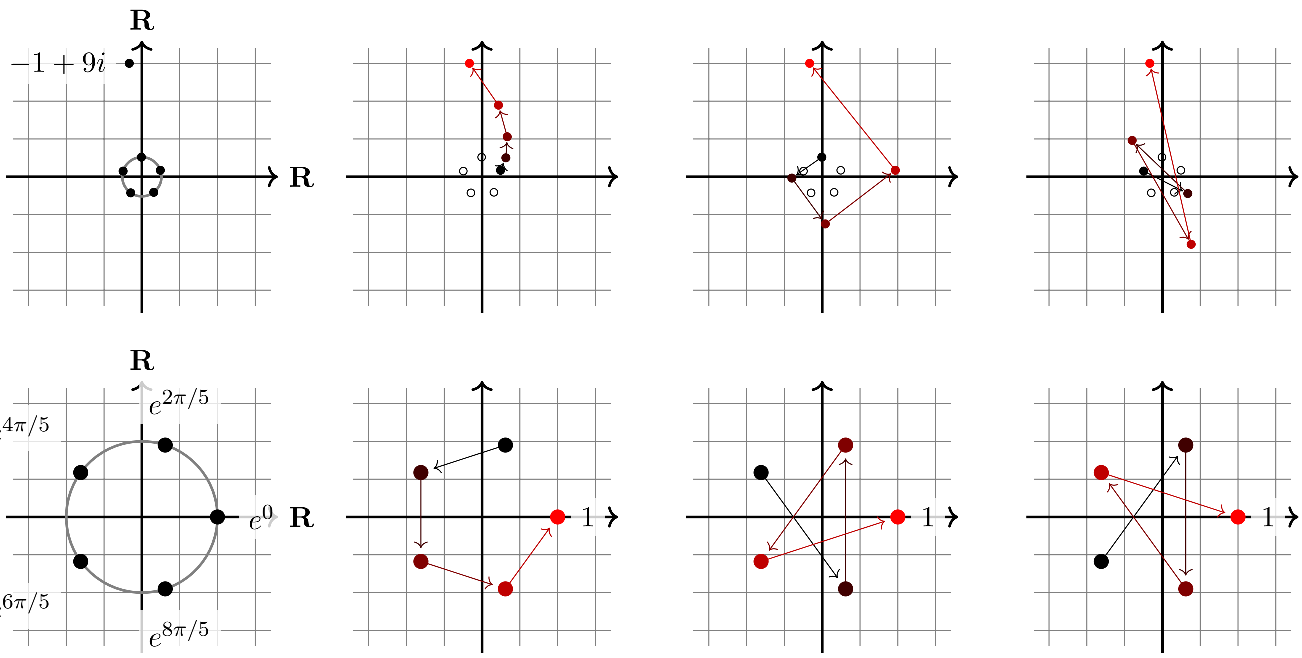
- $z = 5\cos(\pi/4) + 5i\sin(\pi/4)$ in standard form: $z = \frac{5\sqrt{2}}{2} + \frac{5\sqrt{2}}{2}i$
- $w = -\sqrt{3} - i$ in polar form: $w = 2\cos(\pi + \tan(1/3)) + 2i\sin(\pi + \tan(1/3))$
- the 4th roots of $p = 1+i$ in Cartesian coordinates

Remark 20.8. Putting complex numbers into polar coordinates makes computations in standard form much easier. For $z = re^{i\theta}$, we have:

- (De Moivre's theorem) $z^n = (re^{i\theta})^n = r^n e^{in\theta}$
- (complex roots) the nth roots of z are $r^{1/n}e^{i(\theta+2k\pi)/n}$, for every $k = 1, \dots, n-1$.

For the zecond point, when $z = 1+0i$, then the kth root of z is called the *kth root of unity*.

Example 20.9. Below are given the 5th roots of $z = -1+9i$ and the 5th roots of $z = e^0 = 1$, or unity. For some 5th roots ω of z, the complex numbers $\omega, \omega^2, \omega^3, \omega^4, \omega^5 = z$ are also shown. The circle with radius $\sqrt[5]{|z|}$ is given to emphasize that all 5th roots are the same distance from 0.



Remark 20.10. The space of complex numbers is a 2-dimensional vector space over **R** via the identification of Cartesian coordinates. However, it is a 1-dimensional vector space over **C**.

20.2 Complex matrices

Definition 20.11. Let $\mathbf{z} = [z_1 \ \cdots \ z_n]^T \in \mathbf{C}^n$ be a vector. The (*complex*) *conjugate* is the vector $\overline{\mathbf{z}} = [\overline{z_1} \ \cdots \ \overline{z_n}]^T$.

Often we talk about not just the conjugate, but the *conjugate transpose*. The reason for taking both the conjugate of each element and the transpose, when $n = 2$ and $\mathbf{z} = [\begin{smallmatrix} x \\ y \end{smallmatrix}] = x + yi = z$, is to get that

$$\overline{\mathbf{z}}^T \mathbf{z} = \mathbf{z}^* \mathbf{z} = \|\mathbf{z}\|^2 = |z|^2 = \overline{z} z,$$

so the previous notion of length of a vector corresponds with the new notion of absolute value of a complex number. The notation $\mathbf{z}^* = \overline{\mathbf{z}}^T$ is also used for matrices, so that $(A^*)_{ij} = \overline{A_{ji}}$.

Definition 20.12. Let $A \in \mathcal{M}_{n\times n}(\mathbf{C})$. Then

- A is *Hermitian* if $A = A^*$
- A is *unitary* if the columns of A are orthonormal

Proposition 20.13. Let $A \in \mathcal{M}_{n\times n}(\mathbf{C})$ and $\mathbf{z} \in \mathbf{C}^n$. If A is Hermitian, then:

- $\mathbf{z}^* A \mathbf{z}$ is a real number
- every eigenvalue of A is a real number
- eigenvectors (of different eigenvalues are orthogonal

If A is unitary, then:

- $A^* A = I$ and $A^{-1} = A^*$
- every eigenvalue of A is ± 1

Example 20.14. Consider the 2×2 matrix $A = \left[\begin{smallmatrix} 2 & 3-3i \\ 3+3i & 5 \end{smallmatrix}\right]$. This matrix is Hermitian, so should have real eigenvalues and orthogonal eigenvectors by the previous Proposition. Indeed, we find that

$$\det(A - \lambda I) = (2-\lambda)(5-\lambda) - (3-3i)(3+3i) = 10 - 7\lambda + \lambda^2 - 18 = \lambda^2 - 7\lambda - 8 = (\lambda - 8)(\lambda + 1),$$

so the eigenvalues are $\lambda = 8, -1$. For the eigenvectors, we must solve

$$\begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} 8z \\ 8w \end{bmatrix} \quad \iff \quad \begin{aligned} -6z + (3-3i)w &= 0, \\ (3+3i)z - 3w &= 0. \end{aligned}$$

Using the first equation to isolate w, we get

$$w = \frac{6z}{3-3i} = \frac{6z}{3-3i}\frac{3+3i}{3+3i} = \frac{(18+18i)z}{9+9} = (1+i)z,$$

which, when placed into the second equation, gives us $(3+3i)z - 3(1+i)z = 0$, which means there are no constraints on z. So we let $z = 1$ and $w = 1 + i$. Similarly for the second eigenvector we find $z = 2$ and $w = -1 - i$. To check they are orthogonal, we observe that

$$\begin{bmatrix} 1+i \\ 1 \end{bmatrix}^* \cdot \begin{bmatrix} -1-i \\ 2 \end{bmatrix} = \begin{bmatrix} 1-i & 1 \end{bmatrix} \begin{bmatrix} -1-i \\ 2 \end{bmatrix} = (1-i)(-1-i) + 2 = -1 - i + i + i^2 + 2 = -2 + 2 = 0,$$

and we have orthogonality, as desired.

Example 20.15. Consider the *Fourier matrix*, which is a unitary matrix:

$$F = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{2\pi i/3} & e^{4\pi i/3} \\ 1 & e^{4\pi i/3} & e^{2\pi i/3} \end{bmatrix}$$

The columns contain the *cube roots of unity*. More specifically, $F_{ij}^3 = 1$ and

20.3 Exercises

Exercise 20.1. Show that every complex number $z = x + yi$ for which at least one of x and y are not zero has an inverse. That is, find $w \in \mathbf{C}$ for which $zw = 1$.

Exercise 20.2. Prove all the claims of Proposition 20.4, for $z = x + yi, w = a + bi \in \mathbf{C}$:

1. $\overline{z + w} = \overline{z} + \overline{w}$
2. $\overline{zw} = \overline{z}\,\overline{w}$
3. $\overline{\overline{z}} = z$
4. $z + \overline{z} = 2x$
5. $z - \overline{z} = 2yi$
6. $z^{-1} = \overline{z}/|z|^2$ for $z \neq 0$
7. $|z| = 0$ iff $z = 0$
8. $|\overline{z}| = |z|$
9. $|zw| = |z||w|$
10. $|z + w| \leqslant |z| + |w|$

Exercise 20.3. Prove Euler's formula $\cos(\theta) + i\sin(\theta) = e^{i\theta}$ is true by showing that the derivative of $(\cos(\theta) + i\sin(\theta))e^{-i\theta}$ is zero.

Part IV
Extensions

Lecture 21: Fourier topics

Chapters 8.3 and 9.3 in Strang's "Linear Algebra" and IV.1 in Strang's "Learning from Data"

- Fact 1: Every piecewise continuous function on a finite interval can be approximated by sines and cosines
- Fact 2: The discrete Fourier transform makes this approximation faster by using matrix multipliation instead of integration

---

- Skill 1: Express the Fourier coefficients of a piecewise continuous function on $[0, 2\pi]$
- Skill 2: Construct the discrete Fourier transform of evenely-spaced data points.

This lecture is all about things named after Fourier: the Fourier basis for the space of 2π-periodic, the Fourier series for expressing any 2π-periodic function using this basis, and the discrete Fourier transform, which extends this approach to functions which are not completely known.

21.1 The Fourier basis and the Fourier series

Recall that a function $f\colon \mathbf{R} \to \mathbf{R}$ is *piecewise continuous* if f is continuous at all except finitely many points of $\mathbf{R}$. We consider the space $PC[0, 2\pi]$ of piecewise continuous functions defined on $[0, 2\pi]$, using the inner product

$$\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\ dx.$$

We can integrate piecewise continuous functions just like continuous functions, by applying linearity and splitting them up over intervals where they are continuous.

In the first part of this lecture we consider the *Fourier basis* (yet to be shown that it is a basis) $F = \{1\} \cup \{\sin(nx)\ :\ n \in \mathbf{N}\} \cup \{\cos(nx)\ :\ n \in \mathbf{N}\} \subseteq PC[0, 2\pi]$.

Proposition 21.1. The set F is orthogonal.

Proof. It is immediate that the function 1 is orthogonal to all other functions. Indeed,

$$\int_0^{2\pi} \sin(nx)\ dx = \frac{1}{n}\int_0^{2\pi n} \sin(u)\ du = \left.\frac{-\cos(u)}{n}\right|_{u=0}^{u=2\pi n} = \frac{1-\cos(2\pi n)}{n} = 0,$$

as $\cos(2\pi n) = 1$ for all $n \in \mathbf{N}$, and

$$\int_0^{2\pi} \cos(nx)\ dx = \frac{1}{n}\int_0^{2\pi n} \cos(u)\ du = \left.\frac{\sin(u)}{n}\right|_{u=0}^{u=2\pi n} = \frac{\sin(2\pi n)}{n} = 0,$$

as $\sin(2\pi n) = 0$ for all $n \in \mathbf{N}$. To show that the $\cos(nx)$ functions are orthogonal, and that the $\sin(nx)$ functions are orthogonal, we use the sum of angles formula:

$$\begin{aligned}\cos(\theta \pm \varphi) = \cos(\theta)\cos(\varphi) \mp \sin(\theta)\sin(\varphi) \quad &\Longrightarrow \quad \cos(\theta)\cos(\varphi) = \frac{1}{2}\left(\cos(\theta+\varphi) + \cos(\theta-\varphi)\right)\\ &\Longrightarrow \quad \sin(\theta)\sin(\varphi) = \frac{1}{2}\left(\cos(\theta-\varphi) - \cos(\theta+\varphi)\right).\end{aligned}$$

Using substitution we can solve the intergals. To show $\sin(nx)$ is orthogonal to $\cos(mx)$, we use the sum of angles formula for sin. ☐

To finish justifying that F is a basis for $PC[0, 2\pi]$, we need to show that F spans this set. Such a proof is beyond the scope of this course, so we continue with the assumption that F is a basis for $PC[0, 2\pi]$.

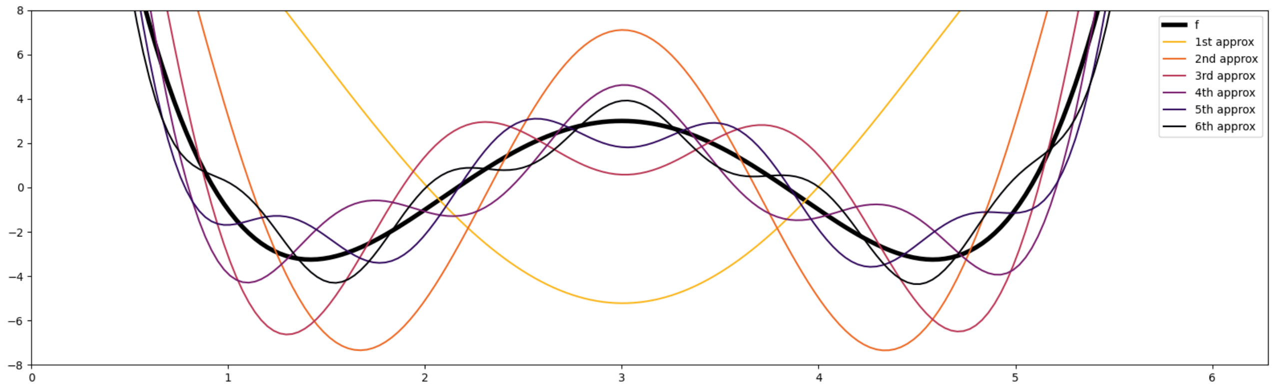
Definition 21.2. Let $f \in PC[0, 2\pi]$. Expressing f using the basis F is the *Fourier series* of f:

$$\begin{aligned} f(x) &= a_0 + a_1 \sin(x) + b_1 \cos(x) + a_2 \sin(2x) + b_2 \cos(2x) + \cdots \\ &= a_0 + \sum_{n=1}^{\infty} (a_n \sin(nx) + b_n \cos(nx)). \end{aligned}$$

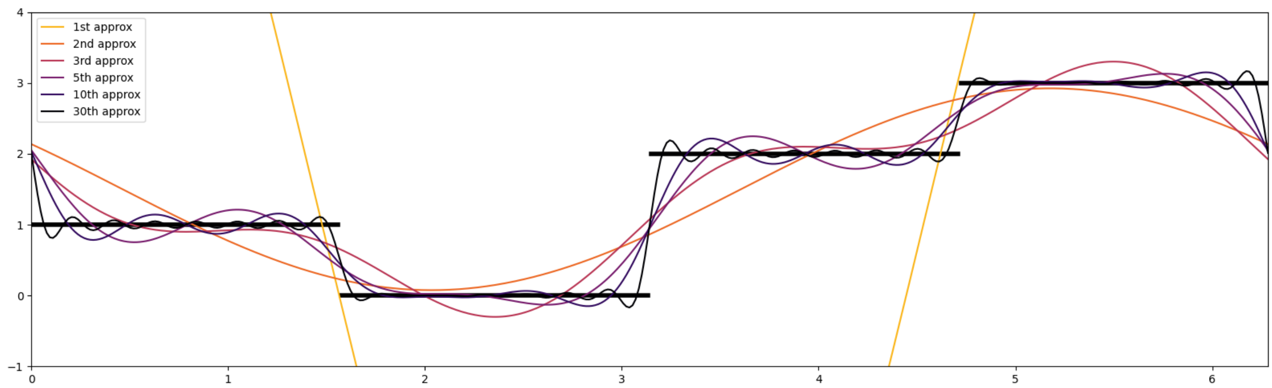
The numbers a_n and b_n are the projections of f onto the vectors spanned by $\sin(nx)$ and $\cos(nx)$, respectively. They are the *Fourier coefficients* of f:

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x)\, dx, \qquad a_n = \frac{\langle f(x), \sin(nx) \rangle}{\|\sin(nx)\|^2}, \qquad b_n = \frac{\langle f(x), \cos(nx) \rangle}{\|\cos(nx)\|^2}.$$

Example 21.3. Unless f is very nice, the sum is usually infinite. Hence we often give only the first few terms in the series to describe f. Here are the first 6 pairs of Fourier coefficients for a simple degree 4 polynomial.



This may seem like overkill, but it is very useful when the orginal function is not continuous everywhere. The Fourier series of any piecewise continuous function will be continuous (and differentiable!) everywhere.



Remark 21.4. The description of F given initially is a bit cumbersome. We can make it simpler using the definition of sin and cos in terms of the exponential function:

$$\sin(\theta) = \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{ie^{-i\theta} - ie^{i\theta}}{2}, \qquad \cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2}.$$

Let $F_e = \{e^{inx} \ : \ n \in \mathbf{Z}\}$. It is immediate that $\text{span}(F) = \text{span}(F_e)$, and checking for orthogonality requires only one step, instead of five. If we have a Fourier series with terms a_0, a_1, a_2, b_1, b_2, then we can express is in the basis F_e as

$$\begin{aligned} f(x) &= a_0 + a_1 \sin(x) + b_1 \cos(x) + a_2 \sin(2x) + b_2 \cos(2x) \\ &= a_0 + a_1 \left(\frac{ie^{-ix} - ie^{ix}}{2} \right) + +b_1 \left(\frac{e^{ix} + e^{-ix}}{2} \right) + a_2 \left(\frac{ie^{-i2x} - ie^{i2x}}{2} \right) + b_2 \left(\frac{e^{i2x} + e^{-i2x}}{2} \right) \\ &= a_0 e^{i0x} + \left(\frac{b_1}{2} - \frac{ia_1}{2} \right) e^{ix} + \left(\frac{b_1}{2} + \frac{ia_1}{2} \right) e^{i(-1x)} + \left(\frac{b_2}{2} - \frac{ia_2}{2} \right) e^{i2x} + \left(\frac{b_2}{2} + \frac{ia_2}{2} \right) e^{i(-2x)}. \end{aligned}$$

Note that the coefficients are now complex numbers.

21.2 The Fourier matrix and the discrete Fourier transform

In Example 21.3 above, we had two functions that were completely known. In the real world, we do not know completely the function we are considering, but only know its value at certain inputs x. A very pertinent queston is then how to convert this discrete data into a continuous function.

Example 21.5. Suppose we have the following data points on the interval $[0, 2\pi]$, evenly spaced out. This could be only part of a signal that we can pick up, or a very sparsely sampled sound:

$$(0, 1), \qquad \left(\frac{\pi}{2}, 2 \right), \qquad (\pi, -2), \qquad \left(\frac{3\pi}{2}, -3 \right).$$

How can we make this data into a continuous function? We could apply the approach from Example 21.3, but we would be assuming the values of the signal at unknown points, and there are several natural ways to extend the discrete signal into a continuous signal.

Definition 21.6. The $n \times n$ *Fourier matrix* $F_n \in \mathcal{M}_{n\times n}(\mathbf{C})$ has $n(F_n)_{ij} = \omega^{(i-1)(j-1)}$, where $\omega = e^{-2\pi i/n}$ is an nth root of unity:

$$F_n = \frac{1}{n} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix}.$$

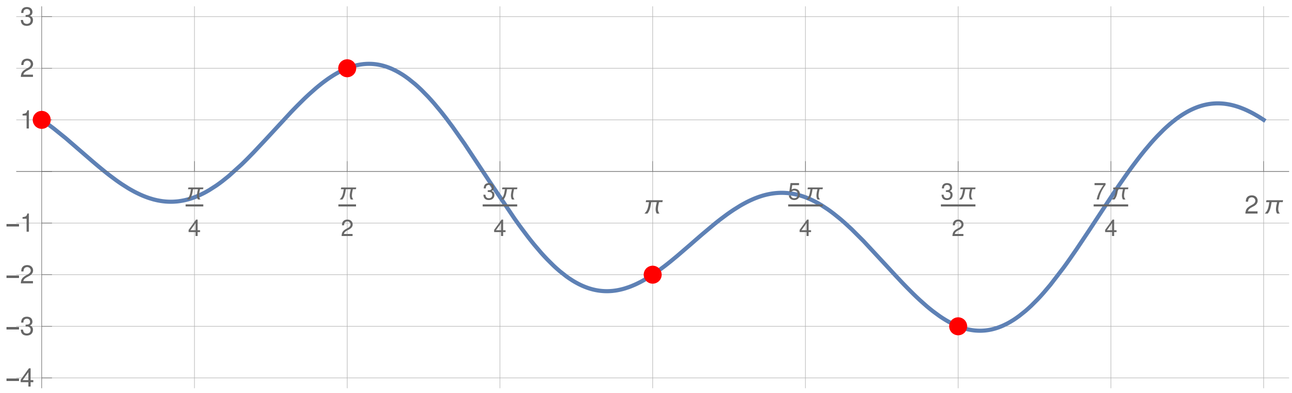
Given $\mathbf{x} \in \mathbf{R}^n$, the vector $F_n\mathbf{x} \in \mathbf{C}^n$ is called the *discrete Fourier transform* of $\mathbf{x}$.

Remark 21.7. This matrix may look familiar - it is the *Vandermonde* matrix from Definition 9.2 in Lecture 9, for $x_1, \ldots, x_n$ the nth roots of unity. In that lecture the Vandermonde matrix was used to create a polynomial that approximates well some given data points, and here we create a periodic function that approximates well some data points.

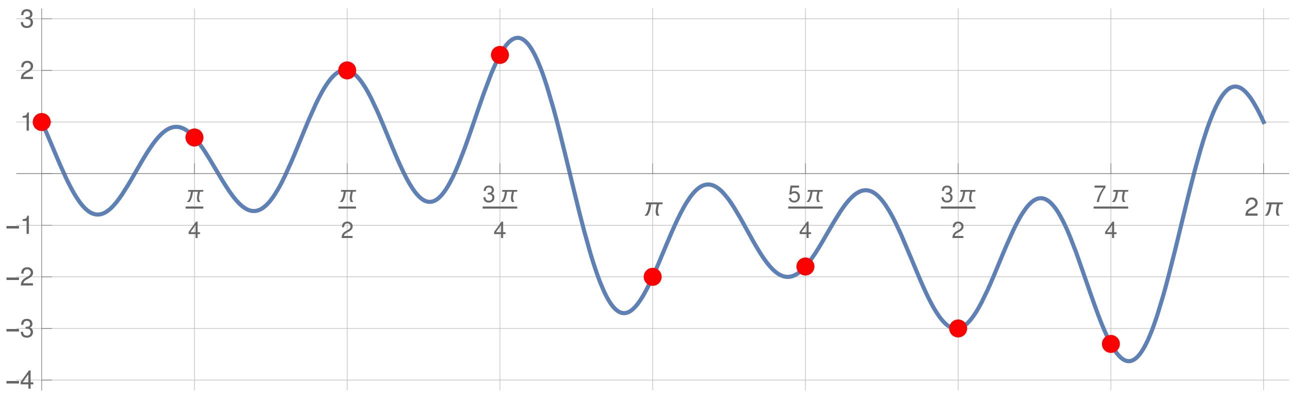
Example 21.8. The reason the Fourier matrix is useful is because it provides the coefficients of the periodic function in the basis F_e that goes through the given data points. So instead of integrating, we simply multiply to get the same result. Consider the data from Example 21.3:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ -2 \\ -3 \end{bmatrix}, \qquad F_4\mathbf{x} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -2 \\ -3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} -2 \\ 3-5i \\ 0 \\ 3+5i \end{bmatrix},$$

which means that $f_4(x) = -\frac{1}{2} + \frac{3-5i}{4} e^{ix} + \frac{3+5i}{4} e^{i3x}$. Plotting the real part of this function (since it is complex-valued) along with the original data points gives the following graph:



If we recieved more data points, spaced $\pi/4$ (instead of $\pi/2$) apart, to get a new data vector $\mathbf{x} \in \mathbf{R}^8$, we could use the Fourier matrix F_8 to reconstruct a continuous function from this data:



We finish off this lecture with some observations about the Fourier matrix F_n.

Remark 21.9. The Fourier matrix F_n is symmetric, which follows immediately from the definition that $(F_n)_{ij} = \omega^{(i-1)(j-1)}$. The matrix is not Hermitian, as both symmetric and Hermitian would imply that everything off the diagonal is zero. As given, F_n is not unitary, but the columns are orthogonal. For example:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}^* \begin{bmatrix} 1 \\ i \\ -1 \\ -i \end{bmatrix} = 1 + i - 1 - i = 0, \qquad \begin{bmatrix} 1 \\ i \\ -1 \\ -i \end{bmatrix}^* \begin{bmatrix} 1 \\ -i \\ -1 \\ i \end{bmatrix} = 1 - 1 + 1 - 1 = 0.$$

If we change the coefficient in front from $\frac{1}{n}$ to $\frac{1}{\sqrt{n}}$, then F_n becomes unitary. As a result of the columns being orthogonal, the columns may be interpreted as eigenvectors. Setting all eigenvalues to be 1, we can construct the matrix that has these eigenvectors, and it turns out to be a permutation matrix P that cycles all the coordinates:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \xrightarrow{P} \begin{bmatrix} y \\ z \\ w \\ x \end{bmatrix} \xrightarrow{P} \begin{bmatrix} z \\ w \\ x \\ y \end{bmatrix} \xrightarrow{P} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} \xrightarrow{P} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}.$$

21.3 Exercises

Exercise 21.1. Find the Fourier coefficients a_n, b_n up until $n = 3$ for $f(x) = \sin(x)\cos^2(x)$.

Exercise 21.2. Consider the function $f \in C[0, 2\pi]$ given by $f(x) = \left\{ \begin{smallmatrix} -1 & x<\pi \\ 1 & x\geqslant\pi \end{smallmatrix} \right.$.

1. Compute the Fourier series of f up to $n = 1$, $n = 3$, and $n = 5$. Plot these three functions together with f.

2. Compute the discrete Fourier transform of f for $n = 4$, using evenly spaced samples $f(x_k)$ for $x_k = 2k\pi/4$, with $k = 0, 1, 2, 3$. Express it as a sum of sin and cos functions using Euler's formula.

3. Plot the real part of the discrete Fourier transform of f for $n = 4, 8, 12$ together with f. As above, take $4, 8, 12$ evenly spaced samples in the interval $[0, 2\pi]$, starting with 0. You do not need to show your computations.

Lecture 22: Graphs

Chapter 10.1 in Strang's "Linear Algebra" and IV.6 in Strang's "Learning from Data"

- Fact 1: Simple graphs are either directed or undirected, and do not have multiple edges or loops.
- Fact 2: Row reducing the incidence matrix gives a spanning tree.
- Fact 3: Taking products of the adjacency matrix counts the number of walks between vertices.

- Skill 1: Compute the four matrices associated to graph, and reconstruct the graph from its adjacency or incidence matrix.
- Skill 2: Count the number of of spanning trees using the incidence matrix.
- Skill 3: Count the number of walks between vertices.

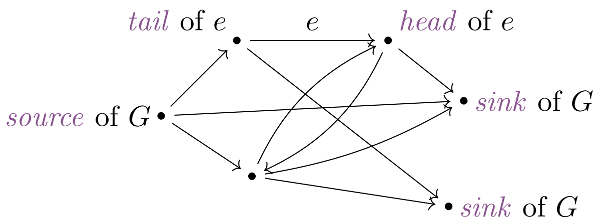
In this lecture we set up a new interpretation of matrices, to be used in the next lecture. We will attempt to apply the tools of matrix algebra already seen so far to graphs.

22.1 Vertices and edges

Definition 22.1. A *graph* G is a pair of sets (V, E), where $V = \{v_1, \ldots, v_n\}$ is a finite set and every element of E is a set $\{v_i, v_j\}$, for $1 \leqslant i < j \leqslant n$. The elements of V are called *vertices* (singular *vertex*) and the elements of E are called *edges*.

The above definition is for an *undirected* graph. For a *directed* graph, or *digraph*, every element of the set E is an ordered set, or pair, (v_i, v_j), with $1 \leqslant i, j \leqslant n$ and $i \neq j$.

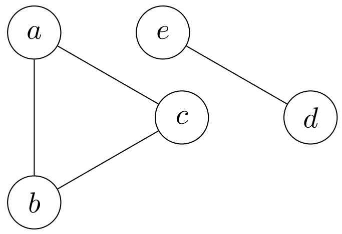
Definition 22.2. Let $G = (V, E)$ be a directed graph. In the edge $e = (t, h) \in E$, the vertex t is called the *tail* and the vertex h is called the *head* of e. If $v \in V$ only appears as a tail in edges, then v is called a *source* of G. If v only appears as a head, then v is called a *sink* of G.



The egde $e = (t, h)$ is an *outgoing* edge of t and an *incoming* edge of h.

In the definitions of both directed and undirected graphs we do not allow repeated edges (since E is a set, it only sees distinct elements) and self loops (an edge $\{v_i, v_i\}$). Graphs without repeated edges and without self loops are called *simple* graphs.

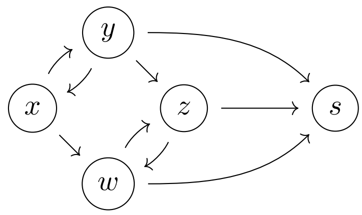
Example 22.3. Here are some examples of graphs.



$$V = \{a, b, c, d, e\}$$

$$E = \{\{a, b\}, \{b, c\}, \{d, e\}, \{c, a\}\}$$

$$\begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$



$$V = \{x, y, z, s\}$$

$$E = \{(x, y), (y, x), (x, w), (y, z), (y, s),$$
$$(w, z), (z, w), (w, s), (z, s)\}$$

$$\begin{array}{c} \\ x \\ y \\ z \\ w \\ s \end{array} \begin{array}{c} \begin{array}{ccccc} x & y & z & w & s \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Each graph has its square *adjacency matrix* $A \in \mathcal{M}_{|V|\times|V|}$ given below it: $A_{ij} = 1$ if an edge from v_i to v_j exists, and is 0 otherwise. In the matrix, every potential edge (even self loops) has a position.

Definition 22.4. Let $G = (V, E)$ be a graph and $A \in \mathcal{M}_{n\times n}$ its adjacency matrix. For $v_k \in V$, the *degree* of v_k is the number of edges in E in which v_k appears. Or, it is the sum

$$\deg(v_k) = \underbrace{\sum_{i=1}^{n} A_{ik}}_{G \text{ undirected}} \qquad \text{or} \qquad \deg(v_k) = \underbrace{\overbrace{\sum_{i=1}^{n} A_{ik}}^{\textit{out-degree}} + \overbrace{\sum_{j=1}^{n} A_{kj}}^{\textit{in-degree}}}_{G \text{ directed}}.$$

The out-degree of v_k is denoted $\text{outdeg}(v_k)$, and the in-degree is denoted $\text{indeg}(v_k)$.

Remark 22.5. There are several other matrices related to a graph $G = (V, E)$:

- the *incidence matrix* $N \in \mathcal{M}_{|E|\times|V|}$, where $N_{ij} = -1$ if vertex j is the tail of edge i, and 1 if it's the head of edge i
- the *degree matrix* $D \in \mathcal{M}_{|V|\times|V|}$, which is diagonal and $D_{ii} = \deg(v_i)$
- the *Laplacian matrix* $L \in \mathcal{M}_{|V|\times|V|}$, defined as $L = N^T N$. If G is undirected, then $L = D - A$.

Most often the adjacency matrix is used, since it is square and the graph can be easily reconstructed from it.

Example 22.6. Consider the following graph G, for which we construct all the related matrices and compute the degrees of all vertices.

graph matrices

Definition 22.7. Let $G = (V, E)$ be a graph. A *subgraph* of G is a graph $G' = (V,', E')$ with $V' \subseteq V$ and $E' \subseteq E$. The subgraph G' is called *induced* if for every $e = (v, w) \in E$ with $v, w \in V'$, we also have $e \in E'$.

Example 22.8. Here is an example of graph with two subgraphs, only one of which is induced.

$$G = \quad H = \quad K =$$

22.2 Patterns in graphs

Definition 22.9. Let $G = (V, E)$ be a graph.

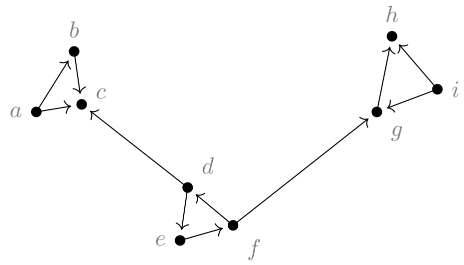
- A *path* in G is an ordered sequence of distinct vertices $v_1, \ldots, v_n$ for which v_i and v_{i+1} form an edge, for every i.
- A *walk* in G is the same as a path, but the vertices do not need to be distinct.
- A *cycle*, or *loop* in G is a path for which v_n and v_1 form an edge.

In directed graphs, the edges of these objects do not need to all be oriented the same way, but often it is assumed they are. To highlight the difference in digraphs, the words *undirected* and *directed* are used in front of each of these objects.

Every one of the objects in Definition 22.9, directed or undirected, is related to a unique sequence of edges. That is, these objects are often given in terms of the edges rather than the vertices.

Example 22.10. Consider the following directed graph and associated sequences of vertices.

$G =$



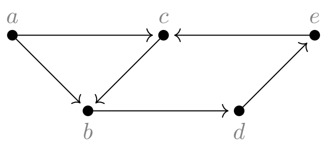
$$\begin{aligned} P_1 &= (d, e, f, g, h) \\ P_2 &= (c, d, e, f, g, h) \\ W_1 &= (f, d, e, f, g, h) \\ W_2 &= (a, b, c, a, c, a, b, c) \\ C_1 &= (a, b, c) \\ C_2 &= (d, e, f) \end{aligned}$$

Here P_1 is a (directed) path, P_2 is an undirected path, but W_1 is not a path, as f appears twice. The sequence W_1 is a (directed) walk and W_2 is an undirected walk. For cycles, C_1 is an undirected cycle (though it is a directed path) and C_2 is a directed cycle (and a directed path).

Row reduction was a key operation in matrices, but so far we have not seen row operations for matrices related to graphs.

Remark 22.11. Let $G = (V, E)$ be the graph given below, with incidence matrix N.

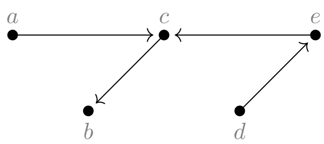
$G =$



$$N = \begin{array}{c} \\ ab \\ ac \\ bd \\ cb \\ de \\ ec \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \end{array}$$

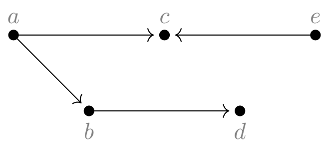
The linearly independent rows of the incidence matrix N of G form a *spanning tree* T of G. That is, $T = (V', E')$ is a subgraph of G with $V' = V$, and T has no cycles (directed or undirected). For G, we have many spanning trees, including T_1 and T_2 given below.

$T_1 =$



$$ab = ac + cb$$
$$bd = -cb - de - ec$$

$T_2 =$



$$cb = ab - ac$$
$$de = ac - ab - bd - ec$$

Proposition 22.12. Let $G = (V, E)$ be a graph with adjacency matrix A and $v_i, v_j \in V$. The number of walks from v_i to v_j of length k is the (i, j)-entry of A^k.

Example 22.13. Consider the following graphs and their powers. Since loops, number of walks always grows.

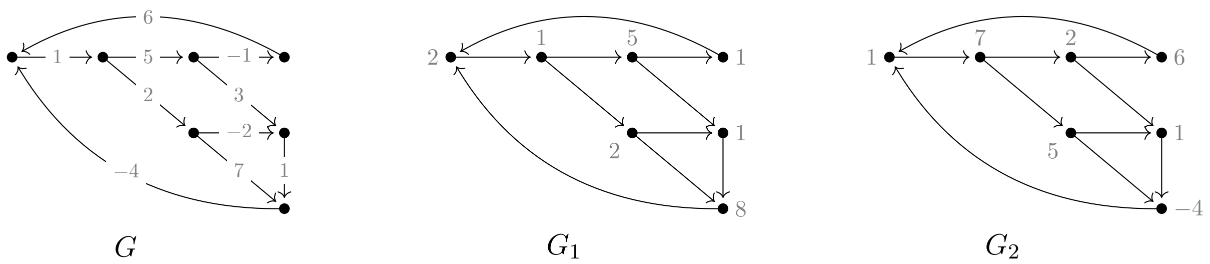
$$G_1 \ directed \ G_2 \ undirected \ table \ of \ walks$$

The adjacency matrix for an undirected graph is symmetric and *binary*, which means the entries are either 1 or 0. For directed graphs, the matrix is still binary, but not symmetric. Matrices that are neither symmetric nor binary are associated to a special type of graph.

Definition 22.14. A graph $G = (V, E)$ is *weighted* when accompanied by a function $w \colon E \to \mathbf{R}$. This is sometimes called an *edge-weighted* graph to distinguish it from a *vertex-weighted* graph, which needs a function $w \colon V \to \mathbf{R}$.

Vertex-weighted directed graphs can be turned into edge-weighted graphs by assigning each edge the weight of its head (or tail). Similarly, an edge-weighted graph can be turned into a vertex weighted graph by assigning each vertex the sum of the weights of all incoming (or outgoing) edges.

Example 22.15. Here is an example of an edge-weighted directed graph G and two vertex-weighted graphs G_1, G_2 that are built following the comment above.



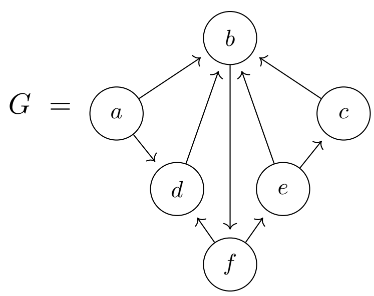
22.3 Exercises

Exercise 22.1. Let $G = (V, E)$ be a directed simple graph, and let A be its adjacency matrix.

1. Just by looking at A, how can you tell which vertices are sinks and which are sources of G?
2. What is the largest number of edges that G can have?

Exercise 22.2. Use induction to prove Proposition 22.12.

Exercise 22.3. Consider the following directed graph:



1. Give the adjacency and incidence matrix for G.
2. Find all $k \in \mathbf{N}$ for which there are no walks of length k from f to f.
3. Find as many spanning trees as you can for G.

Lecture 23: Markov matrices and spectral clustering

Chapter 10.3 in Strang's "Linear Algebra" and IV.7, V.6 in Strang's "Learning from Data"

- Fact 1: Markov matrices have all eigenvalues $|\lambda| \leqslant 1$, and at least one is equal to 1. If all matrix entries are positive, only one eigenvalues is equal to 1.
- Fact 2: The eigenvector corresponding to the Laplacian spectral gap (or the Fiedler eigenvalue) of a directed graph clusters the vertices of the graph into two groups.

- Skill 1: Find the steady state of a Markov matrix with non-zero entries.
- Skill 2: Identify when a Markov matrix does or does not have a steady state.
- Skill 3: Cluster graph vertices usiung the Laplacian spectral gap.

In this lecture we reconsider Markov matrices, also called *stochastic* matrices, which we already saw in Lecture 13. These are related to graphs, and will help us understand big graphs from their small parts, or their *clusters*.

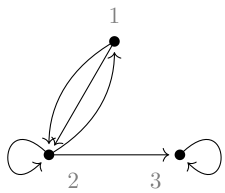
23.1 Markov matrices

Recall that a *Markov matrix* is an element $M \in \mathcal{M}_{n\times n}([0,1])$, for which we ask the sum of every column is 1 or of every row is 1. If columns sum to 1, then M is a *left stochastic* matrix, and if rows sum to 1, then M is a *right stochastic* matrix. The word "stochastic" says that something will change, or will change in an unpredictable manner.

Definition 23.1. Let $G = (V, E)$ be a directed graph, not necessarily simple, with $|V| = n$ and adjacency matrix $A \in \mathcal{M}_{n\times n}$. The *transition probability matrix* is $T \in \mathcal{M}_{n\times n}(\mathbf{Z})$ with

$$T_{ij} = \begin{cases} 1/\text{outdeg}(v_i) & \text{if outdeg}(v_i) \neq 0 \text{ and } A_{ij} = 1, \\ 0 & \text{if outdeg}(v_i) = 0. \end{cases}$$

Remark 23.2. When every vertex of G has an outgoing edge, the transition probability matrix of G is a right stochastic matrix:

$G =$ 

$$T = \begin{bmatrix} 0 & 1 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 \end{bmatrix}$$

Proposition 23.3. A Markov matrix (right stochastic matrix) M always has an eigenvalue 1 with eigenvector $[1\ 1\ \cdots\ 1]^T$. All other eigenvalues λ have $|\lambda| \leqslant 1$.

Proof. Since the rows of M add up to 1, it is immediate that $M[1\ 1\ \cdots\ 1]^T = [1\ 1\ \cdots\ 1]^T$. For the other eigenvalues, suppose that $M\mathbf{x} = \lambda\mathbf{x}$, and take the largest entry in $\mathbf{x}$, suppose it is x_k. Taking the absoulte values on the kth line of $M\mathbf{x} = \lambda\mathbf{x}$, we have $|\lambda x_k| = |\lambda||x_k|$ on the right, and

$$\left|\sum_{j=1}^{n} M_{kj}x_j\right| \leqslant \sum_{j=1}^{n} M_{kj}\,|x_j| \leqslant \sum_{j=1}^{n} M_{kj}\,|x_k| = |x_k|$$

on the left. The first relation follows by the triangle inequality, the second relation follows by assumption, and the third relation follws as the matrix is right stochastic. Hence $|\lambda| \leqslant 1$. ☐

Since the eigenvalues of a matrix and its transpose are the same (see Proposition 13.7), the same result holds for a left stochastic matrix.

Example 23.4. Consider the two Markov matrices $A = \left[\begin{smallmatrix} .2 & .8 \\ .9 & .1 \end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & .5 & .5 \end{smallmatrix}\right]$. We compute their right eigenvectors and eigenvalues to be

$$\mathbf{u}_i = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -8/9 \\ 1 \end{bmatrix}, \ \lambda_i = 1, -\frac{7}{10}, \qquad \mathbf{v}_i = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \ \lambda_i = 1, 1, \frac{1}{2}.$$

Similarly, we can find their left eigenvectors and eigenvalues to be

$$\mathbf{u}_i = \begin{bmatrix} 9/8 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \ \lambda_i = 1, -\frac{7}{10}, \qquad \mathbf{v}_i = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \ \lambda_i = 1, 1, \frac{1}{2}.$$

This shows an important observation: when the entries of M are all strictly positive, there is a unique eigenvector with eigenvalue 1 (this is the *Perron–Frobenius theorem*). When some entries of M are zero, we can have several independent eigenvectors with eigenvalues 1.

Definition 23.5. Let $M \in \mathcal{M}_{n\times n}$ be left stochastic with all positive entries. The unique eigenvector $\mathbf{v} \in \mathbf{R}^n$ of M corresponding to the eigenvalue 1 is the *steady state*, or *Perron–Frobenius eigenvector* of M.

This means that for every vector $\mathbf{x}$, the vector $M^k\mathbf{x}$ eventually converges to a multiple of the vector $\mathbf{v}$.

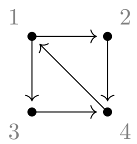
The vector $\mathbf{v}$ corresponds to a left eigenvector of a right stochastic matrix M, that is, $\mathbf{v}^T M = \mathbf{v}$. In both cases, the vector is important, because along with the ability to express any vector $\mathbf{x}$ as a linear combination of eigenvectors $\mathbf{x} = a_1\mathbf{v} + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$, we get that

$$\lim_{k\to\infty} M^k\mathbf{x} = \lim_{k\to\infty} \left(1^k a_1\mathbf{v} + \lambda_2^k a_2\mathbf{v}_2 + \cdots + \lambda_n^k a_n\mathbf{v}_n\right) = a_1\mathbf{v},$$

since $\lim_{k\to\infty} \lambda_i^k = 0$ for all $i \neq 1$, as $|\lambda_i| < 1$.

Example 23.6. There are some Markov matrices that do not have all positive entries, but which still have a steady state to which everything converges to. That is, there is still a unique eigenvector with eigenvalue 1:

$$\begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix}^T \underbrace{\begin{bmatrix} 0 & .5 & .5 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}}_{M} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix}^T, \qquad G_M =$$



The steady state of M will exist when, in the corresponding graph G_M, it is always possible to get from any vertex to any other vertex. The graph G_M has M as its transition probability matrix.

23.2 Spectral clustering

Definition 23.7. Let $G = (V, E)$ be a graph.

- If G is undirected, or considering a directed graph as undirected, the graph is *connected* if for every $v, w \in V$, there exists a path from v to w.
- If G is directed, the graph is *strongly connected* if for every $v, w \in V$, there exists a path from v to w.

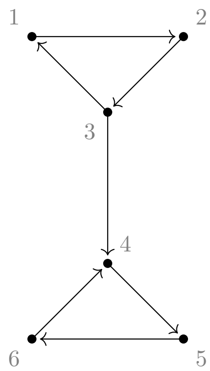
Connectivity can be considered relatively, in terms of how far vertices are from each other.

Definition 23.8. Let $G = (V, E)$ be a graph and N its incidence matrix. The Laplacian $L = N^T N$ has smallest eigenvalue 0, from the vector $\mathbf{v} = [1\ 1\ \cdots\ 1]^T$, as $N\mathbf{v} = 0$. The next smallest eigenvalue λ is the *Fiedler eigenvalue* $\lambda > 0$, and the eigenvector $\mathbf{f}$ corresponding to it is the *Fiedler eigenvector*.

The smallest positive eigenvalue of the Laplacian of a graph G is also called the *Laplacian spectral gap* of G. The Fiedler eigenvector $\mathbf{f}$ is important because it partitions the vertices of G into two *clusters*: those whose entry is positive, and those whose entry is negative.

Example 23.9. Consider the following graph, its incidence matrix, and its Laplacian matrix.

$G =$



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

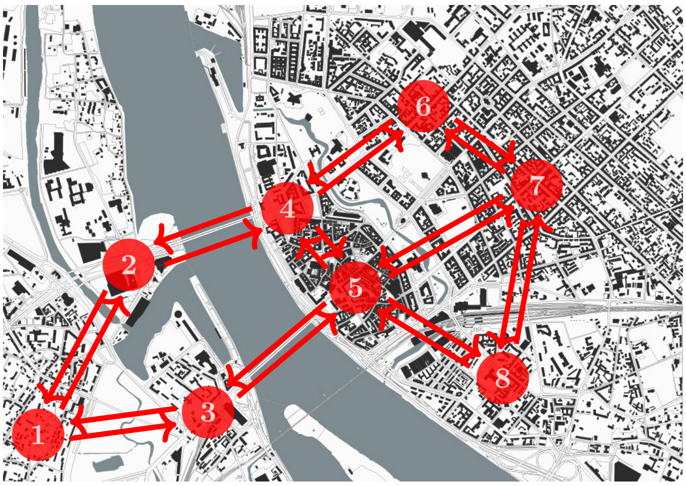
The smallest eigenvalue that is not zero is $\lambda = \frac{5-\sqrt{17}}{2}$, which corresponds to the Fiedler eigenvector $\mathbf{f} = \left[-1\ -1\ (3-\sqrt{17})/2\ (-3+\sqrt{17})/2\ 1\ 1\right]$. Since $\sqrt{17} \in (4, 5)$, the first three entries are negative, and the last three are positive, showing the clustering $\{1, 2, 3\}$ and $\{4, 5, 6\}$ of the vertices of G.

23.3 Exercises

Exercise 23.1. Let $M \in \mathcal{M}_{2\times 2}$.

1. Let $M = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ be Markov (right stochastic) with $d = 0$. Show that if M^3 has only positive entries, then M^2 has only positive entries.
2. Find and example of M (not Markov, but with $|M_{ij}| \leqslant 1$) for which M^3 has only positive entries, but M has at least one negative entry.

Exercise 23.2. Consider the map of Riga below, with vertices as marked.



1. Construct the transition probability matrix R for Riga.
2. Will the matrix have a steady state? If not, explain why. If yes, compute it.
3. Find the Fiedler eigenvector and give the clustering according to it.

Lecture 24: Graph clustering

Chapter IV.7 in Strang's "Learning from Data"

- Fact 1: The clustering coefficient in an undirected graph indicates how tight of a cluster each vertex has with its neighbors.
- Fact 2: A weighted graph can be transformed into a distance matrix.

- Skill 1: Find the smallest cut-set in small graphs.
- Skill 2: Compute weights of vertex and edge sets. Compute the normalized cut and k-cut weight.

In this lecture we continue the idea of clustering, introduced in Lecture 23, on more general data sets that have more than two clusters.

24.1 Graph structures for clustering

Definition 24.1. Let $G = (V, E)$ be a graph with $V' \subseteq V$ and $v \in V$.

- The *induced subgraph* of V is the graph $G' = (V', \{\{u, v\} \in E \ : \ u, v \in V'\} \subseteq G$
- The *neighbors* of V are vertices $u \in V$ for which $\{u, v\} \in V$
- The *open neighborhood* of v is the subgraph $N_G^\circ(v) \subseteq G$ induced by the neighbors of v
- The *closed neighborhood* of v is the subgraph $N_G(v) \subseteq G$ induced by the neighbors of v and v itself.
- The *star* of v is the set of edges $\text{star}(v) \subseteq E$ in the closed neighborhood of v that are not in the open neighborhood of v.

The subscript G is omitted if G is clear from context. If neither "open" nor "closed" are specified in front of "neighbohood," then the neighborhood is assumed to be closed.

A naive way to cluster vertices of a vertex-weighted graph would be to cluster them by weight. This is a type of *filtering*. This approach may not yield the best results, but it highlights that all we need for clustering is a function on the vertices - a filter. We now consider one such more relevant function.

Definition 24.2. Let $G = (V, E)$ be an undirected graph. The *clustering coefficient* of $v \in V$ is

$$\text{cc}(v) = \frac{\text{number of 3-cycles of } G \text{ containing } v}{\text{number of 3-cycles that } v \text{ could be in}} = \frac{\text{number of 3-cycles of } G \text{ containing } v}{\binom{\deg(v)}{2}} \in [0, 1].$$

A value close to 1 indicates that the closed neighborhood of v is almost a complete graph.

Example 24.3. Consider the following graph G. The clustering coefficients are gven next to each vertex.

$$G \ =$$

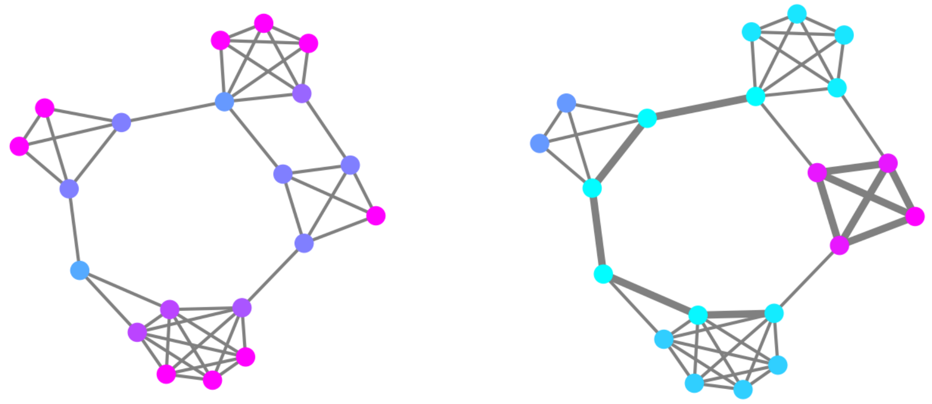
Remark 24.4. We can extend this idea to weighted graphs. Given a weighted graph $G = (V, E)$ with weight function $w \colon E \to \mathbf{R}$ and subsets $V' \subseteq V$, $E' \subseteq E$, we can assign weights to them:

$$w(E) = \sum_{e \in E} w(e), \qquad w(V') = \sum_{v \in V'} \sum_{e \in \text{star}(v)} w(e).$$

Note that if V' contains u, v with $e = \{u, v\} \in E$, then the weight of e will be counted twice in the weight of V'. The *weighted* clustering coefficient is then

$$\text{wcc}(v) = \frac{\text{weights of 3-cycles of } G \text{ containing } v}{\text{weights of 3-cycles that } v \text{ could be in}} = \frac{\text{weights of 3-cycles of } G \text{ containing } v}{(\deg(v) - 1) w(\text{star}(v)) + w(N^\circ(v))} \in [0, 1].$$

Example 24.5. Consider the graph below, which has 4 clear clusters. The clustering coefficient (unweighted) of each vertex is given as a color, ranging from pink (high value) to blue (low value). On the left, the graph is unweighted, so as expected, in each cluster high values appear, with lower values given to vertices that are "closer" to other clusters.



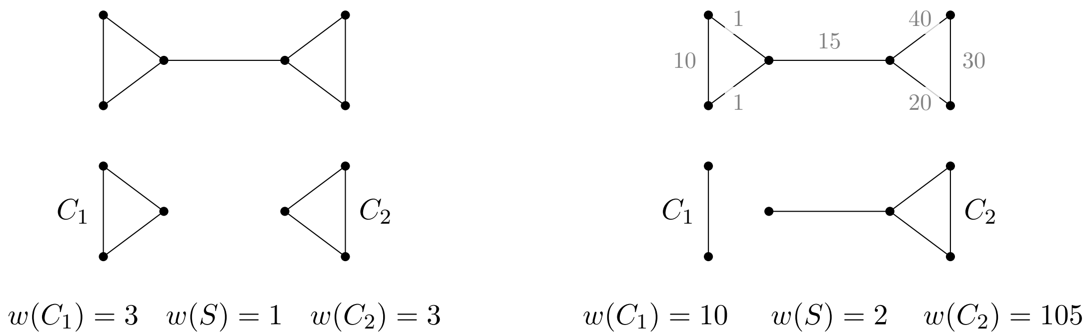
The graph on the right is weighted (higher weights are thicker edges), and triangles of higher weight increase the value of the clustering coefficient, whereas higher weight edges that do not form a triangle seem to decrease the clustering coefficient.

24.2 k-means clustering

In this section we consider one way to cluster vertices of an undirected graph into $k \in \mathbf{N}$ clusters. Most often weighted graphs are used for clustering, but in the case our graph is not weighted, we may simply assign a weight of 1 to each edge.

Definition 24.6. let $G = (V, E)$ be a graph. A *cut* of G is a partition of V into two sets C_1, C_2. If G is connected, a cut can also be defined in terms of the edges $\{e = \{v, w\} \in E \; : \; v \in C_1, w \in C_2\}$ between the two pieces of the partition. Such a set of edges that corresponds to a cut is called a *cut-set* of G.

Example 24.7. One way to cluster vertices of $G = (V, E)$ into two sets is to find the smallest cut-set of G. Using the graph G from Example 23.9, we immediately get the same result we previously had, on the left below.



If the edges of the graph are weighted as on the right, we get a different smallest cut-set. The weight of the cut-set is 2, which is the smallest among all possible cut-set weights.

It is natural to want to minimize the weight of the edges in the cut-set, but also to maximize the weight of the edges left in the clusters. To reflect both of these ideas and to exclude degenerate cases when having a single vetrex cluster would minimize the cut-set weight, we need to *normalize*.

Definition 24.8. For (C_1, C_2) a cut of $G = (V, E)$ with corresponding cut-set S, the *normalized cut* *weight* of this cut is

$$ncw(C_1, C_2) = \frac{w(S)}{w(C_1)} + \frac{w(S)}{w(C_2)}.$$

Example 24.9. Consider the graph $G = (V, E)$ below, with the two cut sets S, T.

$$G =$$

Both cuts have the same weight, as do their resulting partitions. Situations where two clusters are not enough lead to a generalization of the above.

Definition 24.10. Let $G = (V, E)$ be a weighted graph with weight function $w \colon E \to \mathbf{R}$. Let $V' \subseteq V$ and $G' = (V', E') \subseteq G$ be the induced graph of V'. The *star* of V' is

$$\text{star}(V') = \left(\bigcup_{v \in V'} \text{star}(v) \right) \setminus \left(\bigcup_{e \in E'} e \right) \subseteq E.$$

This captures the idea of generalizing a cut-set from two partitions to more than two. A partition $C_1, \dots, C_k$ of V corresponds to the *k-cut*

$$\{e = (u, v) \in E \ : \ u \in C_i, v \in C_j, i \neq j\} = \bigcup_{\ell=1}^{k} \text{star}(C_\ell).$$

The *normalized k-cut weight* of this partition is

$$ncw(C_1, \dots, C_k) = \frac{w(\text{star}(C_1))}{w(C_1)} + \frac{w(\text{star}(C_2))}{w(C_2)} + \cdots + \frac{w(\text{star}(C_k))}{w(C_k)}.$$

Example 24.11. A large value for the normalized k-cut weight indicates a bad cut, a small indicates a better cut. For example, consider the two following 3-cuts of the graph below. The one with the lower valuie is also visually the better of the two cuts.

$$G =$$

Remark 24.12. Recall *distance matrices* from Example 10.7 in Lecture 10. To every weighted graph $G = (V, E)$ we can associate a distance matrix $D \in \mathcal{M}_{|V| \times |V|}$, where D_{ij} is the sum of weights along the shortest path from v_i to v_j. We can recover the *position matrix* X, and from it we can compute the average of each cluster:

$$\begin{pmatrix} \text{weighted} \\ \text{graph} \end{pmatrix} \to \begin{pmatrix} \text{distance} \\ \text{matrix} \end{pmatrix} \to \begin{pmatrix} \text{position} \\ \text{matrix} \end{pmatrix} \to \begin{pmatrix} \text{average } \mathbf{a}_i \\ \text{of each} \\ \text{cluster } C_i \end{pmatrix}$$

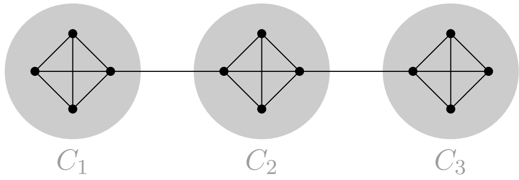
Then the *k-means* partition is the one which minimizes the function

$$\sum_{v \in C_1} \|\mathbf{a}_1 - v\|^2 + \sum_{v \in C_2} \|\mathbf{a}_2 - v\|^2 + \cdots + \sum_{v \in C_k} \|\mathbf{a}_k - v\|^2.$$

Here v is used to represent the position vector associated to the vertex v.

24.3 Exercises

Exercise 24.1. Consider the undirected graph $G = (V, E)$ below, with partition C_1, C_2, C_3 as indicated. Every edge in the subgraphs induced by C_1, C_2, C_3 has weight 1. For every $r \in \mathbf{R}$, extend this weight function to all the edges of G so that the normalized 3-cut weight of C_1, C_2, C_3 is r.



Part V
Answers to lecture exercises

Lecture 1: Vectors and matrices

Exercise 1.1. We solve the equation line by line.

1. From the first line, we have $-3b = -5$, which means $b = 5/3$. From the second line on the left and, using the result $a = 35/18$ with the third line on the right, we have:

$$\begin{aligned} 6a - 4b &= 5 \\ 6a - 20/3 &= 5 \\ 18a - 20 &= 15 \\ 18a &= 35 \\ a &= 35/18 \end{aligned} \qquad\qquad \begin{aligned} -a - 5b + c &= -4 \\ -35/18 - 25/3 + c &= -4 \\ -35 - 150 + 18c &= -72 \\ 18c &= 113 \\ c &= 113/18 \end{aligned}$$

2. The requested equation is

$$\underbrace{\begin{bmatrix} 0 & -3 & 0 \\ 6 & -4 & 0 \\ -1 & -5 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 35/18 \\ 5/3 \\ 113/18 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} -5 \\ 5 \\ 4 \end{bmatrix}}_{\mathbf{y}}.$$

Exercise 1.2. By expressing each vector in terms of its constitutent parts, we see the desired result. Let $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{v} = (v_1, \ldots, v_n)$, and $\mathbf{w} = (w_1, \ldots, w_n)$. Then

$$\begin{aligned} \mathbf{v} \cdot (\mathbf{u} + \mathbf{w}) &= \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \cdot \left(\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \right) && \text{(definition of vectors } \mathbf{u}, \mathbf{v}, \mathbf{w}\text{)} \\ &= \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} u_1 + w_1 \\ \vdots \\ u_n + w_n \end{bmatrix} && \text{(definition of matrix addition)} \\ &= v_1(u_1 + w_1) + \cdots + v_n(u_n + w_n) && \text{(definition of dot product)} \\ &= v_1u_1 + v_1w_1 + \cdots + v_nu_n + v_nw_n && \text{(multiplication of real numbers)} \\ &= (v_1u_1 + \cdots + v_nu_n) + (v_1w_1 + \cdots + v_nw_n) && \text{(rearranging)} \\ &= \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} && \text{(definition of dot product)} \\ &= \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{w}. \end{aligned}$$

Exercise 1.3. Since $\mathbf{w} = (w_1, w_2, w_3)$ and $\mathbf{z} = (z_1, z_2, z_3)$ are perpendicular to $\mathbf{v} = (v_1, v_2, v_3)$, we have that

$$\begin{aligned} 0 &= \mathbf{v} \cdot \mathbf{w} = v_1w_1 + v_2w_2 + v_3w_3, \\ 0 &= \mathbf{v} \cdot \mathbf{z} = v_1z_2 + v_2z_2 + v_3z_3. \end{aligned}$$

The halfway point between $\mathbf{w}$ and $\mathbf{z}$ is $\mathbf{h} = \left(\frac{w_1+z_1}{2}, \frac{w_2+z_2}{2}, \frac{w_3+z_3}{2}\right)$, and for this vector

$$\begin{aligned}
\mathbf{v} \cdot \mathbf{h} &= (v_1, v_2, v_3) \cdot \left(\frac{w_1 + z_1}{2}, \frac{w_2 + z_2}{2}, \frac{w_3 + z_3}{2}\right) \\
&= v_1 \cdot \frac{w_1 + z_1}{2} + v_2 \cdot \frac{w_2 + z_2}{2} + v_3 \cdot \frac{w_3 + z_3}{2} \\
&= \frac{1}{2}\left(v_1w_1 + v_1z_1 + v_2w_2 + v_2z_2 + v_3w_3 + v_3z_3\right) \\
&= \frac{1}{2}\left((v_1w_1 + +v_2w_2 + v_3w_3) + (v_1z_1 + v_2z_2 + v_3z_3)\right) \\
&= \frac{1}{2}\left(\mathbf{v} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{z}\right) \\
&= \frac{1}{2}(0 + 0) \\
&= 0.
\end{aligned}$$

Exercise 1.4. 1. One example is $B = \left[\begin{smallmatrix} 1 & 2 \\ 3 & 6 \\ 0 & 1 \end{smallmatrix}\right]$, for which

$$AB = \begin{bmatrix} 1 & 0 & -2 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 6 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I,$$

and

$$BA = \begin{bmatrix} 1 & 2 \\ 3 & 6 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2 \\ 3 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 7 & -2 & 0 \\ 21 & -6 & 0 \\ 3 & -1 & 1 \end{bmatrix} \neq I.$$

2. No, it is not possible, because the three equations

$$c_{11} + 3c_{12} = 1, \qquad -c_{12} = 0, \qquad -2c_{12} + c_{22} = 0$$

in two unknowns, and none of the equations are multiples of each other. There are no possible solutions to this.

Exercise 1.5. Suppose that C exists with $AC = CA = I$. Multiplying $CA = I$ by B on the right gives

$$\begin{aligned}
(CA)B &= IB \\
C(AB) &= B \\
CI &= B \\
C &= B.
\end{aligned}$$

Exercise 1.6. Let $A, B \in \mathcal{M}_{n\times n}$, with ij-entries a_{ij} and b_{ij}, respectively.

1. Suppose that A, B are lower triangular, so $a_{ij} = 0$ and $b_{ij} = 0$ if $i < j$. In the product, the ij entry of AB, for $i < j$, is

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} = \left(\sum_{k=1}^{i} A_{ik} \underbrace{B_{kj}}_{=0}\right) + \left(\sum_{k=i+1}^{n} \underbrace{A_{ik}}_{=0} B_{kj}\right) = 0.$$

Hence AB is also lower triangular.

2. Suppose that A, B are upper triangular, so $a_{ij} = 0$ and $b_{ij} = 0$ if $i > j$. In the product, the ij

entry of AB, for $i > j$, is

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} = \left(\sum_{k=1}^{j} \underbrace{A_{ik}}_{=0} B_{kj}\right) + \left(\sum_{k=j+1}^{n} A_{ik} \underbrace{B_{kj}}_{=0}\right) = 0.$$

Hence AB is also upper triangular.

3. The result does not have to be triangular, for example:

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

Here we have two non diagonal matrices, whose product is a diagonal matrix:

$$\begin{bmatrix} 6 & -10 \\ 77 & 22 \end{bmatrix} \begin{bmatrix} 2 & 5 \\ -7 & 3 \end{bmatrix} = \begin{bmatrix} 82 & 0 \\ 0 & 451 \end{bmatrix}.$$

Lecture 2: Elimination and inverses

Exercise 2.1. Some examples are given below. Many more exist.

1. An example is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, as all the pivots can be read off the diagonal.

2. An example is $\begin{bmatrix} 1 & 0 & 0 \\ 6 & 2 & 0 \\ 5 & 8 & 3 \end{bmatrix}$, as elimination tells us to:

subtract $\ell_{21} = 6$ of the first row from the second row: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 5 & 8 & 3 \end{bmatrix}$

subtract $\ell_{31} = 5$ of the first row from the third row: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 8 & 3 \end{bmatrix}$

subtract $\ell_{32} = 4$ of the second row from the third row: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

This ends Gaussian elimination, as the matrix is upper triangular, and indicates the pivots on the diagonal.

3. An example is $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, as elimination tells us to:

subtract $\ell_{21} = 1$ of the first row from the second row: $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

subtract $\ell_{31} = 1$ of the first row from the second row: $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

This ends Gaussian elimination, as the matrix is upper triangular, and indicates the pivots on the diagonal. There is no third pivot, since the third row is all zeros.

Exercise 2.2. 1. For the first matrix, elimination tells us to:

$$\text{subtract } \ell_{21} = d/a \text{ of the first row from the second row:} \quad \begin{bmatrix} a & b & c \\ 0 & e - bd/a & f - cd/a \\ g & h & i \end{bmatrix}$$

$$\text{subtract } \ell_{31} = g/a \text{ of the first row from the third row:} \quad \begin{bmatrix} a & b & c \\ 0 & e - bd/a & f - cd/a \\ 0 & h - bg/a & i - cg/a \end{bmatrix}$$

For the multiplier ℓ_{32}, it needs to be

$$(h - bg/a) \cdot (e - bd/a)^{-1} = \frac{h - \frac{bg}{a}}{e - \frac{bd}{a}} = \frac{ah - bg}{ae - bd}.$$

The lower right entry after this step will be $(i - cg/a) - (f - cd/a) \cdot \frac{ah-bg}{ae-bd}$, which we call simply n, because it is very long to write. So elimination tells us to

$$\text{subtract } \ell_{32} = \frac{ah - bg}{ae - bd} \text{ of the second row from the third row:} \quad \begin{bmatrix} a & b & c \\ 0 & e - bd/a & f - cd/a \\ 0 & 0 & n \end{bmatrix}$$

This ends Gaussian elimination, as the matrix is upper triangular, and indicates the pivots $a, e - bd/a, n$ on the diagonal.

For the second matrix, elimination tells us to:

$$\text{subtract } \ell_{32} = h/e \text{ of the second row from the third row:} \quad \begin{bmatrix} 0 & b & c \\ 0 & e & f \\ 0 & 0 & i - fh/e \end{bmatrix}$$

This ends Gaussian elimination, as the matrix is upper triangular, and indicates the pivots $e, i - fh/e$ on the diagonal.

For the third matrix, elimination tells us to:

$$\text{subtract } \ell_{21} = d/a \text{ of the first row from the second row:} \quad \begin{bmatrix} a & b & c \\ 0 & 0 & f - cd/a \\ d & bd/a & i \end{bmatrix}$$

$$\text{subtract } \ell_{31} = d/a \text{ of the first row from the third row:} \quad \begin{bmatrix} a & b & c \\ 0 & 0 & f - cd/a \\ 0 & 0 & i - cd/a \end{bmatrix}$$

This ends Gaussian elimination, as the matrix is upper triangular, and indicates the pivots $a, i - cd/a$ on the diagonal.

For the fourth matrix, elimination tells us to:

$$\text{subtract } \ell_{32} = 1 \text{ of the second row from the third row:} \quad \begin{bmatrix} 0 & b & c \\ 0 & e & ce/b \\ 0 & 0 & 0 \end{bmatrix}$$

This ends Gaussian elimination, as the matrix is upper triangular, and indicates the only pivot e on the diagonal.

2. Here is an example of such a function, in Python, using the input `A[[a,b,c],[d,e,f],[g,h,i]]`. We use the result from the first matrix in part 1. above.

```
def pivots(A):
    a = A[0][0]
```

```
    b = A[0][1]
    c = A[0][2]
    d = A[1][0]
    e = A[1][1]
    f = A[1][2]
    g = A[2][0]
    h = A[2][1]
    i = A[2][2]
    return [a, b*d/a, (i-c*g/a)-(f-c*d/a)*(a*h-b*g)(a*e-b*d)]
```

3. Here is some Python code that produces the range and average as requested, using the function above.

```
import numpy as np
values = []
for i in range(1000):
    M1 = np.random.rand(3,3)
    M2 = 2*M1 - np.ones((3,3))
    values += pivots(M2)
print([min(values), max(values), sum(values)/len(values)])
```

This is the result it prints on one particular run:

```
[-1105.1138842178975, 1650.5842938466174, -0.272518610029052]
```

Exercise 2.3. 1. Yes, the product of all three is a permutation matrix:

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

This is also immediate from knowing that a permutation matrix rearrages rows. Rearranging rows three times is the same as rearranging them once, but with more steps.

2. Yes, the inverses are all permutation matrices. We construct the inverse by observing that if, for example, A sends row 1 to row 2, row 2 to row 3, and row 3 to row 1, then the inverse of A will send row 1 to row 3, row 2 to row 1, and row 3 to row 2 (that is, will put the rows back in their original place. Hence:

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

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

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

In this case, the permutation matrices ar their own inverses (though this is not always true).

Exercise 2.4. Using the observation $(AB)^{-1} = B^{-1}A^{-1}$ $k-1$ times, we find that the inverse is

$$\begin{aligned}(A_1A_2\cdots A_{k-1}A_k)^{-1} &= ((A_1A_2\cdots A_{k-1})A_k)^{-1}\\ &= A_k^{-1}(A_1A_2\cdots A_{k-2}A_{k-1})^{-1}\\ &= A_k^{-1}((A_1A_2\cdots A_{k-2})A_{k-1})^{-1}\\ &= A_k^{-1}A_{k-1}^{-1}(A_1A_2\cdots A_{k-2})^{-1}\\ &\vdots\\ &= A_k^{-1}A_{k-1}^{-1}\cdots A_2^{-1}A_1^{-1}.\end{aligned}$$

Exercise 2.5. We apply row operations to the block matrix $[A\ I] = \begin{bmatrix} 0 & 2 & -1 & 1 & 0 & 0\\ 1 & 0 & -4 & 0 & 1 & 0\\ 2 & 2 & 2 & 0 & 0 & 1\end{bmatrix}$, as below.

swap the first and the second rows to get a first pivot: $\begin{bmatrix} 1 & 0 & -4 & 0 & 1 & 0\\ 0 & 2 & -1 & 1 & 0 & 0\\ 2 & 2 & 2 & 0 & 0 & 1\end{bmatrix}$

subtract $\ell_{31} = 2$ of the first row from the third row: $\begin{bmatrix} 1 & 0 & -4 & 0 & 1 & 0\\ 0 & 2 & -1 & 1 & 0 & 0\\ 0 & 2 & 10 & 0 & -2 & 1\end{bmatrix}$

subtract $\ell_{32} = 1$ of the second row from the third row: $\begin{bmatrix} 1 & 0 & -4 & 0 & 1 & 0\\ 0 & 2 & -1 & 1 & 0 & 0\\ 0 & 0 & 11 & -1 & -2 & 1\end{bmatrix}$

This finishes Gaussian elminaton, so we proceeed with Gauss–Jordan elimination above the diagonal.

subtract $\ell_{23} = -1/11$ of the third row from the second row: $\begin{bmatrix} 1 & 0 & -4 & 0 & 1 & 0\\ 0 & 2 & 0 & 10/11 & -2/11 & 1/11\\ 0 & 0 & 11 & -1 & -2 & 1\end{bmatrix}$

subtract $\ell_{13} = -4/11$ of the third row from the first row: $\begin{bmatrix} 1 & 0 & 0 & -4/11 & 3/11 & 1/11\\ 0 & 2 & 0 & 10/11 & -2/11 & 1/11\\ 0 & 0 & 11 & -1 & -2 & 1\end{bmatrix}$

multiply each row by the inverse of the pivots: $\begin{bmatrix} 1 & 0 & 0 & -4/11 & 3/11 & 1/11\\ 0 & 1 & 0 & 5/11 & -1/11 & 1/22\\ 0 & 0 & 1 & -1/11 & -2/11 & 1/11\end{bmatrix}$

Hence the inverse of A is $A^{-1} = \begin{bmatrix} -4/11 & 3/11 & 1/11\\ 5/11 & -1/11 & 1/22\\ -1/11 & -2/11 & 1/11\end{bmatrix}$.

Exercise 2.6. In Example 2.6, we had the following result:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -1 & 1\end{bmatrix}}_{E_{32}} \cdot \underbrace{\begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ \frac{1}{2} & 0 & 1\end{bmatrix}}_{E_{31}} \cdot \underbrace{\begin{bmatrix} 0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}}_{P_{12}} \cdot \begin{bmatrix} 0 & 6 & -2 & 2\\ 4 & 8 & -4 & 8\\ -2 & 2 & 7 & 12\end{bmatrix} = \begin{bmatrix} 4 & 8 & -4 & 8\\ 0 & 6 & -2 & 2\\ 0 & 0 & 7 & 14\end{bmatrix}$$

For $PA = LDU$ decomposition, we don't need the fourth column $\mathbf{b}$ used in this example. We also note several necessary things:

$$D = \begin{bmatrix} 4 & 0 & 0\\ 0 & 6 & 0\\ 0 & 0 & 7\end{bmatrix}, \qquad E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 1\end{bmatrix}, \qquad E_{31}^{-1} = \begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ -\frac{1}{2} & 0 & 1\end{bmatrix}.$$

For the inverses of elementary matrices, we used the observations from Example 2.11. This gets us almost where we want to be:

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 0 & 6 & -2 \\ 4 & 8 & -4 \\ -2 & 2 & 7 \end{bmatrix}}_{A} = E_{31}^{-1} E_{32}^{-1} \begin{bmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ 0 & 0 & 7 \end{bmatrix}.$$

The lower triangular matrix is

$$L = E_{31}^{-1} E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & 1 & 1 \end{bmatrix},$$

and the product DU is

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -1/3 \\ 0 & 0 & 1 \end{bmatrix}.$$

Putting this all together, we get

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 0 & 6 & -2 \\ 4 & 8 & -4 \\ -2 & 2 & 7 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & 1 & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 4 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 7 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -1/3 \\ 0 & 0 & 1 \end{bmatrix}}_{U}.$$

Lecture 3: The column space and the nullspace

Exercise 3.1. The operations of addition and scalar multiplication clearly exist:

- $c_1(2,1) + c_2(2,1) = (c_1 + c_2)(2,1)$, and $c_1 + c_2 \in \mathbf{R}$
- $c_1 \cdot (c_2(2,1)) = (c_1 c_2)(2,1)$, and $c_1 c_2 \in \mathbf{R}$

The identity element is the zero vector $(0,0) = 0(2,1)$, and every $c(2,1)$ has an inverse $(-c)(2,1)$, for which $c(2,1) + (-c)(2,1) = (c + (-c))(2,1) = 0(2,1) = (0,0)$. Finally, scalar multiplication has the usual identity 1, as $1(c(2,1)) = (1 \cdot c)(2,1) = c(2,1)$. Commutativity, associativity, and distributivity in this space all follow from the same properties of $\mathbf{R}^2$ as a vector space.

Exercise 3.2. To show this, we need to show that every element in W can be expressed an element in V. An arbitrary element of W looks like

$$a(\mathbf{u} + \mathbf{v}) + b(\mathbf{v} + \mathbf{w}),$$

for some $a, b \in \mathbf{R}$. Rearranging, we get

$$a\mathbf{u} + (a + b)\mathbf{v} + b\mathbf{w},$$

which is an element of the span of $\mathbf{u}, \mathbf{v}, \mathbf{w}$, hence in V. Therefore $W \subseteq V$.

Exercise 3.3. 1. First we observe the following linear combinations from the two spans:

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

These are useful because they only have one nonzero entry. That is,

$$\begin{aligned}
\begin{bmatrix} x \\ y \\ z \end{bmatrix} &= x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + (-y) \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \\
&= x \left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \right) + z \left(\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \right) + (-y) \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \\
&= x \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + (x + z - y) \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.
\end{aligned}$$

Hence $\mathbf{R}^3$ is a subspace of $\subseteq V + W$.

2. We take a linear combination of vectors from both spans. Consider

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

This is an element of $V + W$. For it to be an element of $V \cup W$, it must either be in V or in W. This vector is in V if any only if the matrix equation

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

has a solution. However, the first line is the equation $x_1 = 0$ and the last line is $x_2 = -1$, so it must be that $x_1 + x_2 = -1$. But the second line says $x_1 + x_2 = 1$, and these two equations contradict each other, so there is no solution. Similarly, this vector is in W if and only if the matrix equation

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

has a solution. Again, we find the first line $x_1 = 0$ and the third line $x_1 = -1$ cannot both be true at the same time, hence there is no solution. Therefore $\mathbf{u} \notin V$ and $\mathbf{u} \notin W$, so $\mathbf{u} \notin V \cup W$. Since $\mathbf{u} \in V + W$, it follows that $V \cup W \neq V + W$.

Exercise 3.4. 1. There are many such examples, one is $A = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{smallmatrix}\right]$.

2. Here we repeat a column again, for example $B = \left[\begin{smallmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{smallmatrix}\right]$.

3. There are many such examples, one is $C = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$. Here we see that

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -\frac{1}{3} \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Exercise 3.5. The first matrix A is already in RREF, as

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

The first two are pivot columns and the last two are free columns, so it has two special solutions, which define the nullspace as

$$\operatorname{null}(A) = \operatorname{span}\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

The second matrix is almost in RREF, but we can bring it quickly there:

$$B = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This is the identity matrix, which we know has only the zero vector in its nullspace. For the third matrix, we again need to bering to RREF:

$$C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This has two pivot columns and two free columns, so it has two special solutions. The nullspace is the span of these, and is given by

$$\operatorname{null}(C) = \operatorname{span}\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Lecture 4: Completely solving $A\mathbf{x} = \mathbf{b}$

Exercise 4.1. The product is

$$\mathbf{v}\mathbf{w}^T = \begin{bmatrix} a \\ a \\ a \\ a \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} a & a & a & a \\ a & a & a & a \\ a & a & a & a \\ a & a & a & a \end{bmatrix}.$$

if $a = 0$, then we have the zero matrix, which has rank 0. But if a is any nonzero real number, then the the reduced row echelon form of A will be

$$\begin{bmatrix} a & a & a & a \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

which clearly has only one pivot. So in this case, the rank is 1.

Exercise 4.2. First we find the particular solutions. We get these by elimination on the augmented matrix $[A\ \mathbf{b}]$. The first multiplier is $\ell_{21} = 2$:

$$\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & -9 & -3 & 0 & 9 \\ 6 & 0 & -21 & 0 & 2 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & -9 & -3 & 0 & 9 \\ 0 & 0 & -3 & 6 & 2 & -19 \end{bmatrix}.$$

We see the pivots already as $3, -3$. Now we clear the -9 above the -3:

$$\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & -9 & -3 & 0 & 9 \\ 0 & 0 & -3 & 6 & 2 & -19 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 & -21 & -6 & 66 \\ 0 & 0 & -3 & 6 & 2 & -19 \end{bmatrix}.$$

Finally we multiply by the reciprocals of the pivots:

$$\begin{bmatrix} 1/3 & 0 \\ 0 & -1/3 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 & -21 & -6 & 66 \\ 0 & 0 & -3 & 6 & 2 & -19 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -7 & -2 & 22 \\ 0 & 0 & 1 & -2 & -2/3 & 19/3 \end{bmatrix}.$$

We find the particular solution immediately by placing the last column $\mathbf{d}$ in the pivot variable spots, and get $\mathbf{p} = [22\ 0\ 19/3\ 0\ 0]$. The special solutions, which we know there are 3 (as there are 3 free columns), come from considering $R\mathbf{x} = 0$. The three special solutions will have one 1 in each of the free variable spots, and 0 in the other free variable spots.

$$\begin{bmatrix} 1 & 0 & 0 & -7 & -2 \\ 0 & 0 & 1 & -2 & -2/3 \end{bmatrix} \begin{bmatrix} x_1 \\ 1 \\ x_3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \implies x_1 = 0,\ x_3 = 0$$

$$\begin{bmatrix} 1 & 0 & 0 & -7 & -2 \\ 0 & 0 & 1 & -2 & -2/3 \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \\ x_3 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \implies x_1 = 7,\ x_3 = 2$$

$$\begin{bmatrix} 1 & 0 & 0 & -7 & -2 \\ 0 & 0 & 1 & -2 & -2/3 \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \\ x_3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \implies x_1 = 2,\ x_3 = 2/3$$

Hence the complete solution is

$$\mathbf{x} = \begin{bmatrix} 22 \\ 0 \\ 19/3 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 7 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 2 \\ 0 \\ 2/3 \\ 0 \\ 1 \end{bmatrix},$$

for any $x_2, x_4, x_5 \in \mathbf{R}$.

Exercise 4.3. Note the answer is presented in the usual (particular solution) + (special solution) way, with the free column being the second one, since x_2 is the variable. In a particular solution the free variables are zero, which occurs in

$$\begin{bmatrix} 19 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -2 \end{bmatrix} + (-4) \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}.$$

This vector is still on the line of intersection, and we can build $A\mathbf{x} = \mathbf{b}$ from it. The augmented matrix $[R\ \mathbf{d}]$ from the equation $R\mathbf{x} = \mathbf{d}$, obtained via elimination, is

$$\begin{bmatrix} 1 & 3 & 0 & 19 \\ 0 & 0 & 1 & -2 \end{bmatrix}.$$

Here $R = \left[\begin{smallmatrix} 1 & 3 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right]$ could already be A, and $\mathbf{d} = [19\ 0\ -2]^T$ could already be $\mathbf{b}$. Or, we can add rows to get rid of the zeros:

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 6 & 3 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 17 \\ 32 \end{bmatrix}.$$

Exercise 4.4. First we compute the necessary products.

Lecture 5: Independence, basis, dimension

Exercise 5.1. Two choices are given below:

$$S_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}, \qquad S_1 = \left\{ \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

These are linearly independent because using them as columns of a matrix A, one quickly finds the RREF to be the 3×3 identity matrix.

Exercise 5.2. Here we again use Python, and take a 2×2 matrix to be a list of lists `[[a,b],[c,d]]`.

1. The following function takes a 2×2 matrix as input and returns `True` if one column is a multiple of the other, and `False` otherwise. We have an additional function that allows for computer precision up to 10 decimal points.

   ```
   def iszero(n):
       return (abs(n) < 1e-10)


   def twomult(mat):
       ratio1 = mat[0][1] / mat[0][0]
       ratio2 = mat[1][1] / mat[1][0]
       return iszero(ratio1 - ratio2)
   ```

   This does not take into account the possibility that one of the denominators could be zero.

Exercise 5.3. To express $\mathbf{v}$ in terms of the basis B, we solve the matrix equation $A\mathbf{x} = \mathbf{v}$, where the columns of A are the vectors of B. We use Gaussian elimination on the augmented matrix $[A\ \mathbf{v}]$, followed by back substitution:

$$\begin{bmatrix} 1 & -1 & 3 & -3 \\ 2 & 1 & 0 & -1 \\ 3 & -1 & 6 & 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -1 & 3 & -3 \\ 0 & 3 & -6 & 5 \\ 0 & 0 & 1 & 32/3 \end{bmatrix} \longrightarrow \begin{array}{rcl} x_3 = 32/3 & & \\ 3x_2 - 6x_3 = 5 & \implies & x_2 = 23 \\ x_1 - x_2 + x_3 = -3 & \implies & x_1 = -12 \end{array}$$

Hence we find that

$$\begin{bmatrix} -3 \\ -1 \\ 5 \end{bmatrix} = \frac{32}{3} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 23 \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} - 12 \begin{bmatrix} 3 \\ 0 \\ 6 \end{bmatrix}.$$

Exercise 5.4. A plane is 2-dimensional, so it should have two elements in the basis. Note that the defining equation may be expressed as

$$\begin{bmatrix} 2 & -4 & -5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

and bringing the matrix on the left to row reduced form we get

$$A = \begin{bmatrix} 2 & -4 & -5 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & -2 & -\frac{5}{2} \end{bmatrix} = R.$$

The nullspace of these matrices consists of precisely those vectors (x, y, z) which lie in the plane P. Note there are two free columns, so there are two special solutions. We find them quickly to be

$$\mathbf{s}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{s}_2 = \begin{bmatrix} \frac{5}{2} \\ 0 \\ 1 \end{bmatrix},$$

and get that the nullspace of the matrix A is the span of $\mathbf{s}_1$ and $\mathbf{s}_2$. Hence the plane P is the span of these equations. We did not check that $\{\mathbf{s}_1, \mathbf{s}_2\}$ is a linearly independent set, but because we know $\text{span}(\mathbf{s}_1, \mathbf{s}_2) = P$ and we know $\dim(P) = 2$, we must have that $\{\mathbf{s}_1, \mathbf{s}_2\}$ is linearly independent, because there are only two vectors in the set, and every basis of P must have 2 vectors. Hence $\{\mathbf{s}_1, \mathbf{s}_2\}$ is a basis for P.

Exercise 5.5. To find the change of basis matrix, we have to express the vectors of the second basis in terms of vectors from the first basis. This means solving the two matrix equations

$$\begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}, \qquad \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}.$$

Gaussian elimination on the first augmented matrix gives us,

$$\begin{bmatrix} 3 & -1 & -2 \\ 2 & 1 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 1/5 \\ 0 & 1 & 13/5 \end{bmatrix},$$

and on the second gives us

$$\begin{bmatrix} 3 & -1 & 0 \\ 2 & 1 & 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix}.$$

Hence the change of basis matrix is $\begin{bmatrix} 3 & 5 \\ 13/5 & 3 \end{bmatrix}$.

Exercise 5.6. The proof of the first claim follows from first observing that the intersection $U \cap W$ is closed under vector addition and scalar multiplication. Indeed, if $\mathbf{v} \in U \cap W$, the $\mathbf{v} \in U$ (so $c\mathbf{v} \in U$) and $\mathbf{v} \in W$ (so $c\mathbf{v} \in W$). Hence $c\mathbf{v} \in U \cap W$. A similar appproach works for vector addition. The zero element is in both U and W, and so must be in $U \cap W$. Additive inverses are -1 multiples, and so are also in the intersection. The other properties are inherited from U and W similarly.

The proof of the second claim comes from constructing a basis for $U \cap W$ that can be extended to bases of U and W separately.

The proof of the third claim comes by constructing an explicit basis $\{(\mathbf{u}, 0) : \mathbf{u} \in B_U\} \cup \{(0, \mathbf{w}) : \mathbf{w} \in B_W\}$ for $V \oplus W$, where B_U is a basis for U and B_W is a basis for W.

Lecture 6: The rank-nullity theorem

Exercise 6.1. The easiest case is when a, b, c are all nonzero. In that case, the row reduced echelon form of A and its tranpose are

$$\text{rref}(A) = \begin{bmatrix} 0 & 1 & 0 & 0 & a-b-ab+abc & a-b-ab+abc \\ 0 & 0 & 1 & 0 & b-bc & b-bc \\ 0 & 0 & 0 & 1 & c & c \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \text{rref}(A^T) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Lecture 7: Orthogonality

Exercise 7.1. Arbitrary elements in U and V are

$$U \ni \mathbf{u} = a_1\mathbf{u}_1 + \cdots a_k\mathbf{u}_k = \sum_{i=1}^{k} a_i\mathbf{u}_i, \qquad V \ni \mathbf{v} = b_1\mathbf{v}_1 + \cdots b_\ell\mathbf{v}_\ell = \sum_{j=1}^{\ell} b_j\mathbf{v}_j.$$

Their dot product, following the laws of dot products, is

$$\begin{aligned}
\mathbf{u}\cdot\mathbf{v} &= (a_1\mathbf{u}_1 + \cdots a_k\mathbf{u}_k)\cdot(b_1\mathbf{v}_1 + \cdots b_\ell\mathbf{v}_\ell) \\
&= a_1b_1\mathbf{u}_1\cdot\mathbf{v}_1 + a_1b_2\mathbf{u}_1\cdot\mathbf{v}_2 + \cdots + a_kb_k\mathbf{u}_k\cdot\mathbf{v}_\ell \\
&= \sum_{i=1}^{k}\sum_{j=1}^{\ell} a_ib_j\underbrace{\mathbf{u}_i\cdot\mathbf{v}_j}_{0} \\
&= 0
\end{aligned}$$

Exercise 7.2.

Exercise 7.3. For every $\mathbf{x} \in \mathbf{R}^n$, there is a decomposition $\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$, where $x \in \text{row}(A)$ and $w_n \in \text{null}(A)$. The desired function and its inverse are

$$\begin{aligned}
f\colon \text{row}(A) &\to \text{col}(A), & g\colon \text{col}(A) &\to \text{row}(A), \\
\mathbf{v} &\mapsto A\mathbf{v}, & \mathbf{w} &\mapsto A^T\mathbf{w}.
\end{aligned}$$

Lecture 8: Projections and least squares

Exercise 8.1. We expand the expression P^2 to get

$$P^2 = \left(\frac{1}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}\mathbf{v}^T\right)\left(\frac{1}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}\mathbf{v}^T\right) = \frac{(\mathbf{v}\mathbf{v}^T)(\mathbf{v}\mathbf{v}^T)}{(\mathbf{v}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{v})} = \frac{\mathbf{v}(\mathbf{v}^T\mathbf{v})\mathbf{v}^T}{(\mathbf{v}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{v})} = \frac{\mathbf{v}(\mathbf{v}\cdot\mathbf{v})\mathbf{v}^T}{(\mathbf{v}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{v})} = \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}\cdot\mathbf{v}} = P,$$

as desired.

Exercise 8.2.

Exercise 8.3. The normal vector is $\mathbf{n} = (3, 4, 9)$. Following Exercise 5.4, we find the plane $3x+4y-9z$ to be the column space of the matrix

$$A = \begin{bmatrix} -\frac{4}{3} & 3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Exercise 8.4. Recal the nullspace of A is all the vectors $\mathbf{x}$ for which $A\mathbf{x} = 0$. To see $\text{null}(A) \subseteq \text{null}(A^TA)$, suppose that $\mathbf{x} \in \text{null}(A)$. That is, $A\mathbf{x} = 0$, and multiplying by A^T on the left gives $A^TA\mathbf{x} = 0$, which means $\mathbf{x} \in \text{null}(A^TA)$. To see $\text{null}(A^TA) \subseteq \text{null}(A)$, suppose that $\mathbf{y} \in \text{null}(A^TA)$. That is, $A^TA\mathbf{y} = 0$, and multiplying by $\mathbf{y}^T$ on the left gives

$$0 = \mathbf{y}^T(A^TA\mathbf{y}) = (\mathbf{y}^TA^T)(A\mathbf{y}) = (A\mathbf{y})^T(A\mathbf{y}) = \|A\mathbf{y}\|.$$

Since the norm is positive definite, it follow that $A\mathbf{y} = 0$, and so $\mathbf{y} \in \text{null}(A)$.

Exercise 8.5. The mentioned examples look for the solution to

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 7 & 1 \\ 5 & 1 \end{bmatrix}}_{A}\begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}}_{\mathbf{b}}.$$

Using the projection matrix formula, the projection of $\mathbf{b}$ onto col(A) is

$$A(A^TA)^{-1}A^T\mathbf{b} = \frac{1}{14}\begin{bmatrix} 13 & -2 & 3 \\ -2 & 10 & 6 \\ 3 & 6 & 5 \end{bmatrix}\begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 59 \\ 20 \\ 33 \end{bmatrix}.$$

Now we are trying to solve the equation

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 7 & 1 \\ 5 & 1 \end{bmatrix}}_{A}\begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\frac{1}{14}\begin{bmatrix} 59 \\ 20 \\ 33 \end{bmatrix}}_{\mathbf{c}},$$

which requires sending the augmented matrix $[A\ \mathbf{c}]$ to row reduced form:

$$\begin{bmatrix} 1 & 1 & \frac{59}{14} \\ 7 & 1 & \frac{20}{14} \\ 5 & 1 & \frac{33}{14} \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & -\frac{13}{28} \\ 0 & 1 & \frac{131}{28} \\ 0 & 0 & 0 \end{bmatrix}.$$

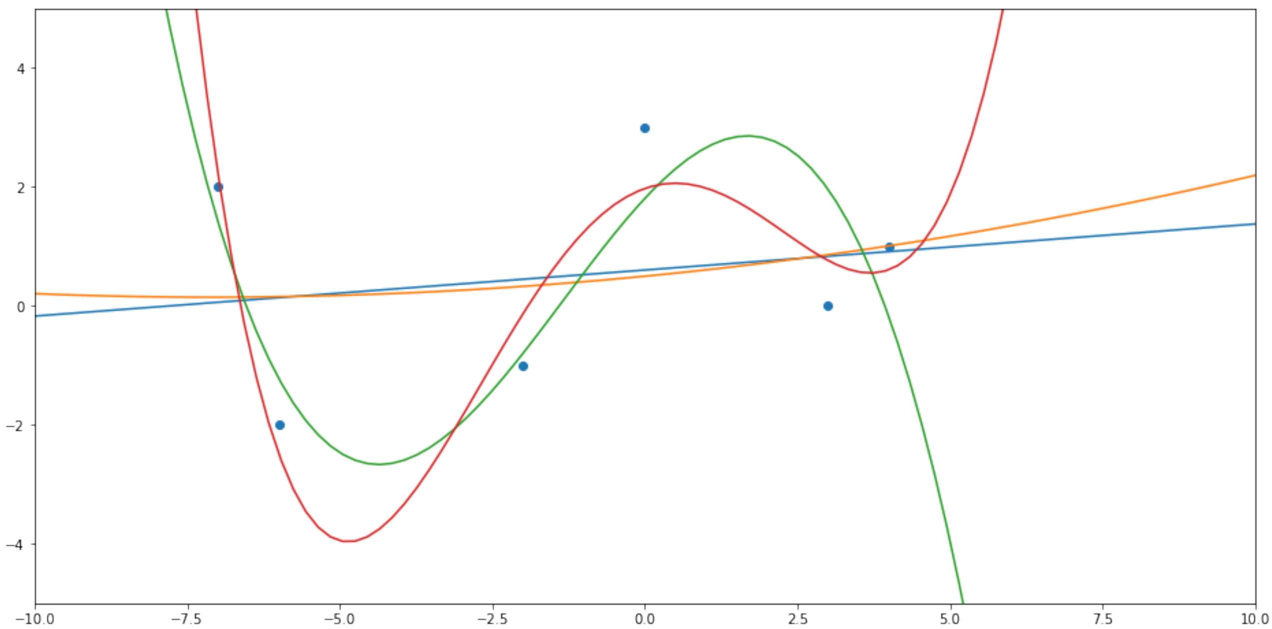
This produces the equation $y = -\frac{13}{28}x + \frac{131}{28}$, exactly as in the conclusion to Example 8.10.

Lecture 9: The Gram–Schmidt process

Exercise 9.1. The least squares polynomials are given below.

$$\begin{aligned} f_1(x) &= \frac{12}{155}x + \frac{187}{310} \\ f_2(x) &= \frac{943}{134556}x^2 + \frac{2229}{2426}x + \frac{67193}{134556} \\ f_3(x) &= -\frac{410881}{8205900}x^3 - \frac{137014}{683825}x^2 + \frac{9060313}{8205900}x + \frac{49420}{27353} \\ f_4(x) &= \frac{64739}{6474480}x^4 + \frac{487007}{55033080}x^3 - \frac{40794619}{110066160}x^2 + \frac{3957479}{11006616}x + \frac{3621483}{1834436} \end{aligned}$$

They are plotted together on the plot below. Note that the higher degree polynomials better approximate the points.



Exercise 9.2. Yes, the columns are orthogonal and orthonormal.

Lecture 10: Inner products and distances

Exercise 10.1. For each we give an example where it fails.

- For $A = B = -I$, the trace is negative, but the inner product is nonnegative.
- For $f = 1$, the derivative is zero, but f is not zero. The inner product is positive definite, so only if f is 0, can the inner product $\langle f, f\rangle$ be zero.
- For $a = 2, b = 4$, by multiplicativity we should have $\langle 2, 4\rangle = \langle 2, 2\cdot 2\rangle = 2\langle 2, 2\rangle$. But intsead we have $\langle 2, 4\rangle = 20$ and $2\langle 2, 2\rangle = 16$.

Exercise 10.2.

Exercise 10.3. As in the Gram–Schmidt process, we first get a basis A', B', C' that is orthogonal, then normalize to get A'', B'', C''. We begin by computing

$$\begin{aligned} A' = A &= \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \\ B' = B - \text{proj}_{A'}(B) &= \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} - \frac{\text{trace}\left(\left[\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right]^T \left[\begin{smallmatrix} 2 & 0 & -1 & 1 \end{smallmatrix}\right]\right)}{\text{trace}\left(\left[\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right]^T \left[\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right]\right)} \left[\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right] \end{aligned}$$

Exercise 10.4.

Lecture 11: Determinants, part 1

Exercise 11.1. There are many examples we may use. The identity matrix I and its negative $-I$ both have nonzero determinant, but their sum has a zero determinant, and so is not invertible:

$$I + (-I) = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\det(I)=1} + \underbrace{\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}}_{\det(-I)=-1} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\det(0)=0}.$$

Exercise 11.2. Since $AB = I$, the properties of the determinant tell us that

$$1 = \det(I) = \det(AB) = \det(A)\det(B).$$

Since $1 \neq 0$, we have that $\det(A) \neq 0$ and $\det(B) \neq 0$. This means that B has an inverse B^{-1} with $BB^{-1} = I$. Also from $AB = I$, we have

$$\begin{aligned} BAB &= B && \text{(multiply by } B \text{ on the left)} \\ BABB{-1} &= BB^{-1} && \text{(multiply by } B^{-1} \text{ on the right)} \\ BAI &= I && \text{(definition of inverse)} \\ BA &= I. && \text{(properties of the identity)} \end{aligned}$$

Exercise 11.3. There are 9 terms in the recursive definition, and we see that many cofactors are zero:

$$\det(A^{11}) = \begin{vmatrix} 2 & 0 \\ 2 & 0 \end{vmatrix} = 0 - 0 = 0 \quad \det(A^{12}) = \begin{vmatrix} 2 & 0 \\ 1 & 0 \end{vmatrix} = 0 - 0 = 0 \quad \det(A^{13}) = \begin{vmatrix} 2 & 2 \\ 1 & 1 \end{vmatrix} = 2 - 2 = 0$$
$$\det(A^{21}) = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = 0 - 1 = -1 \quad \det(A^{22}) = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = 0 - 1 = -1 \quad \det(A^{23}) = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = 1 - 1 = 0$$
$$\det(A^{31}) = \begin{vmatrix} 1 & 1 \\ 2 & 0 \end{vmatrix} = 0 - 2 = -2 \quad \det(A^{32}) = \begin{vmatrix} 1 & 1 \\ 2 & 0 \end{vmatrix} = 0 - 2 = -2 \quad \det(A^{33}) = \begin{vmatrix} 1 & 1 \\ 2 & 2 \end{vmatrix} = 2 - 2 = 0$$

That is, 5 of 9 cofactors are zero. For the 4 that are not zero, $A_{21} = 2$, $A_{22} = 2$, $A_{31} = 1$, $A_{32} = 1$, all of the multipliers are nonzero. Hence 4 of the 9 terms in the determinant definition are nonzero.

Exercise 11.4. Suppose that $A\mathbf{x} = 0$ for nonzero $\mathbf{x}$. This means that there is at least one nonzero vector in the nullspace of A, so $\dim(\text{null}(A)) \geqslant 1$. By the rank-nulltiy theorem, it follows that $\text{rank}(A) \leqslant n - 1$. Since the rank is the dimension of the row space, it follows that there is some linear dependence among the rows. By row operations, we can row reduce A to having at least one zero row. By the multilinear property of the determinant, it follows that $\det(A) = 0$.

Now suppose that $\det(A) = 0$. This means that the parallelotope with edges given by the columns of A has volume zero. This means that at least three of the edges lie on the same plane, which means there is linear dependence among the columns. Hence we can find an $\mathbf{x}$ for which $A\mathbf{x} = 0$.

Lecture 12: Determinants, part 2

Exercise 12.1. This follows by expressing the “swapping” operation as multiplying by a permutation matrix.

Exercise 12.2. The permutation σ is a single transposition $2 \leftrightarrow 3$, so the parity is -1. The permutation σ is a composition of $(1 \leftrightarrow 3)$ followed by $(2 \leftrightarrow 1)$, hence its parity is 1. The determinant of A is then

$$\begin{aligned}
\det(A) &= \text{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} A_{3\sigma(3)} A_{4\sigma(4)} + \text{sgn}(\rho) A_{1\rho(1)} A_{2\rho(2)} A_{3\rho(3)} A_{4\rho(4)} \\
&= (-1) A_{11} A_{23} A_{32} A_{44} + A_{13} A_{21} A_{32} A_{44} \\
&= (-1) \cdot 7 \cdot 2 \cdot (-2) \cdot 1 + (-1) \cdot 3 \cdot (-2) \cdot 1 \\
&= 28 + 6 \\
&= 34.
\end{aligned}$$

Note that no other term appears in the determinant, because starting out with A_{11} means we cannot have A_{21} as a factor in that term, and similarly for starting out with A_{13} does not allow A_{23} to be in the term.

Lecture 13: Eigenvalues and eigenvectors

Exercise 13.1. 1. The eigenvalues are the roots of

$$\det(A - \lambda I) = (6 - \lambda)(-2 - \lambda) + 25 = -12 - 4\lambda + \lambda^2 + 25 = \lambda^2 - 4\lambda + 13,$$

which has roots $\lambda = \frac{4 \pm \sqrt{16-52}}{2} = \frac{4 \pm \sqrt{-36}}{2} = 2 \pm 3i$. The associated eigenvectors are solved by the matrix equation

$$\begin{bmatrix} 6 & -5 \\ 5 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (2+3i) \begin{bmatrix} x \\ y \end{bmatrix} \iff \begin{array}{rcl} 6x - 5y &=& (2+3i)x \\ 5x - 2y &=& (2+3i)y \end{array} \implies \begin{array}{rcl} x &=& 4+3i \\ y &=& 5 \end{array}.$$

The second eigenvector is solved similarly, to produce the system

$$\lambda_1 = 2+3i, \quad \mathbf{v}_1 = \begin{bmatrix} 4+3i \\ 5 \end{bmatrix}, \qquad \lambda_2 = 2-3i, \quad \mathbf{v}_2 = \begin{bmatrix} 4-3i \\ 5 \end{bmatrix}.$$

2. The product is real:

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \begin{bmatrix} 4+3i \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 4-3i \\ 5 \end{bmatrix} = (4+3i)(4-3i)+25 = 16-12i+12i-9i^2+25 = 41+9 = 50.$$

Exercise 13.2. 1. There are many options. and the easiest ones are with the eigenvalues on the diagonal. Two such examples are:

$$A = \begin{bmatrix} -3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \qquad B = \begin{bmatrix} -3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

2. The eigenvectors are the three standard basis vectors of $\mathbf{R}^3$,

$$\mathbf{e}_1 = \left[\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right], \mathbf{e}_2 = \left[\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right], \mathbf{e}_3 = \left[\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right].$$

3. If $\lambda_3 = -2$, then $\lambda_3 = \lambda_2$. So for an arbitrary combination $a\mathbf{v}_2 + b\mathbf{v}_3, a, b \in \mathbf{R}$, we have

$$A(a\mathbf{v}_2 + b\mathbf{v}_3) = aA\mathbf{v}_2 + bA\mathbf{v}_3 = a\lambda_2\mathbf{v}_2 + b\lambda_3\mathbf{v}_3 = a\lambda_3\mathbf{v}_2 + b\lambda_3\mathbf{v}_3 = \lambda_3(a\mathbf{v}_2 + b\mathbf{v}_3).$$

Hence $a\mathbf{v}_2 + b\mathbf{v}_3$ is an eigenvector with eigenvalue λ_3.

Exercise 13.3. Just from the block description, the matrix A will have (at least) eigenvalues $-1, 2, 5$ by creating an eigenvector $\left[\begin{smallmatrix} \mathbf{v} \\ 0 \end{smallmatrix}\right] \in \mathbf{R}^6$ where $\mathbf{v}$ is an eigenvector of B.

Exercise 13.4. We have to do this in reverse. Let $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ be the answer to this question, which will then satisfy

$$A\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = \lambda\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \iff \left[\begin{smallmatrix} ax+by \\ cx+dy \end{smallmatrix}\right] = \left[\begin{smallmatrix} \lambda x \\ \lambda y \end{smallmatrix}\right] \qquad \text{and} \qquad A\left[\begin{smallmatrix} z \\ w \end{smallmatrix}\right] = \mu\left[\begin{smallmatrix} z \\ w \end{smallmatrix}\right] \iff \left[\begin{smallmatrix} az+bw \\ cz+dw \end{smallmatrix}\right] = \left[\begin{smallmatrix} \mu z \\ \mu w \end{smallmatrix}\right].$$

We could do back substitution, or we could write this as a matrix equation:

$$\begin{bmatrix} x & y & 0 & 0 \\ 0 & 0 & x & y \\ z & w & 0 & 0 \\ 0 & 0 & z & w \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \\ \mu z \\ \mu w \end{bmatrix}.$$

Row reducing the augmented matrix we find solutions to a, b, c, d in the last column:

$$\begin{bmatrix} x & y & 0 & 0 & \lambda x \\ 0 & 0 & x & y & \lambda y \\ z & w & 0 & 0 & \mu z \\ 0 & 0 & z & w & \mu w \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{\mu yz - \lambda xw}{yz - xw} \\ 0 & 1 & 0 & 0 & \frac{\mu xz - \lambda xz}{xw - yz} \\ 0 & 0 & 1 & 0 & \frac{\lambda yw - \mu yw}{xw - yz} \\ 0 & 0 & 0 & 1 & \frac{\mu xw - \lambda yz}{xw - yz} \end{bmatrix}.$$

Hence the matrix

$$\begin{bmatrix} \dfrac{\mu yz - \lambda xw}{yz - xw} & \dfrac{\mu xz - \lambda xz}{xw - yz} \\ \dfrac{\lambda yw - \mu yw}{xw - yz} & \dfrac{\mu xw - \lambda yz}{xw - yz} \end{bmatrix}$$

will have eigenvector $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ with eigenvalue λ, and eigenvector $\left[\begin{smallmatrix} z \\ w \end{smallmatrix}\right]$ with eigenvector μ. Note that the denominators cannot be zero, so there are relations that must be satisfied among the x, y, z, w for such a matrix to even exist.

Lecture 14: Diagonalization

Exercise 14.1. For the matrix A, note that the rows are multiples of each other, so $\lambda_1 = 0$. Since there are 2 eigenvalues (as it is a 2×2 matrix), and the sum of the eigenvalues is the trace, it follows that $\lambda_1 + \lambda_2 = 2 + 5 = 7$, so $\lambda_2 = 7$. For the eigenvectors, we eliminate the augmented matrices

$$\begin{bmatrix} 2 & 2 & 0 \\ 5 & 5 & 0 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} -5 & 2 & 0 \\ 5 & -2 & 0 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & -2/5 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So the eigenvectors are $\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$ for $\lambda_1 = 0$ and $\left[\begin{smallmatrix} 2/5 \\ 1 \end{smallmatrix}\right]$ for $\lambda_2 = 7$, giving the decomposition

$$\begin{bmatrix} 2 & 2 \\ 5 & 5 \end{bmatrix} = \begin{bmatrix} -1 & 2/5 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} -1 & 2/5 \\ 1 & 1 \end{bmatrix}^{-1} \quad \text{where} \quad \begin{bmatrix} -1 & 2/5 \\ 1 & 1 \end{bmatrix}^{-1} = \frac{-5}{7} \begin{bmatrix} 1 & -2/5 \\ -1 & -1 \end{bmatrix}.$$

For the matrix B, the eigenvalues are on the diagonal, but the eigenvectors are not so immediate. For $\lambda_1 = 1$ we have $\mathbf{e}_1$, but for $\lambda_2 = 4$ we need

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4x \\ 4y \\ 4z \end{bmatrix} \implies \begin{aligned} z &= 0 \\ 4y &= 4y \\ -3x &= -2y \end{aligned} \implies \mathbf{v}_2 = \begin{bmatrix} 2/3 \\ 1 \\ 0 \end{bmatrix}.$$

Similarly for $\lambda_3 = 6$, we need

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6x \\ 6y \\ 6z \end{bmatrix} \implies \begin{aligned} 6z &= 6z \\ -2y &= -5z \\ -5x &= -2y - 3z \end{aligned} \implies \mathbf{v}_3 = \begin{bmatrix} 8/5 \\ 5/2 \\ 1 \end{bmatrix}.$$

Hence the decomposition is

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2/3 & 8/5 \\ 0 & 1 & 5/2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2/3 & 8/5 \\ 0 & 1 & 5/2 \\ 0 & 0 & 1 \end{bmatrix}^{-1},$$

where

$$\begin{bmatrix} 1 & 2/3 & 8/5 \\ 0 & 1 & 5/2 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -2/3 & 1/15 \\ 0 & 1 & -5/2 \\ 0 & 0 & 1 \end{bmatrix}.$$

Exercise 14.2.

1. The eigenvector matrix X has the eigenvectors as columns, and the eigenvalues matrix Λ has the eigenvalues on the diagonal:

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

2. First we get the inverse of X by row reduction:

$$\begin{bmatrix} 1 & 0 & -1 & 1 & 0 & 0 \\ 2 & 1 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}.$$

Hence the matrix A is

$$X\Lambda X^{-1} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & -1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix} X^{-1}$$
$$= \begin{bmatrix} -1 & 0 & 3 \\ -2 & 2 & 3 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ -1 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -3 & 0 & 2 \\ -5 & 2 & -1 \\ 0 & 0 & -1 \end{bmatrix}.$$

Exercise 14.3.

Lecture 15: Special matrices

Exercise 15.1. 1. The eigenvalues are the roots of the charatceristic polynomial $(-\lambda)^2 + a^2$, which has roots $\lambda = \pm ia$.

2. The eigenvalues are the roots of the charatceristic polynomial

$$(-\lambda)\begin{vmatrix} ia-\lambda & 0 \\ 0 & -\lambda \end{vmatrix} + a\begin{vmatrix} 0 & ia-\lambda \\ -a & 0 \end{vmatrix} = \lambda^2(ia-\lambda) + a^2(ia-\lambda) = (\lambda^2 - a^2)(ia-\lambda),$$

which has roots $\lambda = \pm ia$. The root ia is a root of multiplicity 2.

3. We follow the pattern presented, and claim the matrix

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

has all imaginary eigenvalues. This matrix is clearly skew-symmetric, and its characteristic polynomial, expanding along the first then the last row, is

$$(-\lambda)\begin{vmatrix} -\lambda & a & 0 \\ -a & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix} - a\begin{vmatrix} 0 & -\lambda & a \\ 0 & -a & -\lambda \\ -a & 0 & 0 \end{vmatrix} = (-\lambda)^2\begin{vmatrix} -\lambda & a \\ -a & -\lambda \end{vmatrix} + (-a)^2\begin{vmatrix} -\lambda & a \\ -a & -\lambda \end{vmatrix}$$
$$= (\lambda^2 + a^2)(\lambda^2 + a^2),$$

which again has roots $\lambda = \pm ia$, both of which have multiplicity 2.

4. There are many examples, one is

$$A = \begin{bmatrix} a & a & a \\ a & 2a & 2a \\ a & 2a & 3a \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} a & a & a \\ 0 & a & a \\ 0 & 0 & a \end{bmatrix},$$

so all three pivots are a. The matrix is clearly symmetric.

Exercise 15.2. This follows from the rules of matrix multiplication. Recall that for $A \in \mathcal{M}_{m\times n}$ and $B \in \mathcal{M}_{n\times \ell}$, the product AB has ij-entry given by

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}.$$

And for A^T, we have $(A^T)_{ij} = A_{ji}$. Hence for $B = A^T$, the product AA^T has entries

$$(AA^T)_{ij} = \sum_{k=1}^{n} A_{ik}(A^T)_{kj} = \sum_{k=1}^{n} A_{ik}A_{jk},$$
$$(AA^T)_{ji} = \sum_{k=1}^{n} A_{jk}(A^T)_{ki} = \sum_{k=1}^{n} A_{jk}A_{ik}.$$

Since both lines are the same, the matrix is symmetric. The proof for A^TA is analogous.

Exercise 15.3. 1. It must be that $b = 0$, which is a $\frac{1}{7}$ chance. The values a, c can be anything, so the probability is $\frac{1}{7}$.

2. For A to be positive definite, the pivots on the diagonals must be positive. For each of a, c, this is a $\frac{3}{7}$ chance, so the probability is $\frac{9}{49}$.

Exercise 15.4. 1. For the matrix A, we have

$$\begin{bmatrix} 1 & 2 & 2 \\ 2 & a & 2 \\ 2 & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 2 \\ 0 & a-4 & -2 \\ 0 & -2 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 2 \\ 0 & a-4 & -2 \\ 0 & 0 & -3-\frac{4}{a-4} \end{bmatrix}.$$

For the pivots to be positive, we need $a - 4 > 0 \iff a > 4$ and

$$-3 - \frac{4}{a-4} > 0 \iff -3a + 12 - 4 > 0 \iff -3a > -8 \iff a > 8/3.$$

The two conditions $a > 4$ and $a > 8/3$ are satsified on $a > 4$.

For the matrix B, we have

$$\begin{bmatrix} b & 2 & 0 \\ 2 & b & 3 \\ 0 & 3 & b \end{bmatrix} \to \begin{bmatrix} b & 2 & 0 \\ 0 & b-\frac{4}{b} & 3 \\ 0 & 3 & b \end{bmatrix} \to \begin{bmatrix} b & 2 & 0 \\ 0 & b-\frac{4}{b} & 3 \\ 0 & 0 & b-\frac{9}{b-\frac{4}{b}} \end{bmatrix}.$$

For the pivots to be positive, we need $b > 0$ in the first row,

$$b - \frac{4}{b} > 0 \iff b^2 - 4 > 0 \iff b^2 > 4 \iff b > 2 \text{ or } b < -2$$

in the second row, and

$$b - \frac{9}{b-\frac{4}{b}} > 0 \iff b^2 - 4 - 9 > 0 \iff b^2 > 13 \iff b > \sqrt{13} \text{ or } b < -\sqrt{13}$$

in the third row. The intersection of all these conditions is $b > \sqrt{13}$.

2. For the eigenvalues, we find the roots of the characteristic polynomial:

$$\begin{aligned} \det(A - \lambda I) &= (\lambda^2 - \lambda(3+a) + 3a - 8)(\lambda + 1) \\ &\implies \lambda \in \left\{-1, \frac{3 + a \pm \sqrt{(3+a)^2 - 12a + 32}}{2}\right\}, \\ \det(B - \lambda I) &= (\lambda^2 - 2b\lambda + b^2 - 13)(\lambda - b) \\ &\implies \lambda \in \left\{b, \frac{2b \pm \sqrt{4b^2 - 4b^2 + 52}}{2}\right\} = \left\{b, b \pm \sqrt{13}\right\}. \end{aligned}$$

For A, the first eigenvalue -1 does not depend on a, so it will be negative irrespective of the value of a. For B, we need $b > 0$, and $b > -\sqrt{13}$ or $b > \sqrt{13}$, the intersection of which is

$b > \sqrt{13}$.

3. The upper left determinants of A are:

$$\begin{aligned}
\begin{vmatrix} 1 \end{vmatrix} &= 1, \\
\begin{vmatrix} 1 & 2 \\ 2 & a \end{vmatrix} &= a - 4, \\
\begin{vmatrix} 1 & 2 & 2 \\ 2 & a & 2 \\ 2 & 2 & 1 \end{vmatrix} &= \begin{vmatrix} a & 2 \\ 2 & 1 \end{vmatrix} - 2\begin{vmatrix} 2 & 2 \\ 2 & 1 \end{vmatrix} + 2\begin{vmatrix} 2 & a \\ 2 & 2 \end{vmatrix} \\
&= a - 4 - 2(2-4) + 2(4-2a) \\
&= a - 4 - 4 + 8 + 8 - 4a \\
&= 8 - 3a,
\end{aligned}$$

which all will be positive if $a - 4 > 0$ (or $a > 4$) and $8 - 3a > 0$ (or $a < 8/3$). Both of these conditions can not be satsified at the same time, so there are no values a for which all the uper left determinants will be positive. For B:

$$\begin{aligned}
\begin{vmatrix} b \end{vmatrix} &= b, \\
\begin{vmatrix} b & 2 \\ 2 & b \end{vmatrix} &= b^2 - 4, \\
\begin{vmatrix} b & 2 & 0 \\ 2 & b & 3 \\ 0 & 3 & b \end{vmatrix} &= b\begin{vmatrix} b & 3 \\ 3 & b \end{vmatrix} - 2\begin{vmatrix} 2 & 3 \\ 0 & b \end{vmatrix} \\
&= b(b^2 - 9) - 2(2b - 0) \\
&= b^3 - 9b - 4b \\
&= b(b^2 - 13),
\end{aligned}$$

which will all be positive if $b > 0$, $b^2 - 4 > 0$ (or $b > 2$ or $b < -2$), and $b(b^2 - 13) > 0$. The last condition equals 0 when $b = 0$ and $b = \pm\sqrt{13}$, so we have that:

$$\begin{aligned}
b < -\sqrt{13} &\implies b(b^2 - 13) < 0, \\
-\sqrt{13} < b < 0 &\implies b(b^2 - 13) > 0, \\
0 < b < \sqrt{13} &\implies b(b^2 - 13) < 0, \\
b > \sqrt{13} &\implies b(b^2 - 13) > 0.
\end{aligned}$$

We get this by observing that $b(b^2 - 13)$ has a positive leading coefficient (as a polynomial in b), so its value will be negative as $b \to -\infty$, and poisitive as $b \to +\infty$. Since all the roots have multiplicity 1, the sign changes at every root. We then must satisfy the conditions

$$b > 0, \qquad b > 2 \text{ or } b < -2, \qquad b \in (-\sqrt{13}, 0) \text{ or } b > \sqrt{13}.$$

Since $\sqrt{13} > 2$, the intersection of all these conditions is $b > \sqrt{13}$.

4. We choose $b = 4 > \sqrt{13}$. We already know the eigenvalues will be $\{4, 4 \pm \sqrt{13}\}$ from part (b). For the eigenvectors, we row reduce matrices of the eigenvector equations. For $\lambda_1 = 4$, we row have

$$\begin{bmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & 3 & 0 \\ 0 & 3 & 0 & 0 \end{bmatrix} \xrightarrow{\text{row ops.}} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 3/2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Hence the associated eigenvector is $[\,-\frac{3}{2}\ 0\ 1\,]^T$, which normalized is $[\,-\frac{3}{\sqrt{13}}\ 0\ \frac{2}{\sqrt{13}}\,]^T$. For the eigenvalues $4 \pm \sqrt{13}$, we have

$$\begin{bmatrix} \pm\sqrt{13} & 2 & 0 & 0 \\ 2 & \pm\sqrt{13} & 3 & 0 \\ 0 & 3 & \pm\sqrt{13} & 0 \end{bmatrix} \xrightarrow{\text{row ops.}} \begin{bmatrix} 1 & 0 & -2/3 & 0 \\ 0 & 1 & \pm\sqrt{13}/3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Hence the associated eigenvector to $4 \pm \sqrt{13}$ is $[\,\frac{2}{3}\ \mp\frac{\sqrt{13}}{3}\ 1\,]$, which normalized is $[\,\frac{\sqrt{2}}{\sqrt{13}}\ \mp\frac{1}{\sqrt{2}}\ \frac{3}{\sqrt{26}}\,]^T$. Hence the $Q\Lambda Q^T$-decomposition of B, for $b = 4$ is

$$Q = \frac{1}{\sqrt{13}} \begin{bmatrix} -3 & \sqrt{2} & \sqrt{2} \\ 0 & -\sqrt{\frac{13}{2}} & \sqrt{\frac{13}{2}} \\ 2 & \frac{3}{\sqrt{2}} & \frac{3}{\sqrt{2}} \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4+\sqrt{13} & 0 \\ 0 & 0 & 4-\sqrt{13} \end{bmatrix}.$$

Lecture 16: Singular value decomposition

Exercise 16.1. 1.

Exercise 16.2. 1. First we compute these matrices as

$$AA^T = \begin{bmatrix} 2a^2 & 0 \\ 0 & 4a^2 \end{bmatrix}, \qquad A^TA = \begin{bmatrix} a^2 & 0 & a^2 & 0 \\ 0 & 0 & 0 & 0 \\ a^2 & 0 & a^2 & 0 \\ 0 & 0 & 0 & 4a^2 \end{bmatrix}.$$

The eigenvalue / eigenvector pairs of AA^T are evidently $\lambda_1 = 4a^2$ with $\mathbf{u}_1 = [\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}]$ and $\lambda_2 = 2a^2$ with $\mathbf{u}_2 = [\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}]$. We sort them this way because $4a^2 > 2a^2$. For the SVD we only need the eigenvectors of A^TA corresponding to these two eigenvalues. It is immediate that $\mathbf{v}_1 = [0\ 0\ 0\ 1]^T$ and $\mathbf{v}_2 = [1\ 0\ 1\ 0]^T$, which normalizes to $[1/\sqrt{2}\ 0\ 1/\sqrt{2}\ 0]^T$. Hence the SVD of A is

$$\begin{aligned} A &= 2a \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} + \sqrt{2}a \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 2a & 0 & 0 & 0 \\ 0 & \sqrt{2}a & 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \end{bmatrix}}_{V^T}. \end{aligned}$$

2. The dimensions of the four fundamental subspaces are given by the number of rows and columns in the matrices U, Σ, V^T. We get that:

- $\dim(\text{col}(A)) = \text{rank}(A) = 2$
- $\dim(\text{null}(A^T)) = (\text{number of zero rows in } \Sigma) = 0$
- $\dim(\text{row}(A)) = \text{rank}(A) = 2$
- $\dim(\text{null}(A)) = (\text{number of zero columns in } \Sigma) = 2$

Exercise 16.3. 1. There are many examples, one is $A = \left[\begin{smallmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{smallmatrix}\right]$. The matrix $AA^T \in \mathcal{M}_{3\times 3}$ is diagonal with $1, 4, 9$ on its diagonal, so those are its eigenvalues. The singular values of A are the positive square roots of these numbers, and those are $1, 2, 3$.

2. Take the left singular vectors to be same as the right ones. Let $\sigma_1 = 4$ (to clear denominators) be the only signular value (because rank is 1). By SVD we get

$$A = \underbrace{\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}}_{V^T} = \begin{bmatrix} 2 & 0 \\ 2\sqrt{3} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix}.$$

3. Since A is symmetric, its singular values are its eigenvalues. Since there are many zeros, the eigenvalue / eigenvector pairs can be found by sight:

$$\begin{array}{lllll} \lambda_1 & = & \sigma_1 & = & 2 \\ \lambda_2 & = & \sigma_2 & = & 1 \\ \lambda_3 & = & \sigma_3 & = & 1 \end{array} \qquad \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

The last eigenvalue is zero because the matrix has two equal rows (so the determinant is 0). The approximations then are:

$$\text{rank } 1: \quad \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T = 2 \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

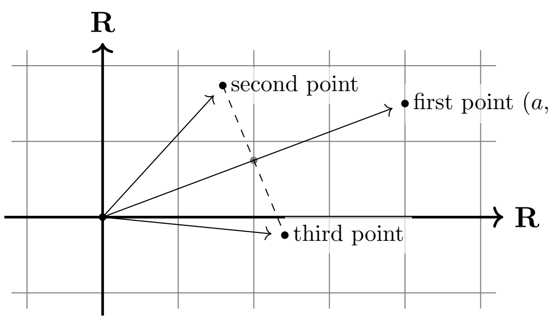
$$\text{rank } 2: \quad \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{other rank } 2: \quad \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_3 \mathbf{u}_3 \mathbf{v}_3^T = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Lecture 17: Principal component analysis

Exercise 17.1.

Exercise 17.2. Since the first principal component solves the perpendicular least squares problem, we choose one point to be exactly $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$, and the other two to lie the same distance on either side of this eigenvector. We choose the distance to be $\ell = \sqrt{a^2 + b^2}/4$ so that the two other points do not dominate the first point. The idea is given in the picture below.



We now construct these points explicitly and perform PCA on the data to confirm that the result will be as desired. To find the coordinates of the other two points, note that the slope of the line to (a, b)

is $\frac{b}{a}$. So the two other points lie on the line with slope $\frac{-a}{b}$ which goes through $(\frac{a}{2}, \frac{b}{2})$. The equation of the line is given by

$$\frac{-a}{b} = \frac{y - \frac{b}{2}}{x - \frac{a}{2}} \quad \Longleftrightarrow \quad f(x) = y = \frac{-a}{b}x + \left(\frac{a^2}{2b} + \frac{b}{2}\right).$$

To find the points a distance ℓ along this line from $(\frac{a}{2}, \frac{b}{2})$, we solve for x in the equality

$$\begin{aligned}
\frac{\sqrt{a^2+b^2}}{4} &= \sqrt{\left(\frac{a}{2} - x\right)^2 + \left(\frac{b}{2} - f(x)\right)^2} \\
&= \sqrt{\left(\frac{a}{2} - x\right)^2 + \left(\frac{b}{2} - \left(\frac{-a}{b}x + \frac{a^2}{2b} + \frac{b}{2}\right)\right)^2} \\
&= \sqrt{\left(\frac{a}{2} - x\right)^2 + \left(\frac{a}{b}x - \frac{a^2}{2b}\right)^2} \\
&= \sqrt{\left(\frac{a}{2} - x\right)^2 + \frac{a^2}{b^2}\left(x - \frac{a}{2}\right)^2} \\
&= \sqrt{\left(\frac{a}{2} - x\right)^2 \left(1 + \frac{a^2}{b^2}\right)}.
\end{aligned}$$

This simplifies to $x = \frac{2a \pm b}{4}$, so the data we have is

$$A = \begin{bmatrix} a & \frac{2a-b}{4} & \frac{2a+b}{4} \\ b & f(\frac{2a-b}{4}) & f(\frac{2a+b}{4}) \end{bmatrix} = \begin{bmatrix} a & \frac{2a-b}{4} & \frac{2a+b}{4} \\ b & \frac{2b+a}{4} & \frac{2b-a}{4} \end{bmatrix}.$$

For PCA, we need to mean-center the data first. The mean of x-coordinates is $2a/3$ and the mean of the y-coordinates is $2b/3$, so after subtracting $2a/3$ from the first row and $2b/3$ from the second row, we get the mean centered data to be

$$M = \begin{bmatrix} \frac{a}{3} & \frac{-2a-3b}{12} & \frac{3b-2a}{12} \\ \frac{b}{3} & \frac{3a-2b}{12} & \frac{-3a-2b}{12} \end{bmatrix} \quad \Longrightarrow \quad S = \frac{MM^T}{2} = \begin{bmatrix} \frac{4a^2+3b^2}{48} & \frac{ab}{48} \\ \frac{ab}{48} & \frac{3a^2+4b^2}{48} \end{bmatrix}.$$

With the help of a computer, we find the eigenvalues and eigenvectors of this symmetric matrix to be

$$\lambda_1 = \frac{a^2+b^2}{12}, \quad \mathbf{u}_1 = \begin{bmatrix} a/b \\ 1 \end{bmatrix}, \qquad \lambda_2 = \frac{a^2+b^2}{16}, \quad \mathbf{u}_2 = \begin{bmatrix} -b/a \\ 1 \end{bmatrix}.$$

It looks like we are done, but the eigenvector $\begin{bmatrix} a/b \\ 1 \end{bmatrix}$ is for the mean-centered data, so we need to shift it back. Hence the first eigenvector for the original data is

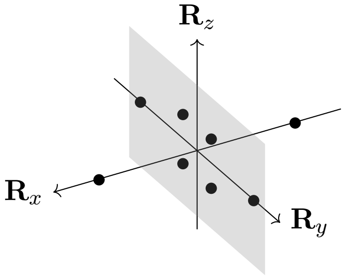
$$\begin{bmatrix} a/b \\ 1 \end{bmatrix} + \begin{bmatrix} 2a/3 \\ 2b/3 \end{bmatrix} = \begin{bmatrix} \frac{2ab+3a}{3b} \\ \frac{2b+3}{3} \end{bmatrix} = \frac{2b+3}{3b}\begin{bmatrix} a \\ b \end{bmatrix},$$

which is indeed a multiple of $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$, as desired.

Exercise 17.3. 1. We folow a similar method as in Question 1, placing what were the second and third points in the second eigenvector direction. We make some other changes:

- Since we need at least 4 points, but cannot place three points in a line, we split up what were the second and third points.
- To ensure the second principal component is $\left[\begin{smallmatrix} 0 & 1 & 0 \end{smallmatrix}\right]^T$, we place further points along the second principal component axis.
- To ensure that the data is mean-centered, we mirror all the points.

This construction is demonstrated in the picture below left (the plane $x = 0$ is emphasized in gray), with the points in the matrix below right.



$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 1 & -1 \\ 0 & 0 & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & 0 & 0 \end{bmatrix}$$

It is evident that no three points lie on a line. The mean-centered matrix M is the same as A, since the mean of each row is 0. The sample covariance matrix S and its eigenvectors are

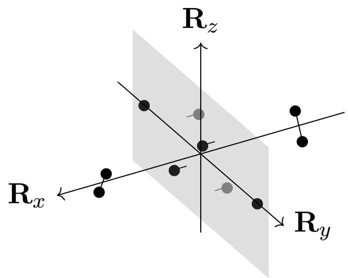
$$S = \begin{bmatrix} \frac{2}{7} & 0 & 0 \\ 0 & \frac{9}{28} & 0 \\ 0 & 0 & \frac{1}{28} \end{bmatrix}, \qquad \begin{aligned} \mathbf{u}_1 &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T \\ \mathbf{u}_2 &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \\ \mathbf{u}_3 &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \end{aligned} .$$

Hence the presented data satisfies the given conditions.

2. Columns 1,2,7,8 lie in plane $z = 0$ and columns 3-8 lie in plane $x = 0$ (emphasized in the picture above). To fix theses issues, we take two steps:

   - For the first issue, we split the points $(1, 0, 0)$ and $(-1, 0, 0)$ into two points just above and below the x-axis. We shift them in equal but opposite diections along the y-axis so that the new points are not on a plane.
   - For the the second issue, we move the four points in columns 3-6 by equal but opposite distances in the x-direction.
   - To ensure the two solutions do not conflict, the shifting magnitudes are different.

   The new data is given below left (with lines indicating shifts from the previous data), and the new matrix is given below right.



$$A = \begin{bmatrix} 1 & 1 & -1 & -1 & \frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & \frac{1}{8} & 0 & 0 \\ \frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & \frac{1}{16} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 1 & -1 \\ \frac{1}{8} & -\frac{1}{8} & \frac{1}{8} & -\frac{1}{8} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & 0 & 0 \end{bmatrix}$$

   By sight we confirm that no four points of these samples line in the same plane. The data is still mean-centered (since we added equal but opposite values to each row), and the sample covariance matrix with its eigenvectors is

$$S = \begin{bmatrix} \frac{65}{144} & 0 & 0 \\ 0 & \frac{145}{576} & 0 \\ 0 & 0 & \frac{5}{144} \end{bmatrix}, \qquad \begin{aligned} \mathbf{u}_1 &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T \\ \mathbf{u}_2 &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \\ \mathbf{u}_3 &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \end{aligned}.$$

   Hence all the conditions are satisfied.

Lecture 18: Linear transformations

Exercise 18.1. 1. T_1 is linear, and its matrix is a permutation matrix:

$$T_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

T_2 is not linear, as $T_2 \left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 2e^2 \\ 0 \end{smallmatrix}\right] \neq \left[\begin{smallmatrix} 4e \\ 0 \end{smallmatrix}\right] = 2T_2 \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$.

T_3 is not linear, as $T_3 \left[\begin{smallmatrix} 3 \\ 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 9 \\ 0 \end{smallmatrix}\right] \neq \left[\begin{smallmatrix} 3 \\ 0 \end{smallmatrix}\right] = 3T_3 \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$.

T_4 is not linear, as $T_4 \left[\begin{smallmatrix} \sqrt{\pi} \\ 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right] \neq \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] = \sqrt{2} T_4 \left[\begin{smallmatrix} \sqrt{\pi/2} \\ 0 \end{smallmatrix}\right]$.

T_5 is not linear, as $T_5 \left[\begin{smallmatrix} 0 \\ 0 \\ 3 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 \\ 0 \\ 9 \end{smallmatrix}\right] \neq \left[\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right] = 3T_5 \left[\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right]$.

T_6 is linear, and its matrix is the zero matrix:

$$T_6 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

T_7 is linear, and its matrix can be found by what it does to each variable:

$$T_7 = \begin{bmatrix} -3 & 0 & 0 \\ -0 & 1 & 1 \end{bmatrix}.$$

T_8 is linear, and its matrix can be found by what it does to each variable:

$$T_8 = \begin{bmatrix} 2 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

2. The three conditions that are given can be simplified using the following observations:

$$T_5 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad T_8 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \qquad T_8 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Using this, we get a clearer description of what S does to $\mathbf{R}^3$:

$$S \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad S \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \qquad S \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

To get the matrix of S, we first describe what S does on the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Note that

$$S \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = S \left(\frac{1}{2} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) = \frac{1}{2} S \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{2} S \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$$

for $\mathbf{e}_1$, and

$$S \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = S \left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) = S \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} S \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{2} S \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 0 \\ 1/2 \end{bmatrix}$$

for $\mathbf{e}_3$. For $\mathbf{e}_2$ we already know what happens. Applying the proof of Theorem 18.9 on the construction of the matrix associated to a linear transformation, we get that the matrix of S is

$$S = \begin{bmatrix} -1/2 & 1 & 3/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}.$$

Exercise 18.2. We show the proof for the kernel. The additive inverse exists, because if $\mathbf{x} \in \ker(f)$, then $f(\mathbf{x}) = 0$, and $f(-\mathbf{x}) = -f(\mathbf{x}) = -0 = 0$. The kernel is closed under addition, as $\mathbf{x}, \mathbf{y} \in \ker(f)$ means $f(\mathbf{x}) = f(\mathbf{y}) = 0$, so

$$f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) = 0 + 0 = 0 \quad \Longrightarrow \quad \mathbf{x}+\mathbf{y} \in \ker(f).$$

Exercise 18.3.

Exercise 18.4. 1. The lengths are $\|\mathbf{x}\| = \sqrt{10}$, $\|\mathbf{y}\| = \sqrt{10}$, and $\|\mathbf{z}\| = 2$. Hence the unit vectors are

$$\hat{\mathbf{x}} = \begin{bmatrix} 1/\sqrt{10} \\ 0 \\ 3/\sqrt{10} \end{bmatrix}, \qquad \hat{\mathbf{y}} = \begin{bmatrix} 3/\sqrt{10} \\ 0 \\ -1/\sqrt{10} \end{bmatrix}, \qquad \hat{\mathbf{z}} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}.$$

2. For this we use the diagonalization equation $A = Q\Lambda Q^T$, where Q has orthonormal columns.

We note that

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{z} = \mathbf{y} \cdot \mathbf{z} = 0,$$

so the vectors re all orthogonal. Normalizing them, as we have done above, makes them orthonormal. We choose eigenvalues 1,2,3 for $\mathbf{x}, \mathbf{y}, \mathbf{z}$, respectively. The eigenvector matrix is then

$$Q = \begin{bmatrix} 1/\sqrt{10} & 3/\sqrt{10} & 0 \\ 0 & 0 & -1 \\ 3/\sqrt{10} & -1/\sqrt{10} & 0 \end{bmatrix},$$

and the requested symmetric matrix is

$$A = Q\Lambda A^T = \begin{bmatrix} 1/\sqrt{10} & 3/\sqrt{10} & 0 \\ 0 & 0 & -1 \\ 3/\sqrt{10} & -1/\sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{10} & 3/\sqrt{10} & 0 \\ 0 & 0 & -1 \\ 3/\sqrt{10} & -1/\sqrt{10} & 0 \end{bmatrix}^T = \begin{bmatrix} \frac{19}{10} & 0 & -\frac{3}{10} \\ 0 & 3 & 0 \\ -\frac{3}{10} & 0 & \frac{11}{10} \end{bmatrix}.$$

3. We follow the proof of Theorem 18.9 to get a matrix for this linear transformation. First we compute what $f(\mathbf{e}_1)$, $f(\mathbf{e}_2)$, and $f(\mathbf{e}_3)$ will be, from the given data. We apply the two properties of linearity:

$$f\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = f\left(\frac{1}{10}\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + \frac{3}{10}\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}\right) = \frac{1}{10}f\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + \frac{3}{10}f\begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{3}{10}\begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1/5 \\ -1/5 \\ -3/10 \end{bmatrix},$$

$$f\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = f\left(-\frac{1}{2}\begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix}\right) = -\frac{1}{2}f\begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1/2 \\ 1/2 \end{bmatrix},$$

$$f\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = f\left(\frac{3}{10}\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} - \frac{1}{10}\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}\right) = \frac{3}{10}f\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} - \frac{1}{10}\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} = \frac{3}{10}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{10}\begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2/5 \\ 2/5 \\ 1/10 \end{bmatrix}.$$

Hence the matrix of f is

$$f\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right] = \begin{bmatrix} -1/5 & 0 & 2/5 \\ -1/5 & -1/2 & 2/5 \\ -3/10 & 1/2 & 1/10 \end{bmatrix}.$$

Exercise 18.5.

Exercise 18.6. Let $f\colon V \to W$ and $g\colon W \to Z$ be linear. Take any $\mathbf{x}, \mathbf{y} \in V$, for which

$$(g \circ f)(\mathbf{x} + \mathbf{y}) = g(f(\mathbf{x} + \mathbf{y})) = g(f(\mathbf{x}) + f(\mathbf{y})) = g(f(\mathbf{x})) + g(f(\mathbf{y})) = (g \circ f)(\mathbf{x}) + (g \circ f)(\mathbf{y}).$$

The second equality follows from the linearity of f, and the third equality follows from the linearity of g. Similarly, for any $c \in \mathbf{R}$, we have that

$$(g \circ f)(c\mathbf{x}) = g(f(c\mathbf{x})) = g(cf(\mathbf{x})) = cg(f(\mathbf{x})) = c(g \circ f)(\mathbf{x}),$$

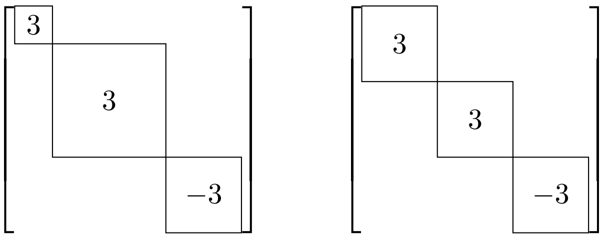
where again the second equality follows from the linearity of f, and the third equality follows from the linearity of g. Hence $g \circ f$ satsifies the linearity conditions, and is a linear map $V \to Z$.

Lecture 19: Jordan form

Exercise 19.1.

Exercise 19.2. 1. The eigenvalue -3 contributes a 2×2 Jordan block, since the algebraic multiplicity is 2 (so all its Jordan blocks together have 2 rows and 2 columns) and the geometric multiplicity is 1 (so there is only one Jordan block corresponding to this eigenvalue). Similarly, the Jordan blocks for the eigenvalue 3 take up 4 rows and 4 columns, and there are 2 of them. Hence:

- there are 3 Jordan blocks
- their sizes are either 1,3,2 or 2,2,2:



2. For the matrix B, we need to find an invertible 6×6 matrix C for which $B = CJC^{-1}$, as the J and B will be similar. We need B to have no zero entries, and generating several random matrices with entries in the range $\{-1, 0, 1\}$, we quickly find one (there is not a unique answer). We see that

$$B = \underbrace{\begin{bmatrix} -1 & -1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 & -1 & 1 \\ 1 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 \\ 1 & -1 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & -3 & 1 \\ 0 & 0 & 0 & 0 & 0 & -3 \end{bmatrix}}_{J} \underbrace{\begin{bmatrix} 2 & 1 & 2 & 3 & 1 & -1 \\ -2 & -1 & -1 & -2 & -1 & 2 \\ -3 & -2 & -2 & -3 & -1 & 3 \\ 4 & 2 & 3 & 5 & 1 & -2 \\ -3 & -2 & -2 & -4 & -1 & 2 \\ 3 & 2 & 2 & 3 & 1 & -2 \end{bmatrix}}_{C^{-1}} = \begin{bmatrix} 12 & 5 & 6 & 16 & 3 & -6 \\ -39 & -23 & -26 & -45 & -13 & 26 \\ 15 & 11 & 13 & 20 & 5 & -10 \\ -3 & -2 & -2 & -6 & -1 & 2 \\ -16 & -11 & -10 & -15 & -3 & 12 \\ -14 & -10 & -9 & -13 & -5 & 13 \end{bmatrix}.$$

3. Applying Theorem 19.10 from the lecture notes and the fact that $J = C^{-1}BC$, we get that the generalized eigenvectors of B are the columns of C.

Lecture 20: Complex numbers and complex matrices

Exercise 20.1. We place z in the denominator and multiply by the conjugate:

$$\frac{1}{z} = \frac{1}{x+yi} = \frac{1}{x+yi}\frac{x-yi}{x-yi} = \frac{x-yi}{x^2+y^2} = \frac{x}{x^2+y^2} + \frac{-y}{x^2+y^2}i.$$

Exercise 20.2. 1.

$$\begin{aligned} \overline{z+w} &= \overline{(x+yi)+(a+bi)} \\ &= \overline{(x+a)+(y+b)i} \\ &= (x+a)-(y+b)i \\ &= (a-yi)+(a-bi) \\ &= \overline{z}+\overline{w} \end{aligned}$$

2. *This equation originally looked like "$\overline{zw} = \overline{zw}$", where the space is not visible between the bars over the symbols on the right side. An answer stating that that both sides are the same is

acceptable.*

$$\begin{aligned}
\overline{zw} &= \overline{(x+yi)(a+bi)} \\
&= \overline{xa + xbi + yai - yb} \\
&= \overline{(xa - yb) + (xb + ya)i} \\
&= (xa - yb) - (xb + ya)i \\
&= xa - yb - xbi - yai \\
&= (x - yi)a - (x - yi)bi \\
&= (x - yi)(a - bi) \\
&= \overline{z}\ \overline{w}
\end{aligned}$$

3.
$$\overline{\overline{z}} = \overline{\overline{x+yi}} = \overline{x - yi} = x + yi = z$$

4.
$$z + \overline{z} = (x+yi) + (x-yi) = (x+x) + (y-y)i = 2x$$

5.
$$z - \overline{z} = (x+yi) - (x-yi) = (x-x) + (y+y)i = 2yi$$

6. Since $zz^{-1} = 1$, we have that

$$z^{-1} = \frac{1}{z} = \frac{1}{x+yi} = \frac{1}{x+yi}\frac{x-yi}{x-yi} = \frac{x-yi}{x^2+y^2} = \frac{\overline{z}}{|z|^2}.$$

7. Suppose that $|z| = 0$. Then

$$0 = |z| = \sqrt{x^2+y^2} \quad \Longrightarrow \quad 0 = x^2 + y^2.$$

Since $x^2 \geqslant 0$ and $y^2 \geqslant 0$, but their sum is equal to zero, it must be that $x = y = 0$, so $z = 0$. Conversely, suppose that $z = 0$. Then $|z| = \sqrt{0^2} = 0$.

8.
$$|\overline{z}| = |\overline{x+yi}| = |x - yi| = \sqrt{x^2 + (-y)^2} = \sqrt{x^2+y^2} = |x+yi| = |z|$$

9.

$$\begin{aligned}
|zw| &= |(x+yi)(a+bi)| \\
&= |xa + xbi + yai - yb| \\
&= |(xa - yb) + (xb + ya)i| \\
&= \sqrt{(xa-yb)^2 + (xb+ya)^2} \\
&= \sqrt{(xa)^2 - 2xayb + (yb)^2 + (xb)^2 + 2xbya + (ya)^2} \\
&= \sqrt{(xa)^2 + (yb)^2 + (xb)^2 + (ya)^2} \\
&= \sqrt{(x^2+y^2)(a^2+b^2)} \\
&= \sqrt{x^2+y^2}\sqrt{a^2+b^2} \\
&= |z||w|
\end{aligned}$$

10. For this question we work backwards, doing invertible operations (adding / subtracting, multipliying / dividing by nonzero numbers):

$$
\begin{aligned}
& |z+w| \leqslant |z|+|w| && \\
\iff\ & |(x+yi)+(a+bi)| \leqslant |x+yi|+|a+bi| && \text{(expanding)}\\
\iff\ & |(x+a)+(y+b)i| \leqslant |x+yi|+|a+bi| && \text{(expanding)}\\
\iff\ & \sqrt{(x+a)^2+(y+b)^2} \leqslant \sqrt{x^2+y^2}+\sqrt{a^2+b^2} && \text{(definition)}\\
\iff\ & (x+a)^2+(y+b)^2 \leqslant x^2+y^2+2\sqrt{(x^2+y^2)(a^2+b^2)}+a^2+b^2 && \text{(squaring)}\\
\iff\ & x^2+2ax+a^2+y^2+2yb+b^2 \leqslant x^2+y^2+2\sqrt{(x^2+y^2)(a^2+b^2)}+a^2+b^2 && \text{(expanding)}\\
\iff\ & 2ax+2yb \leqslant 2\sqrt{(x^2+y^2)(a^2+b^2)} && \text{(cancelling)}\\
\iff\ & ax+yb \leqslant \sqrt{(x^2+y^2)(a^2+b^2)} && \text{(dividing by 2)}\\
\iff\ & (ax)^2+2axyb+(yb)^2 \leqslant x^2a^2+x^2b^2+y^2a^2+y^2b^2 && \text{(squaring)}\\
\iff\ & 2axyb \leqslant x^2b^2+y^2a^2 && \text{(cancelling)}\\
\iff\ & 0 \leqslant x^2b^2-2axyb+y^2a^2 && \text{(rearranging)}\\
\iff\ & 0 \leqslant (xb-ya)^2 && \text{(rearranging)}
\end{aligned}
$$

This last line is clearly a true statement, and since all operations were reversible, the first line is also true.

Exercise 20.3.

Lecture 21: Fourier topics

Exercise 21.1.

Exercise 21.2.

Lecture 22: Graphs

Exercise 22.1. 1. A sink will have all zeros in its row (no outgoing edges) and a source will have all zeros in its columns (no incoming edges).

2. If all the entries of A were 1, except the diagonal (because that would imply self loops, which we forbid in simple graphs), G would have $|V|(|V|-1)$ edges.

Exercise 22.2.

Exercise 22.3.

Lecture 23: Markov matrices and spectral clustering

Exercise 23.1.

Exercise 23.2.

Lecture 24: Graph clustering

Exercise 24.1.

Index