Introduction to Linear Algebra

Jānis Lazovskis

Updated July 19, 2023

Preface

These notes were created to accompany the course *Introduction to Linear Algebra* for the BITL program at RTU Riga Business School. They have been used in the Fall 2021, Spring 2022, and Spring 2023 semesters. The text may contain mistakes - please send any you find to `janis.lazovskis@rbs.lv`.

This course broadly follows 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.

Throughout the text, there are highlighted **Definitions** in green, **Inquiries** in blue, and **Algorithms** in red. The definitions are meant as key points that should be understood, if nothing else. The inquiries are meant as guiding questions to connect and unify ideas. The algorithms are meant as step-by-step instructions for complicated ideas.

At the end of each lecture there are exercises, with some solutions provided at the end of the text. Exercises which require the use of a computer are marked with the symbol ⋈ .

Every inquiry and exercise is marked with a symbol ✠ and a number next to it, indicating the standard that it refers to. A full list of standards is given further below.

Contents

Standards

This list of standards, each indicated by ✠ in the text, is a collection of the standards indicated at the beginning of each lecture. In case of conflict between this list and the list kept on the course website, the list on the course website will be taken to be the correct one.

Part	Number	Standard
I. Matrices and row reduction	1.01	Add vectors, multiply them by scalars, take their dot products
	1.02	Compute the angle between vectors
	1.03	Add, multiply, and transpose matrices (including block matrices)
	1.04	Multiply vectors with matrices
	1.05	Construct vectors and matrices in Python and perform operations with them
	1.06	Draw lines in the plane representing $m \times 2$ matrix equations (rows and columns)
	1.07	Understand row operations as matrix multiplication
	1.08	Bring a matrix to (reduced) row echelon form using Gaussian elimination
	1.09	Construct the inverse of a matrix A by eliminating the block matrix $[A\ I]$
	1.10	Decompose matrices A and PA as the products LU and LDU
	1.11	Identify when the matrix equation $A\mathbf{x} = \mathbf{b}$ has no solutions or has infinitely many solutions
	1.12	Construct the column space and nullspace of a matrix as spans
	1.13	Describe solutions to $A\mathbf{x} = \mathbf{b}$ using the language of vector spaces
	1.14	Construct the particular, special, and complete solutions to $A\mathbf{x} = \mathbf{b}$, for any $m \times n$ matrix A
	1.15	Identify the row rank, column rank, and rank of a matrix
II. Vector spaces	2.01	Determine if something is a vector space or a subspace
	2.02	Describe a vector space as a span of vectors
	2.03	Identify linearly independent subsets in a given set of vectors
	2.04	Express the same vector in different bases
	2.05	Find a basis and the dimension of a vector space
	2.06	Find the intersection of two planes
	2.07	Describe a hyperplane as a span of vectors
	2.08	Find the bases of the four fundamental subspaces of a matrix
	2.09	Determine if the columns of a matrix are orthogonal
	2.10	Determine if two subspaces are orthogonal
	2.11	Compute the projection of a vector onto another vector
	2.12	Compute the projection of a vector onto a subspace
	2.13	Find the least squares solution to a matrix equation
	2.14	Find the degree-d polynomial that approximates a collection of points in $\mathbf{R}^2$

	2.15	Apply the Gram–Schmidt process to a set of vectors
	2.16	Extend a set of linearly independent vectors to a basis
	2.17	Determine if something is or is not an inner product space
	2.18	Compute the length, angle, projections of vectors in arbitrary inner product spaces
	2.19	Determine whether or not a function is a linear transormation
	2.20	Construct a matrix for a linear transformation, knowing what it does to a basis
	2.21	Construct the image and kernel of a linear transformation, as vector spaces
III. Eigentheory	3.01	Compute the determinant using both the recursive and combinatorial definitions
	3.02	Use the multilinearity and alternating properties to infer results for special matrices
	3.03	Compute determinants of products, inverses, transposes
	3.04	Prove simple properties of the determinant
	3.05	Find eigenvalues and eigenvectors of a matrix
	3.06	Given only eigenvalues and eigenvectors of A, compute $A\mathbf{x}$ for any $\mathbf{x}$
	3.07	Given only eigenvalues and eigenvectors, construct a matrix having them
	3.08	Compute trace, determinant, eigensystems of matrices
	3.09	Diagonalize a matrix with lin. indep. eigenvectors, and identify when it is not possible
	3.10	Given a matrix A, construct and identify matrices similar to A
	3.11	Identify symmetric and positive definite matrices, directly and indirectly
	3.12	Express a symmetric matrix as a sum of rank one matrices
	3.13	Compute the rank r approximation to a matrix A
	3.14	Decompose a non-square matrix by the SVD
	3.15	Normalize and center a matrix of n samples on its mean
	3.16	Identify the principal components of A, in terms of the total covariance of A
	3.17	Solve the perpendicular least squares problem using SVD
IV. Extensions	4.01	Express a complex number in one of four different ways
	4.02	Translate known properties of vectors and matrices to Hermitian vectors and matrices
	4.03	Construct the four matrices associated to a graph
	4.04	Find a spanning tree of a graph using row reduction on the incidence matrix

Part I

Matrices and row reduction

Lecture 1: Vectors

Chapters 1.1 and 1.2 in Strang's "Linear Algebra"

- Fact 1: The dot product of a vector with itself is the square of its length.
- Fact 2: A plane in $\mathbf{R}^3$ is defined by an equation in x, y, z.

✠ Standard 1.01: Add vectors, multiply them by scalars, take their dot products.

✠ Standard 1.02: Compute the angle between vectors.

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 The algebra of vectors

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

The superscript "n" in $\mathbf{R}^n$ is described in full in detail in Lecture 10 (it is the "dimension" of the "space"), but for now can be taken to be simply the number of elements in the vector. When $n = 1$, the number is usually omitted, as in Definition 1.1. The elements, or *components*, of a vector $\mathbf{v}$ are denoted by a subscript, that is,

$$\mathbf{v} = (v_1, v_2, \ldots, v_n) \qquad \text{and} \qquad \mathbf{v}_i = (v_1, v_2, \ldots, v_n)_i = v_i,$$

for every $i = 1, 2, \ldots, n$.

Example 1.2. A vector $\mathbf{v}$ with three elements 3, and -1, and $\sqrt{2}$, is denoted equivalently as

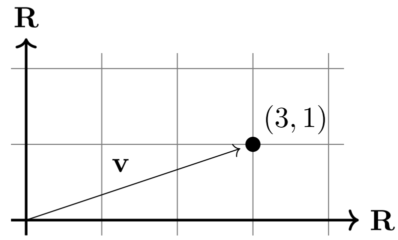
$$\begin{bmatrix} 3 \\ -1 \\ \sqrt{2} \end{bmatrix} \quad \text{or} \quad \left[\begin{smallmatrix} 3 \\ -1 \\ \sqrt{2} \end{smallmatrix}\right] \quad \text{or} \quad \begin{bmatrix} 3 & -1 & \sqrt{2} \end{bmatrix}^T \quad \text{or} \quad \begin{pmatrix} 3 \\ -1 \\ \sqrt{2} \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 3 & -1 & \sqrt{2} \end{pmatrix}^T \quad \text{or} \quad (3, -1, \sqrt{2}).$$

In this class, the comma "," is used to separate elements in a list, and the point "." is used as a decimal separator.

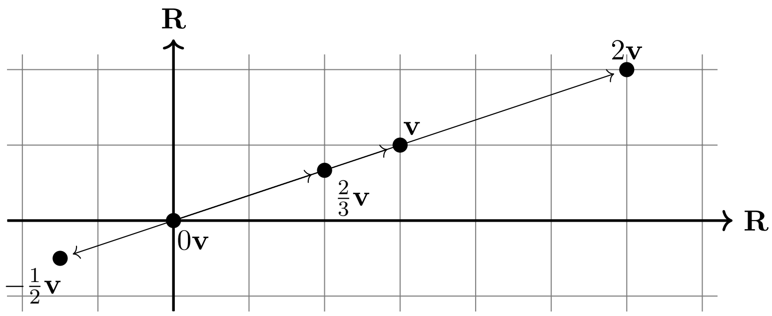
The *zero vector*, or a *trivial vector*, denoted 0, is vector for which all elements are 0. Vectors that are not the zero vector are called *nontrivial*. A vector is usually thought of as a column of numbers, or a point in n-dimensional space, or the arrow to that point. All notions of a vector will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

Definition 1.3: Let $c \in \mathbf{R}$ be a real number and $\mathbf{v} \in \mathbf{R}^n$ be a vector. The *product* or *scalar product* $c\mathbf{v}$ is a vector in R^n with $(c\mathbf{v})_i = c \cdot v_i$.

Example 1.4. 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.

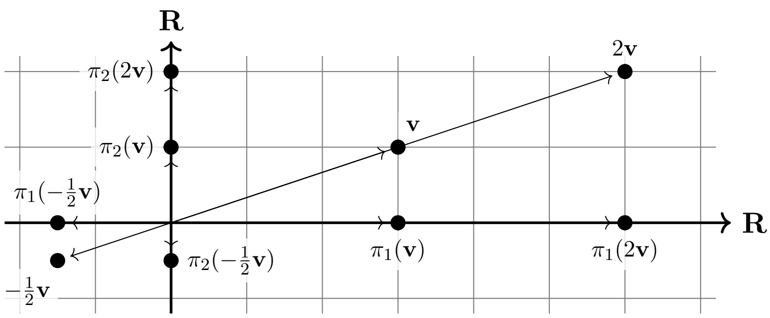


Multiplying the vector by elements of $\mathbf{R}$ we get other vectors "going in the same direction" as $\mathbf{v}$.



> **Definition 1.5:** The numbers $v_1, \ldots, v_n \in \mathbf{R}$ in the vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbf{R}^n$ are called the *components* of the vector $\mathbf{v}$. For each component v_i, there is a unique function $\pi_i \colon \mathbf{R}^n \to \mathbf{R}$ called the *projection*, with $\pi_i(\mathbf{v}) = v_i$.

Example 1.6. Projection and multiplication by a number can be rearranged.



That is, $\pi_i(c\mathbf{v}) = c \cdot \pi_i(\mathbf{v})$ for all real numbers c and indices i. We will consider projections in more detail in Lecture 13.

Vectors are combined together with each other and with numbers in *linear combinations*.

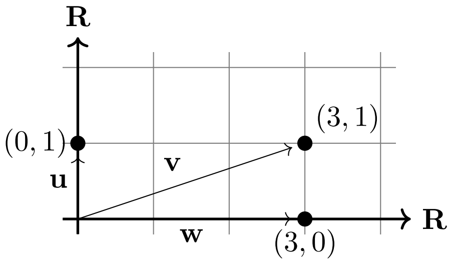
> **Definition 1.7:** A *linear combination* of vectors is a vector $\mathbf{v} \in \mathbf{R}^n$ when it is expressed as a sum of other vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k \in \mathbf{R}^n$, and *scalars* $a_1, a_2, \ldots, a_k \in \mathbf{R}$ multiplying them. That is,
>
> $$\mathbf{v} = a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_k\mathbf{w}_k.$$
>
> When $k = 1$, the linear combination of one vector $a_1\mathbf{w}_1$ is called a *multiple* of the vector $\mathbf{w}_1$.

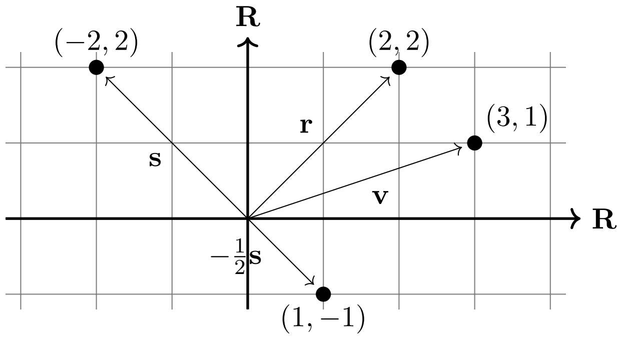
Multiplying vectors by a number is distributive over vector addition, as demonstrated in Inquiry 1.11. That is, for any $c \in \mathbf{R}$ and $\mathbf{v}, \mathbf{w} \in \mathbf{R}^n$, we have

$$c(\mathbf{v} + \mathbf{w}) = c\mathbf{v} + c\mathbf{w}.$$

Example 1.8. Every vector in the plane is a linear combination of (at most) two vectors, representing the x-direction and y-direction.

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





The entries of vectors, and the numbers multiplying them, do not need to be numbers - they simply need to be elements of a *field*, denoted $\mathbf{F}$ in general. Unless otherwise noted, we will always use the field $\mathbf{R}$.

Example 1.9. 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).

Definition 1.10: The *dot product*, or *inner product* of two vectors $\mathbf{v} = (v_1, \dots, v_n)$ and $\mathbf{w} = (w_1, \dots, w_n) \in \mathbf{R}^n$ is the real number $\mathbf{v} \bullet \mathbf{w} := v_1 w_1 + \cdots + v_n w_n \in \mathbf{R}$.

In other words, the dot product is a function $\mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$.

Inquiry 1.11 (✠1.01): Consider the vectors $\mathbf{v} = (1, 3, -1)$ and $\mathbf{w} = (2, 2, 0)$ in $\mathbf{R}^3$.

1. Compute the dot products $\mathbf{v} \bullet \mathbf{w}$, $\mathbf{v} \bullet (2\mathbf{w})$, and $\mathbf{v} \bullet (3\mathbf{w})$. What will be $\mathbf{v} \bullet (c\mathbf{w})$, for any real number c?
2. Compute the projections $\pi_i(\mathbf{v} + \mathbf{w})$ for $i = 1, 2, 3$. Do there exist vectors $\mathbf{x}, \mathbf{y}$ with $\pi_i(\mathbf{x} + \mathbf{y}) \neq \pi_i(\mathbf{x}) + \pi_i(\mathbf{y})$?
3. Give an alternative definition of the dot product using the projection maps π_i.

1.2 The geometry of vectors

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.

Definition 1.12: A *plane* in $\mathbf{R}^3$ is all the points $(x, y, z) \in \mathbf{R}^3$ that satisfy an equation $ax + by + cz = d$, for some $a, b, c, d \in \mathbf{R}$. A *line* in $\mathbf{R}^3$ is all the points in $\mathbf{R}^3$ that are in two different planes that intersect.

We are often interested in planes that go through the origin $(0, 0, 0)$. They have $d = 0$ for their defining equation.

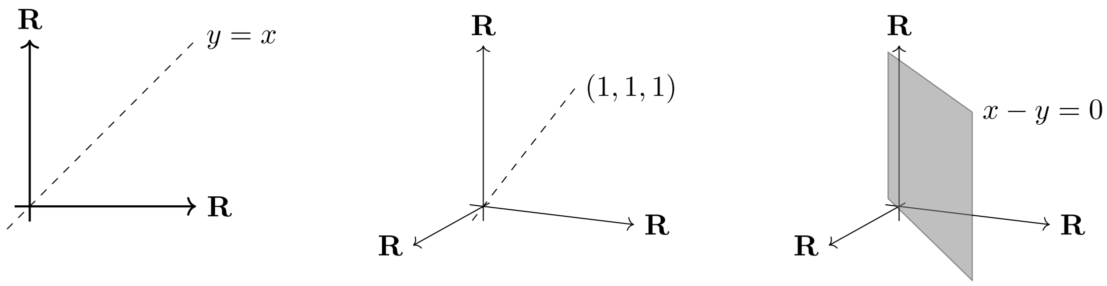
Example 1.13. Linear combinations can be described geometrically. For example:

- 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 $\mathbf{R}^3$
- 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,

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

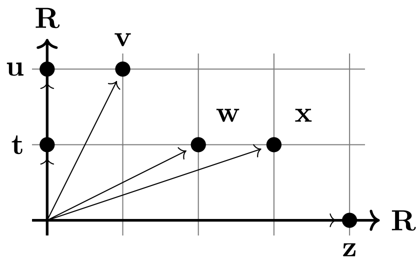
- Linear combinations of $(1,1,1)$, $(1,1,0)$, $(0,1,1)$, and $(1,0,1)$ still fill all of $\mathbf{R}^3$. For example,

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



> **Inquiry 1.14 (✠1.01):** Consider the vector $\mathbf{v} = (1,1,1)$ and the plane P defined by $x+y+z = 3$.
>
> 1. It is clear that the plane defined by $2x+2y+2z = 6$ is the same as P. In general, given two planes defined by $a_1x+b_1y+c_1z = d_1$ and $a_2x+b_2y+c_2z = d_2$, how can you tell just from these equations that they are "different"? Use the variables $a_i, \ldots, d_i$ when explaining.
> 2. The point $\mathbf{v}$ lies on the plane P but $0 = (0,0,0)$ does not. Can you find two different planes that contain both $(1,1,1)$ and $(0,0,0)$?

Example 1.15. Consider the following vectors in $\mathbf{R}^2$.



Given these five different vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x}, \mathbf{z}, \mathbf{t}$, there are several relationships among them:

$$\mathbf{v} = \mathbf{u} + \frac{1}{4}\mathbf{z}, \qquad \mathbf{z} = 2\mathbf{w} - \mathbf{u}, \qquad \mathbf{z} + \mathbf{v} = \mathbf{w} + \mathbf{x}.$$

These are not the only ones - there are many more. The given relationships become clearer when the vectors are split into their x- and y-components:

$$\mathbf{u} = 2\begin{pmatrix}0\\1\end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix}1\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\end{pmatrix}, \quad \mathbf{w} = 2\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}, \quad \mathbf{x} = 3\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}, \quad \mathbf{z} = 4\begin{pmatrix}1\\0\end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix}0\\1\end{pmatrix}.$$

Inquiry 1.16 (✠1.01): In Example 1.15, the relationships given were among three or four vectors.

1. Can any two of the five given vectors be related by an equation? Try to find a counterexample.
2. Explain how the equation $\mathbf{z} + \mathbf{v} = \mathbf{w} + \mathbf{x}$ is actually the sum of two "smaller" equations, with three vectors each.
3. Can any three of the five given vectors be related by an equation? Use the x- and y-decompositions to help you out.

Definition 1.17: The dot product of a vector $\mathbf{v}$ with itself is the square of the *norm*, or *length*, or *distance* of the vector $\mathbf{v}$, denoted $\|\mathbf{v}\|$. That is,

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

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

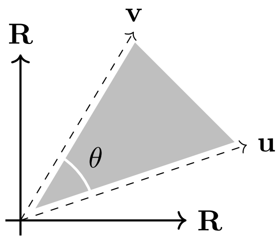
Proposition 1.19. For any $\mathbf{u}, \mathbf{v}$ nonzero in $\mathbf{R}^n$:

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 $\frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \cos(\theta)$
3. The *Cauchy–Schwarz inequality* holds: $|\mathbf{u} \bullet \mathbf{v}| \leqslant \|\mathbf{u}\|\|\mathbf{v}\|$
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

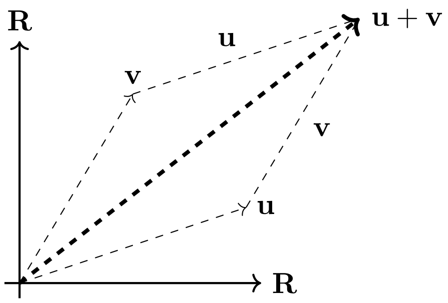
$$\left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\|^2 = \frac{\mathbf{v}}{\|\mathbf{v}\|} \bullet \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{1}{\|\mathbf{v}\|^2}(\mathbf{v} \bullet \mathbf{v}) = \frac{1}{\|\mathbf{v}\|^2}\|\mathbf{v}\|^2 = 1.$$

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



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

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.



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

□

As a result of part 2. of the proof above, 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 $\mathbf{u}\bullet\mathbf{v}=0$.

Definition 1.20: 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 $\mathbf{v}\bullet\mathbf{w}=0$, 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 12.

Inquiry 1.21 (✠1.02): This inquiry uses vectors in $\mathbf{R}^n$ for several different n.

1. Find values of c such that the vectors $(1,1,c-2)$ and $(1,c-2,1)$ in $\mathbf{R}^3$ make an angle of $\pi/3$ with each other.
2. Two vectors $\mathbf{u}(t)=(2\cos(t),2\sin(t)),\mathbf{v}(t)=(\cos(t),\sin(-t))$ are moving in $\mathbf{R}^2$, as $t\in\mathbf{R}$ changes. What is the angle between them, as a function of t?
3. Let $\mathbf{u},\mathbf{v}\in\mathbf{R}^n$. Use Proposition 1.19 to show that the angle between $\mathbf{u}$ and $\mathbf{v}$ is the same as the angle between $\mathbf{u}$ and $2\mathbf{v}$. Can you generalize this to any scalar multiples of $\mathbf{u}$ and $\mathbf{v}$?

1.3 Exercises

Exercise 1.1. (✠1.01) Consider the four vectors $\mathbf{v}=\begin{bmatrix}0\\6\\-1\end{bmatrix}$, $\mathbf{w}=\begin{bmatrix}-3\\-4\\-5\end{bmatrix}$, $\mathbf{z}=\begin{bmatrix}0\\0\\1\end{bmatrix}$, $\mathbf{y}=\begin{bmatrix}-5\\5\\-4\end{bmatrix}$. Find $a,b,c\in\mathbf{R}$ with $a\mathbf{v}+b\mathbf{w}+c\mathbf{z}=\mathbf{y}$.

Exercise 1.2. (✠1.01) Check that the dot product from Definition 1.1 is *distributive* over vector addition. That is, show that $\mathbf{v}\bullet(\mathbf{u}+\mathbf{w})=\mathbf{v}\bullet\mathbf{u}+\mathbf{v}\bullet\mathbf{w}$, for any $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbf{R}^n$.

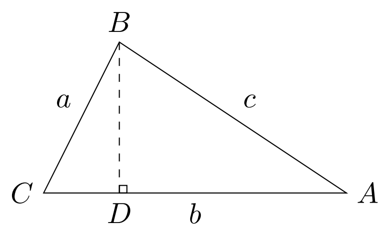
Exercise 1.3. (✠1.01) Let $\mathbf{v} = (1,1,1)$, $\mathbf{w} = (2,-2,0)$ and $\mathbf{z} = (-3,1,2)$ be vectors in $\mathbf{R}^3$.

1. Using a linear equation in three variables, describe the plane of points $\mathbf{R}^3$ that are equidistant from $\mathbf{v}$ and $\mathbf{w}$.
2. Using two equations, describe the line of points in $\mathbf{R}^3$ that are equidistant from $\mathbf{v}$, $\mathbf{w}$, $\mathbf{z}$. Hint: A line is the intersection of two planes.

Exercise 1.4. (✠1.01) Let S be the subset $[-5,5] \times [-3,3] \subseteq \mathbf{R}^2$.

1. Identify all the points in S that correspond to linear combinations $a\left[\begin{smallmatrix}3\\0\end{smallmatrix}\right] + b\left[\begin{smallmatrix}-1\\2\end{smallmatrix}\right]$, for $a, b \in \mathbf{Z}$.
2. Which of the points from part (a) lie a distance of more than 2 but less than 3 from the origin?

Exercise 1.5. The proof of Proposition 1.19 used the "law of cosines", which itself was not proved, so we prove it here. Consider the triangle below:



1. Find the formulas for $\cos(C)$ and $\sin(C)$ in the triangle BCD.
2. Rewrite $\cos(C)$ from above so it has the number $b = AC$. Use the fact that $CD = AC - b$.
3. Express the Pythagorean theorem of triangle ABD.
4. Replace the sides from part (c) with the formular from parts (a) and (b). Simplify to get the law of cosines.

Exercise 1.6. (✠1.01, 1.02) Let $\mathbf{v} \in \mathbf{R}^3$ be non-trivial, and let $\mathbf{w}, \mathbf{z} \in \mathbf{R}^3$ be non-trivial vectors both perpendicular to $\mathbf{v}$. Show that the halfway point between $\mathbf{w}$ and $\mathbf{z}$ is also perpendicular to $\mathbf{v}$.

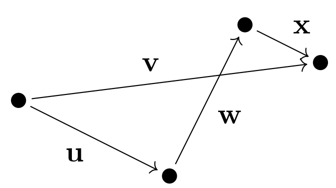
Exercise 1.7. (✠1.02) This question is about orthogonality of vectors in Euclidean space $\mathbf{R}^n$.

1. Find $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^3$ nonzero for which $\mathbf{u}$ is perpendicular to $\mathbf{v}$, $\mathbf{v}$ is perpendicular to $\mathbf{w}$, and $\mathbf{u}$ is perpendicular to $\mathbf{w}$.
2. Find $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^3$ nonzero for which $\mathbf{u}$ is perpendicular to $\mathbf{v}$, $\mathbf{v}$ is perpendicular to $\mathbf{w}$, and $\mathbf{u}$ is colinear to $\mathbf{w}$.
3. **Bonus:** Explain why it is not possible to have $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x} \in \mathbf{R}^3$ nonzero with every pair of vectors orthogonal to each other.

Exercise 1.8. (✠1.01) This question is about the Cauchy–Schwarz inequality, $|\mathbf{v} \cdot \mathbf{w}| \leqslant \|\mathbf{v}\| \cdot \|\mathbf{w}\|$.

1. Suppose that there exists $c \in \mathbf{R} \setminus \{0\}$ with $\mathbf{w} = c \cdot \mathbf{v}$. Show that the Cauchy–Schwarz inerquality holds with equality.
2. Suppose that the Cauchy–Schwarz inequality holds with equality. Show that there exists $c \in \mathbf{R} \setminus \{0\}$ with $\mathbf{w} = c \cdot \mathbf{v}$.

Exercise 1.9. (✠1.01) Use the triangle inequality to show that vector $\mathbf{v}$ is shorter than the sum of the lengths of the vectors $\mathbf{u}, \mathbf{w}, \mathbf{x}$. That is, show with the triangle inequality that $\|\mathbf{v}\| \leqslant \|\mathbf{u}\| + \|\mathbf{w}\| + \|\mathbf{x}\|$.



Lecture 2: Matrices

Chapter 1.3 in Strang's "Linear Algebra"

- Fact 1: Matrix multiplication is associative and distributive, but not commutative.
- Fact 2: Not every matrix has an inverse.
- Fact 3: Conclusions may be made about matrices without knowing all their entries.

✠ Standard 1.03: Add, multiply, and transpose matrices (including block matrices)

✠ Standard 1.04: Multiply vectors with matrices

✠ Standard 1.05: Construct vectors and matrices in Python and perform operations with them

---

2.1 Types of matrices

Definition 2.1: Let $m, n \in \mathbf{N}$. An $m \times n$ *matrix* over $\mathbf{R}$ is an ordered set M 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.
- The size, or dimensions of a matrix, is the pair (m, n). By convention, the number of rows comes first.

The elements of a matrix are called its *entries*. The entry in row i, column j is called the ij-entry.

Comparing Definition 2.1 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 $\mathbf{R}$. Two matrices of particular importance are the *zero matrix* 0 (all entries are zero) and the *identity matrix* I (all entries are zero except the diagonal, which is all 1's), given by

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

The identity matrix is square, but the zero matrix does not have to be square. 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 given specfic numbers for constructing a matrix, you are given other matrices.

Definition 2.2: A matrix $M \in \mathcal{M}_{m\times n}$ is a *block matrix* if its entries are matrices instead of numbers

Example 2.3. 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. As before, the matrix I will always be square, but 0 can be any shape.

Finally, there are three special types of matrices (not necessarily square, though most often they are), called *triangular 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 $i \neq j$

These matrices are drawn square, but they do not have to be square (though if not mentioned, thy are assumed to be square). The symbol "$*$" represents any number, and all the "$*$" entries do not have to be the same. They could even all be zero.

Inquiry 2.4 (✠1.03): Let X, Y, Z, W be matrices, and consider the block matrix $A = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix}$.

1. Suppose you know that $X \in \mathcal{M}_{2\times 4}$ and $W \in \mathcal{M}_{3\times 1}$. How many rows and columns do each of Y, Z have?
2. If A is the identity matrix I, write out Z explicitly.
3. Suppose that each of X, Y, Z, W are upper triangular. Is A upper triangular? How many nonzero entries can A have?

2.2 Operations on matrices

Definition 2.5: 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 $(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}$

Example 2.6. Since vectors are special matrices (and numbers are special vectors), these operations work for multiplying matrices, vectors, and numbers. For example,

$$\underbrace{\begin{bmatrix} 2 & 0 & -2 \\ 4 & -1 & 3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1 & -3 \\ 2 & 2 \\ 0 & -1 \end{bmatrix}}_{B} = \begin{bmatrix} \sum_{j=1}^3 A_{1j}B_{j1} \\ \sum_{j=1}^3 A_{2j}B_{j2} \end{bmatrix} = \begin{bmatrix} 2\cdot 1 + 0 \cdot 2 + (-2)\cdot 0 & 2\cdot(-3) + 0\cdot 2 + (-2)\cdot(-1) \\ 4\cdot 1 + (-1)\cdot 2 + 3\cdot 0 & 4\cdot(-3) + (-1)\cdot 2 + 3\cdot(-1) \end{bmatrix} = \begin{bmatrix} 2 & -4 \\ 2 & -17 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 2 & 3 & -1 \\ 8 & -2 & 0 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix}}_{\mathbf{x}} = \begin{bmatrix} \sum_{j=1}^2 C_{1j}x_1 \\ \sum_{j=1}^2 C_{2j}x_2 \end{bmatrix} = \begin{bmatrix} 2\cdot 3 + 3\cdot 1 + (-1)\cdot(-2) \\ 8\cdot 3 + (-2)\cdot 1 + 0\cdot(-2) \end{bmatrix} = \begin{bmatrix} 11 \\ 22 \end{bmatrix}$$

$$\underbrace{7}_{D} \underbrace{\begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix}}_{E} = \begin{bmatrix} 7\cdot 2 & 7\cdot(-1) \\ 7\cdot 3 & 7\cdot(-2) \end{bmatrix} = \begin{bmatrix} 14 & -7 \\ 21 & -14 \end{bmatrix}.$$

Remark 2.7. Matrix addition has the following properties, for A, B, C are matrices of the appropriate size, $c \in \mathbf{R}$, and $\mathbf{x}$ a vector:

- 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 by a matrix is *distributive* over addition: $C(A+B) = CA+CB$ and $(A+B)C = AC + BC$
- multiplication by a matrix or vector is *assocative*: $A(BC) = (AB)C$ and $A(B\mathbf{x}) = (AB)\mathbf{x}$

Multiplication of matrices is not always *commutative*: $AB \neq BA$.

Example 2.8. 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.

Inquiry 2.9 (✠1.03, 1.05): This inquiry is about the observations from Example 2.8.

1. Suppose that there is a 2×2 matrix J for which $AJ = A$ for every 2×2 matrix A. Show that J must be the identity matrix!
   Hint: write out the elements of J explicitly, and use special matrices A to create four equations from $AJ = A$.
2. In Python, construct a random 5×3 matrix A using the `np.random.randint` function and the identity matrix using the `np.eye` function. Using the function `np.matmul`, compute the products IA and AI. What do you get? Is I the same both times?

Remark 2.10. When multiplying block matrices, extra care has to be taken with non-commutativity. For example, if A, B, C are matrices, then

$$\begin{bmatrix} A & I \\ B & C \end{bmatrix} \begin{bmatrix} I & C \\ D & D \end{bmatrix} = \begin{bmatrix} A + D & AD + D \\ B + CD & BC + CD \end{bmatrix}.$$

The lower right entry cannot be simplified as $C(B+D)$, because it is not always true that $BC = CB$.

Definition 2.11: Let A be an $m \times n$ matrix. The *transpose* of A is written A^T, and has ij-entry $(A^T)_{ij} = A_{ji}$.

The transpose plays well with matrix operations:

$$(A+B)^T = A^T + B^T,$$
$$(A\mathbf{x})^T = \mathbf{x}^T A^T,$$
$$(AB)^T = B^T A^T.$$

These results follow from how the sum and product were defined in Definition 2.5.

Definition 2.12: 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 2.8, the inverse of the identity matrix is itself: $II = I$, so $I^{-1} = I$.

Example 2.13. If $A \in \mathcal{M}_{n\times n}$ is a diagonal matrix with nonzero entries on its diagonal, then its inverse is the same, but with reciprocals on the diagonal:

$$A = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}, \qquad A^{-1} = \begin{bmatrix} \frac{1}{a_{11}} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{a_{22}} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{a_{33}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{a_{nn}} \end{bmatrix}.$$

Inquiry 2.14 (✠1.03): Consider the diagonal matrix and its inverse from Example 2.13.

1. If $A = \left[\begin{smallmatrix} 3 & d \\ 0 & -2 \end{smallmatrix}\right]$, compute the product $\left[\begin{smallmatrix} 3 & d \\ 0 & -2 \end{smallmatrix}\right] \left[\begin{smallmatrix} \frac{1}{3} & 0 \\ 0 & \frac{-1}{2} \end{smallmatrix}\right]$. This is not quite the identity matrix I. What should change for this product to be the identity? This, what is A^{-1}?
2. Let $B = \left[\begin{smallmatrix} 3 & 0 & 0 \\ 0 & -5 & d \\ 0 & 0 & -1 \end{smallmatrix}\right]$, with $d \neq 0$. Find B^{-1}.
   Hint: $(B^{-1})_{ij} = 0$ *precisely when* $B_{ij} = 0$.
3. Let $C = \left[\begin{smallmatrix} 3 & 0 & 0 & 0 \\ 0 & -5 & d & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 3 \end{smallmatrix}\right]$ with $d \neq 0$. Find C^{-1}.
4. Generalize the above example with $C_{ij} = d \neq 0$ instead of C_{23}, with the condition that $i < j$ (that is, C_{ij} is above the diagonal). What if C_{ij} is below the diagonal?

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 2.15. 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 $\mathbf{x} = (x_1, x_2, x_3, x_4)$:

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

Definition 2.16: Let $A \in \mathcal{M}_{m\times n}$ be a matrix, and $\mathbf{x} \in \mathbf{R}^n$, $\mathbf{b} \in \mathbf{R}^m$ be vectors. The equation $A\mathbf{x} = \mathbf{b}$ is a *matrix equation*, and consists of m individual equations:

$$A\mathbf{x} = \mathbf{b} \quad \Longleftrightarrow \quad \begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n & = & b_2 \\ & \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n & = & b_m. \end{array}$$

Finding the inverse of a matrix A is related to finding the solution $\mathbf{x}$ to a matrix equation $A\mathbf{x} = \mathbf{b}$. Indeed, if A has an inverse, then we immediately see that

$$A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b} \quad \Longleftrightarrow \quad A^{-1}A\mathbf{x} = A^{-1}\mathbf{b} \quad \Longleftrightarrow \quad I\mathbf{x} = A^{-1}\mathbf{b} \quad \Longleftrightarrow \quad \mathbf{x} = A^{-1}\mathbf{b}.$$

Inquiry 2.17 (✠1.04): Let A be a matrix.

1. Suppose you know that $A\left[\begin{smallmatrix}2\\4\end{smallmatrix}\right] = \left[\begin{smallmatrix}2\\0\end{smallmatrix}\right]$ and $A\left[\begin{smallmatrix}-1\\-4\end{smallmatrix}\right] = \left[\begin{smallmatrix}0\\2\end{smallmatrix}\right]$. What is the inverse matrix A^{-1}? Hint: If $A\mathbf{x} = \mathbf{b}$, then putting $\mathbf{x}$ as the first column of a 2×2 matrix $[\,\mathbf{x}\ *\,]$, we get $A[\,\mathbf{x}\ *\,] = [\,\mathbf{b}\ *\,]$.

2. In general for $A \in \mathcal{M}_{m\times n}$, suppose that for any vector $\mathbf{b} \in \mathbf{R}^m$, you are able to find $\mathbf{x} \in \mathbf{R}^n$, which depends on $\mathbf{b}$, such that $A\mathbf{x} = \mathbf{b}$. Explain which vectors $\mathbf{b}$ you would choose to construct the inverse of the matrix A.

3. Is the collection of vectors $\mathbf{b}$ from the previous part unique? Is there a minimum number of vectors? Give two different collections of vectors $\mathbf{b}$ that would work.

Example 2.18. The *cyclic matrix* C does not have an inverse. That is, there is no vector $\mathbf{x}$ for which

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

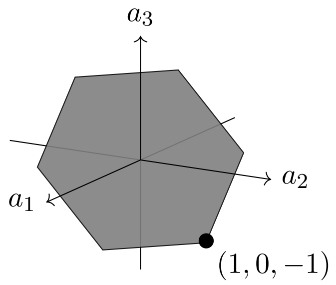
for any chosen $\mathbf{a}$. It is immediate that $\mathbf{a} = 0$ has a solution, when $x_1 = x_2 = x_3$. But it is also immediate that $\mathbf{a} = (1, 2, 3)$ is not a solution, because adding the three equations

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

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:



2.3 Vectors and matrices in Python

The operations you have seen so far can be replicated in Python using the *NumPy* packcage (which may be taken to stand for "Numerical Python"). To begin, the *NumPy* package is loaded by executing the following code:

```
import numpy as np
```

To input matrices and multiply them, such as in Exmaple 2.6, the following three lines are executed, which produces the given result:

```
A = np.array([[2,0,-2],[4,-1,3]])
B = np.array([[1,-3],[2,2],[0,-1]])
np.matmul(A,B)

array([[  2,  -4],
       [  2, -17]])
```

Matrices full of ones or zeros can be created with the `ones` or `zeros` command:

```
np.ones((3,6),dtype=int)

array([[1, 1, 1, 1, 1, 1],
       [1, 1, 1, 1, 1, 1],
       [1, 1, 1, 1, 1, 1]])
```

Generating "random" matrices with integer entries is straightforawrd, as is taking their upper (or lower triangular parts), with `triu` (or `tril`, respectively).

```
A = np.random.randint(low=-9,high=10,size=(4,7))
np.triu(A)

array([[-2,  3,  0, -1,  3, -5,  9],
       [ 0, -5,  2, -1, -4, -8, -1],
       [ 0,  0,  4, -6,  9,  6,  4],
       [ 0,  0,  0, -7,  8, -3, -5]])
```

2.4 Exercises

Exercise 2.1. (✠1.03) 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 2.2. (✠1.03) Let A, B, C, D be $n \times n$ matrices that are invertible. Find the inverses of the following block matrices.

1. $\begin{bmatrix} I & 0 \\ 0 & D \end{bmatrix}$
2. $\begin{bmatrix} I & B \\ 0 & D \end{bmatrix}$
3. $\begin{bmatrix} A & 0 \\ I & D \end{bmatrix}$

Exercise 2.3. (✠1.03) 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 2.4. (✠1.03) This question is about *triangular* matrices.

1. Show that the product of two lower triangular matrices is lower triangular.

2. Show that the product of two upper triangular 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?

Exercise 2.5. (✠1.04) Let $A \in \mathcal{M}_{2\times 3}$ and $\mathbf{v} = \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right] \in \mathbf{R}^2, \mathbf{w} = \left[\begin{smallmatrix} -1 \\ 1 \\ 2 \end{smallmatrix}\right] \in \mathbf{R}^3$.

1. Suppose you know that $A\mathbf{w} = \left[\begin{smallmatrix} -2 \\ 3 \end{smallmatrix}\right]$ and $\mathbf{v}^T A = \left[\begin{smallmatrix} -1 & 1 & 3 \end{smallmatrix}\right]$. What could the entries of A be to satisfy these relations?

2. Only referencing the sizes of $A, \mathbf{v}, \mathbf{w}$ (not the numbers within them), explain why A cannot be uniquely determined just by knowing $A\mathbf{w}$ and $\mathbf{v}^T A$.

Exercise 2.6. (✠1.05) For each part of this question, construct a Python function with the given name.

1. Make a function `ones_counter(matrix)` that takes in a matrix, in the form of a `numpy` array, and returns the number of entries that are `1`.

2. Make a function `thick_diagonal(rownum, colnum)` that takes in two positive integers and returns a matrix, in the form of a `numpy` array, having `rownum` rows and `colnum` columns, and zero everywhere except on the diagonal and just above and just below it. For example, `thick_diagonal(5, 10)` should return the following matrix:

```
array([[1, 1, 0, 0, 0, 0, 0, 0, 0, 0],
       [1, 1, 1, 0, 0, 0, 0, 0, 0, 0],
       [0, 1, 1, 1, 0, 0, 0, 0, 0, 0],
       [0, 0, 1, 1, 1, 0, 0, 0, 0, 0],
       [0, 0, 0, 1, 1, 1, 0, 0, 0, 0]])
```

Lecture 3: Elimination

Chapters 2.1, 2.2 in Strang's "Linear Algebra"

- Fact 1: Row operations are matrix multiplications.
- Fact 2: Solving a matrix equation can be understood in terms of the rows or the columns.

✠ Standard 1.06: Draw lines in the plane representing $m \times 2$ matrix equations (rows and columns).

✠ Standard 1.07: Understand row operations as matrix multiplication.

This lecture reviews how to solve linear systems, and goes into more detail. Recall the three *elementary row operations*, which will be here presented as matrix multiplication:

multiply a row by a nonzero number:
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 7 & 3 & 4 \\ 0 & 2 & -1 & 3 \\ -1 & 2 & 5 & 2 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 7 & 3 & 4 \\ 0 & 4 & -2 & 6 \\ -1 & 2 & 5 & 2 \end{array}\right]$$

swap two rows:
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 7 & 3 & 4 \\ 0 & 2 & -1 & 3 \\ -1 & 2 & 5 & 2 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 7 & 3 & 4 \\ -1 & 2 & 5 & 2 \\ 0 & 2 & -1 & 3 \end{array}\right]$$

add a multiple of one row to another row:
$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 7 & 3 & 4 \\ 0 & 2 & -1 & 3 \\ -1 & 2 & 5 & 2 \end{array}\right] = \left[\begin{array}{ccc|c} 0 & 9 & 8 & 6 \\ 0 & 4 & -2 & 6 \\ -1 & 2 & 5 & 2 \end{array}\right]$$

The reason for interpreting these as matrix operations is to formalize the algorithm that row reduces a matrix and to build the inverse of a matrix.

3.1 The column and row pictures

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, $\mathbf{R}^3$)[1].

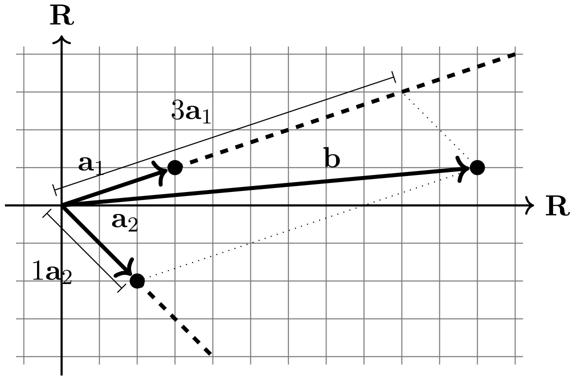
Example 3.1. Let $A = \left[\begin{smallmatrix} 3 & 2 \\ 1 & -2 \end{smallmatrix}\right] = [\,\mathbf{a}_1\ \mathbf{a}_2\,]$ and $\mathbf{b} = \left[\begin{smallmatrix} 11 \\ 1 \end{smallmatrix}\right]$, with $\mathbf{x} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$. As *columns* of A, we have a linear combination

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

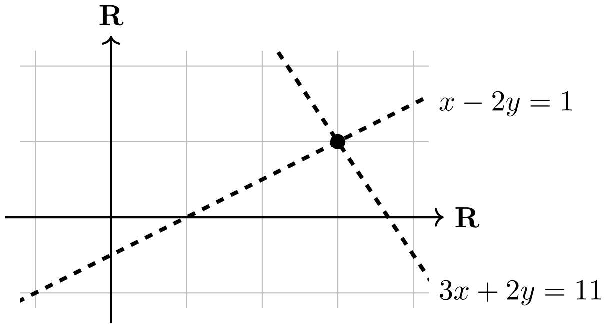
The solution to the matrix equation is the pair of coefficients x, y that satisfy the matrix equation. That is, we want to find how far along $\mathbf{a}_1$ we need to go, so that going a certain distance along $\mathbf{a}_2$ will

[1] It is more precise to say *hyperplane* to describe all the points in $\mathbf{R}^n$ satisfying a single equation. See Definition 3.5.

lead us to $\mathbf{b}$. We find a solution $\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]=\left[\begin{smallmatrix}3\\1\end{smallmatrix}\right]$:



As *rows* of A, we have two equations $3x+2y=11$ and $x-2y=1$, which we may interpret as lines in $\mathbf{R}^2$. This looks like the following picture (note that these are not the same lines as in the previous picture):



The two lines intersect at $(x,y)=(3,1)$, which is the solution $\mathbf{x}$ that solves the given matrix equation $A\mathbf{x}=\mathbf{b}$. Both the column and row pictures give the same answer! This is good.

Inquiry 3.2 (More than two rows ✠1.06): Suppose that we added another row to the matrix A and another row to the vector $\mathbf{b}$. Note that we **do not** add another row to the variable vector $\mathbf{x}$.

1. Where do the lines meet? Evaluate the matrix equation $A\mathbf{x}=\mathbf{b}$ at these points.
2. Is there a solution?
3. Draw a picture for which there is a solution.
4. What does this say about the third line? (Must be a linear combination of first two)

Remark 3.3. For the previous example, in the *row* picture:

- If the two 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 $2x-4y=2$.

Inquiry 3.4 (✠1.06): Follow the set up for drawing $A\mathbf{x}=\mathbf{b}$ from Example 3.1 for this inquiry.

1. Draw the row and column pictures for $\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]\mathbf{x}=\left[\begin{smallmatrix}3\\-4\end{smallmatrix}\right]$. What is the solution $\mathbf{x}$?
2. Draw the row and column pictures for $\left[\begin{smallmatrix}-1&1\\2&-4\end{smallmatrix}\right]\mathbf{x}=\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$. What is the solution $\mathbf{x}$?

3. What if $A = \left[\begin{smallmatrix} -1 & 1 \\ 2 & -2 \end{smallmatrix}\right]$ for the previous point? Is there more than one solution?
4. Interpret $A\mathbf{x} = 0$ having more than one solution, as a relationship between the columns (or rows) of A.

We now set up a specific algorithm (this will be the *Gaussian elimination* algorithm you may have seen earlier) for finding the solution vector $\mathbf{x}$ to a matrix equation $A\mathbf{x} = \mathbf{b}$.

Definition 3.5: Let $A\mathbf{x} = \mathbf{b}$ be a matrix equation with $A \in \mathcal{M}_{m\times n}$ and $\mathbf{b} \in \mathbf{R}^m$. The *augmented matrix* associated to this equation is the $m \times (n+1)$ matrix

$$\left[\; A \;\middle|\; \mathbf{b} \;\right] = \left[\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} & b_m \end{array}\right].$$

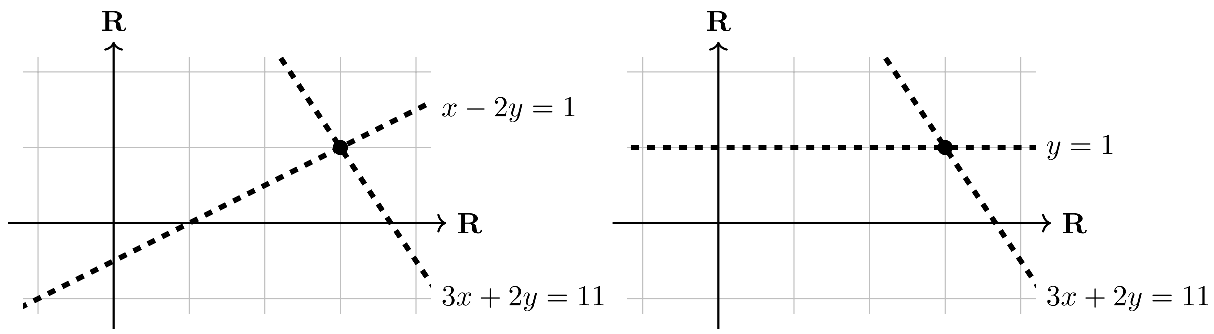
Sometimes the line separating the last two columns is not drawn. Each line $i = 1, \dots, m$ of the augmented matrix represents an equation

$$a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + a_{in}x_n = b_i$$

in n variables $x_1, \dots, x_n$ and defines a *hyperplane* in $\mathbf{R}^n$.

As you saw in Inquiry 3.4, having $A = I$ in your matrix equation makes it very easy to solve. That will be our goal now - to modify the matrix equation so that we get I instead of A. Firest, we need to make sure that this does not change the solution to the equation.

Example 3.6. Consider the augmented matrix $\left[\begin{smallmatrix} 3 & 2 & 11 \\ 1 & -2 & 1 \end{smallmatrix}\right]$ from Example 3.1. To get the first two columsn to be $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, we first will make the $(2,1)$-entry equal to zero. In the *row* picture, this means we are making the second equation flat (it will not change as x changes). The intersection of the two lines stays the same:



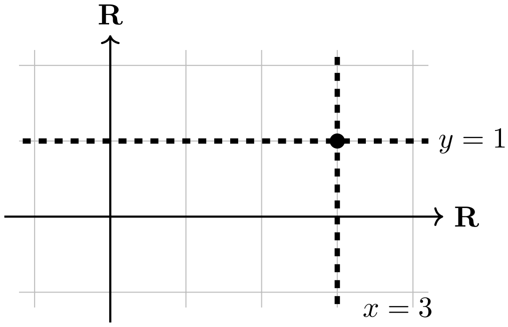
Here we added $-\frac{1}{3}$ of the first line to the second line:

$$\begin{bmatrix} 1 & 0 \\ \frac{-1}{3} & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 & 11 \\ 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 11 \\ 0 & -2 - \frac{2}{3} & 1 - \frac{11}{3} \end{bmatrix} = \begin{bmatrix} 3 & 2 & 11 \\ 0 & \frac{-8}{3} & \frac{-8}{3} \end{bmatrix},$$

so technically the second equation is $-\frac{8}{3}y = -\frac{8}{3}$. Multiplying the second row by $-\frac{3}{8}$ gives the equation as we would like it to be:

$$\begin{bmatrix} 1 & 0 \\ 0 & -\frac{3}{8} \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 & 11 \\ 0 & \frac{-8}{3} & \frac{-8}{3} \end{bmatrix} = \begin{bmatrix} 3 & 2 & 11 \\ 0 & 1 & 1 \end{bmatrix}.$$

Adding -2 of the second line to the first line makes the $(2,1)$-entry 0, and makes the two lines perpendicular:



The matrix multiplication corresponding to this will give us $3x = 9$, so we simplify as well:

$$-2 \text{ times second row plus first row: } \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 & 11 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 9 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\frac{1}{3} \text{ times first row: } \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 & 9 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 1 \end{bmatrix}$$

We put all the matrices together from tall the steps:

$$\underbrace{\begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -\frac{3}{8} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{3} & 1 \end{bmatrix}}_{\text{row operations}} \underbrace{\begin{bmatrix} 3 & 2 & 11 \\ 1 & -2 & 1 \end{bmatrix}}_{[A \mid \mathbf{b}]} = \underbrace{\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 1 \end{bmatrix}}_{[I \mid \mathbf{c}]}.$$

Inquiry 3.7 (✠1.06): This inquiry is about extending Example 3.6.

1. Multiply together the row operation matrices to get a matrix B. Compute BA and AB. What do you get? What can you conclude about B?
2. Repeat the steps and draw the pictures for the example, but use the *column* perspective instead of the *row* perspective.

3.2 The matrices doing the work

The above steps to change a given A matrix to the identity matrix (or to something as close as possible to it) are called *Gaussian elimination*.

Definition 3.8: Let $A\mathbf{x} = \mathbf{b}$ be a matrix equation. For each row i of the augmented matrix $[A \mid \mathbf{b}]$, before any operations are done with row i,

- if $A_{ii} \neq 0$, then A_{ii} is the ith *pivot*; if $A_{ii} = 0$, then the ith pivot does not exist,
- if $A_{ii} \neq 0$, for each $k > i$, the ratio $-\frac{A_{ki}}{A_{ii}}$ is the ki-*multiplier* ℓ_{ki}.

Definition 3.9: Each step of Gaussian elimination is performed by an *elementary matrix*, which is one of:

$$P_{13} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad E_{31} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{2}{5} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad D_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{10} \end{bmatrix}$$

swaps rows 1 and 3
permutation matrix

subtracts $\ell_{31} = \frac{2}{5}$ times row 1 from row 3
elimination matrix

multiplies row 4 by $\frac{1}{10}$
diagonal matrix

Shorthand notation for these matrices is given next to them above. When performed on a matrix A, these steps are equivalent to notation you may have seen before:

$$P_{13}A \text{ is } R_1 \leftrightarrow R_3 \qquad E_{31}A \text{ is } R_3 \to R_3 - \tfrac{2}{5}R_1 \qquad D_4A \text{ is } R_4 \to \tfrac{1}{10}R_4$$

In general, any $n \times n$ matrix that is just I with the rows rearranged is a *permuation matrix*. The steps of Gaussian elimination performed in the reverse order, starting from the bottom left and clearing zeros *above* each pivot is called *Gauss–Jordan elimination*, which is usually used to compute the inverse A^{-1}. Together the two are simply called *elimination*.

3.3 Exercises

Exercise 3.1. (✠1.03) This question is about the three permutation matrix examples given in Definition 3.9.

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

Exercise 3.2. (✠1.03) 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 $A_1A_2\cdots A_k$?

Exercise 3.3. (✠1.07) Let $a, b, c \in \mathbf{R}$ be nonzero numbers, and consider the matrix $A = \left[\begin{smallmatrix} a & b & c \\ a & 2b & 3c \\ a & 3b & 6c \end{smallmatrix}\right]$.

1. Give the elementary matrices which, when they are multiplied on the left of A, leave A with zeros below the diagonal (not above).
2. Let E be the product of the elementary matrices you computed in the first part of this question, and suppose that you began with an equation $A\mathbf{x} = \mathbf{b}$. If $E\mathbf{b} = \left[\begin{smallmatrix} 47 \\ 4 \\ 7 \end{smallmatrix}\right]$, what is $\mathbf{b}$? What are a, b, c?

Lecture 4: The Gaussian algorithm

Chapters 2.3, 2.4 in Strang's "Linear Algebra"

- Fact 1: There is an algorithm to row reduce a matrix
- Fact 2: Irrespective of the steps taken to row reduce a matrix, the result will always be the same.

✠ Standard 1.08: Bring a matrix to (reduced) row echelon form using Gaussian elimination

✠ Standard 1.09: Construct the inverse of a matrix A by eliminating the block matrix $[A\ I]$

This lecture is about the Gaussian algorithm and Gauss–Jordan elimination, to solve systems of linear equations and to find the inverse of a matrix.

4.1 Gaussian elimination to clear entries below the diagonal

We now formalize Example 3.6 into a proper *algorithm* that transforms the augmented matrix $[A \mid \mathbf{b}]$ into the augmented matrix $[I \mid \mathbf{c}]$, or at least as close as possible (it may be that some elements on the diagonal of I may be zero instead of 1):

Algorithm 1 (The Gaussian algorithm to row reduce a matrix):

1. Look at the $(1,1)$-entry A_{11}.
   (a) If $A_{11} \neq 0$: Make all entries below A_{11} zero.
      i. Add $-\frac{A_{21}}{A_{11}}$ of row 1 to row 2.
      ii. Add $-\frac{A_{31}}{A_{11}}$ of row 1 to row 3, and keep going until everything below A_{11} is zero.
   (b) If $A_{11} = 0$:
      i. If $A_{21} \neq 0$, swap row 1 and row 2 so that the new $(1,1)$-entry is not zero. Go back to Step 1.
      ii. Else if $A_{21} = 0$ and $A_{31} \neq 0$, swap row 1 with row 3, so that the new $(1,1)$-entry is not zero. Go back to Step 1.
      iii. Else if $A_{21} = A_{31} = 0$ and $A_{41} \neq 0$, swap row 1 with row 4.
      iv. $\vdots$
      v. If the 1st column is all zeros, go to Step 2.
2. Look at the $(2,2)$-entry A_{22}.
   (a) Repeat steps (a) and (b) above, increasing all the row and column indices by 1. That is, get zeros below A_{22}.
3. Repeating this for every, the matrix A is now upper triangular. That is, the (i,j)-entry should be 0 for $i > j$.
4. Multiply each row by the reciprocal of its first nonzero term.

This algorithm brings the augmented matrix $[A \mid \mathbf{b}]$ to *row echelon form*.

Example 4.1. Let $A\mathbf{x} = \mathbf{b}$ be a matrix equation with $A = \begin{bmatrix} 0 & 6 & -2 \\ 4 & 8 & -4 \\ -2 & 2 & 7 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 8 \\ 12 \end{bmatrix}$. This represents the intersection of three planes in $\mathbf{R}^3$. For the associated 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 $\mathbf{c} = \left[\begin{smallmatrix} 8 \\ 2 \\ 14 \end{smallmatrix}\right]$. The letter "$U$" is used for "upper triangular". We then have three equations:

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

To find the vector $\mathbf{x}$ which solves this system, we can use back substitution from the bottom row up to find $z = 2$, $y = 1$, $x = 2$.

Inquiry 4.2 (✠1.08): So far we have seen nice matrix equations with unique solutions.

1. Use Gaussian elimination to solve $\left[\begin{smallmatrix} 2 & 0 & -1 \\ 4 & -2 & 5 \\ -2 & 2 & -6 \end{smallmatrix}\right] \mathbf{x} = \left[\begin{smallmatrix} 2 \\ 0 \\ 2 \end{smallmatrix}\right]$. Interpret the result as a statement about planes intersecting in $\mathbf{R}^3$.
2. Use Gaussian elimination to solve $\left[\begin{smallmatrix} -3 & 1 \\ 1 & 1 \\ 6 & 2 \end{smallmatrix}\right] \mathbf{x} = \left[\begin{smallmatrix} 0 \\ -3 \\ -4 \end{smallmatrix}\right]$. Interpret the result as a statement about lines intersecting in $\mathbf{R}^2$.

Note that $\mathbf{x}$ in part 1. has three components, while in part 2. it has two components.

4.2 Gauss–Jordan elimination to find the inverse

In this case, we use the block matrix $[A\ I]$, and clear both above and below the diagonal. Performing Gauss–Jordan elimination on $[A\ I]$ will result in the matrix $[I\ A^{-1}]$.

Algorithm 2 (The Gauss–Jordan algorithm to find the inverse of a square matrix):

1. Perform Gaussian elimination on the block matrix $[A\ I]$.
2. Do the same steps as in Algorithm 1, but start from the bottom right and move upwards (as opposed to starting from the top right and moving downwards).
   (a) If the result is $[I\ *]$: The matrix A has an inverse, it is on the right of $[I\ *]$.
   (b) If the result is not $[I\ *]$: The matrix A does not have an inverse.

Remark 4.3. As observed in Inquiry 3.7, the elementary matrices together form the inverse. Below are 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

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

Example 4.4. Let $A = \left[\begin{smallmatrix} 4 & 8 & -4 \\ 0 & 6 & -2 \\ -2 & 2 & 1 \end{smallmatrix}\right]$, 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 4.1, 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 4.5. We now have a new, equivalent definition of $A \in \mathcal{M}_{n\times n}$ not having an inverse: If 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.

Inquiry 4.6 (✠1.09): This inquiry is about elimination using block matrices. Let $A, B \in \mathcal{M}_{2\times 2}$ have inverses.

1. Let $C = \left[\begin{smallmatrix} A & 0 \\ 0 & I_2 \end{smallmatrix}\right]$ be a block matrix. Find the inverse 4×4 matrix C^{-1}.
2. Let $D = \left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right]$ be a block matrix. Find the inverse matrix D^{-1}.
3. Will $\left[\begin{smallmatrix} A & B \\ 0 & B \end{smallmatrix}\right]$ have an inverse? How do you know? What about $\left[\begin{smallmatrix} A & I_2 \\ I_2 & B \end{smallmatrix}\right]$?

4.3 Exercises

Exercise 4.1. (✠1.08) Consider the matrix equation $A\mathbf{x} = \mathbf{b}$, given by $\left[\begin{smallmatrix} 3 & -1 & 2 \\ 9 & -3 & 2 \\ 1 & -1 & -1 \end{smallmatrix}\right] \left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right] = \left[\begin{smallmatrix} 5 \\ 5 \\ -3 \end{smallmatrix}\right]$. Use Gaussian elimination on the augmented matrix $[A \mid \mathbf{b}]$ to solve for x, y, z.

Exercise 4.2. (✠1.07, 1.08) 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 4.3. (✠1.08) 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 4.4. (✠1.09) Using Gauss–Jordan elimination, find the inverse matrix of $A = \begin{bmatrix} 0 & 2 & -1 \\ 1 & 0 & -4 \\ 2 & 2 & 2 \end{bmatrix}$.

Lecture 5: Factorization

Chapters 2.5-2.7 in Strang's "Linear Algebra"

- Fact 1: Every matrix A can be decomposed as $A = LU$ into lower and upper triangular factors.
- Fact 2: Inverses of elementary matrices are elementary matrices.

✠ Standard 1.10: Decompose the matrices A and PA as the products LU and LDU.

✠ Standard 1.11: Identify when the matrix equation $A\mathbf{x} = \mathbf{b}$ has no solutions or infinitely many solutions.

This lecture is about *factorization*, or *decomposition*, for a square matrix $A \in \mathcal{M}_{n\times n}$. Similarly to factoring an integer as the product of two factors (such as $12 = 3 \cdot 4$), we will factor A as the product of two triangular matrices. We will do this in four ways:

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

The matrix L is lower triangular, 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.

5.1 Lower and upper factors

To get the lower factor L and the upper factor U, we apply the Gaussian and Gauss–Jordan algorithms from Section 4.1. First we make an observation about the inverse of elementary matrices.

Remark 5.1. The elementary matrix E_{ki} from Gaussian elimination representing the row operation that substracts ℓ_{ki} times row i from row k is just the identity with $-\ell_{ki}$ in the (ki)-position. Its inverse is similarly the identity, but with ℓ_{ki} in the same (ki)-position:

$$E_{31} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{2}{5} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad E_{31}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{2}{5} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad E_{31}E_{31}^{-1} = E_{31}^{-1}E_{31} = I.$$

The same works if $-\ell_{ki}$ is above the diagonal for Gauss–Jordan elimination (that is, $k < i$).

> **Inquiry 5.2 (✠1.10):** Consider Remark 5.1 about the inverses of elementary matrices. Let $A \in \mathcal{M}_{4\times 4}$.
>
> 1. Let E be an elementary matrix with the number -2 in its $(3,2)$-position. What row reduction (elimination) step does the multiplication EA represent?
> 2. Give an example of A so that $(EA)_{32} = 0$.
> 3. Does the inverse matrix E^{-1} represent a row operation? If yes, which one?

Example 5.3. Let $A = \left[\begin{smallmatrix} 3 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 4 & 1 \end{smallmatrix}\right]$, which is eliminated as:

$$\underbrace{\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}}_{E_{32}} \underbrace{\begin{bmatrix} 3 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 4 & 1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 3 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{A_1}}_{\text{Gaussian elimination}}, \qquad \underbrace{\underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{E_{13}} \underbrace{\begin{bmatrix} 3 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{A_1} = \underbrace{\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{A_2}}_{\text{Gauss–Jordan elimination}}.$$

This is the end of elimination, because we have a diagonal matrix. The first multiplier was $\ell_{32} = 2$ and the second multiplier was $\ell_{13} = -1$. The decomposition comes from putting these two steps together and taking inverses:

$$\begin{aligned} E_{13}E_{32}A &= A_2 \\ E_{32}A &= E_{13}^{-1}A_2 \\ A &= E_{32}^{-1}E_{13}^{-1}A_2 \end{aligned} \qquad \underbrace{\begin{bmatrix} 3 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 4 & 1 \end{bmatrix}}_{A} = \underbrace{\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}}_{E_{32}^{-1}}}_{L} \underbrace{\underbrace{\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{E_{13}^{-1}} \underbrace{\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{A_2}}_{U} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 3 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{U}$$

Remark 5.4. We make several observations about the $A = LU$ decomposition:

- The lower triangular matrix L represents the steps of Gaussian elimination, and has 1's on the diagonal.
- The upper triangular matrix U represents the steps of Gauss–Jordan elimination, and has the pivots of A on the diagonal.

Inquiry 5.5 (✠1.10): This is about extending $A = LU$ into $A = LDU$.

1. In the $A = LU$ factorization from Example 5.3, the upper triangular matrix U has numbers that are not 1's on the diagonal. Do the row reduction steps on U that make all ements on the diagonal be 1. What are the corresponding elementary matrices?
2. Express U from the previous point as $U = DU'$, where D is diagonal and U' is upper triangular with 1's on the diagonal.
3. Generalize the above point: If $U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$ with nonzero diagonal elements, decompose it as $U = DU'$.

Remark 5.6. Elimination is the same for a matrix A or an augmented matrix $[A\ \mathbf{b}]$, but the lack of pivots for the augmented matrix indicates one of two situations: if elimination produces a row with

- 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.

The implication is that if we applied the elimination algorithm to just the matrix A, then we would get a row of zeros in both cases.

Definition 5.7: A matrix $A \in \mathcal{M}_{m\times n}$ is *singular* if elimination returns at least one row of zeros. If there are no zero rows after elimination, then A is *non-singular*.

Example 5.8. Consider the matrix equation from $A\mathbf{x} = \mathbf{b}$ from Example 3.1, but change it slightly: $\begin{bmatrix} 3 & 2 \\ -3 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ 1 \end{bmatrix}$. For elimination we subtract -1 times the first row from the second row:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \left[\begin{array}{cc|c} 3 & 2 & 11 \\ -3 & -2 & 1 \end{array}\right] = \left[\begin{array}{cc|c} 3 & 2 & 11 \\ 0 & 0 & 12 \end{array}\right].$$

In the row picture, we are looking for the intersection of $3x + 2y = 11$ and $0x + 0y = 12$, or $0 = 12$. Since $0 = 12$ is a contradiction, no solution exists. Alternatively, if we changed both A and $\mathbf{b}$ to the equation $\begin{bmatrix} 3 & 2 \\ -3 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ -11 \end{bmatrix}$, then the same step of Guassian elimination would give a full row of zeros:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\left[\begin{array}{cc|c} 3 & 2 & 11 \\ -3 & -2 & -11 \end{array}\right] = \left[\begin{array}{cc|c} 3 & 2 & 11 \\ 0 & 0 & 0 \end{array}\right].$$

The row picture asks for the intersection of $3x + 2y = 11$ and $0x + 0y = 0$. We quickly see that every vector $\mathbf{x} = \begin{bmatrix} x \\ \frac{1}{2}(11-3x) \end{bmatrix}$, for any $x \in \mathbf{R}$, will satisfy the equation $A\mathbf{x} = \mathbf{b}$. Hence we have infintely many solutions.

Inquiry 5.9 (✠1.11): Under the *row* picture, as described in Example 3.1, the solution to the matrix equation $\begin{bmatrix} 3 & 2 \\ -3 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ -11 \end{bmatrix}$ from the above example asks for the intersection of the lines $3x + 2y = 11$ and $-3x - 2y = -11$.

1. The given equation is the same as $\begin{bmatrix} 3 & 2 \\ c & -2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ -11 \end{bmatrix}$ for $c = -3$. What happens if $c \neq -3$? How many solutions does the system have?

2. The given equation is the same as $\begin{bmatrix} 3 & 2 \\ -3 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ d \end{bmatrix}$ for $d = -11$. What happens if $d \neq -11$? How many solutions does the system have?

3. Following the variants in the two previous points, replace some of the number(s) in the given equation to a variable so that the equation still has infinitely many solutions, even when the variable is changed.

The elimination algorithm from Section 4.1 was made more complicated by the fact that not all pivots may exist, in which case we need to swap rows so that we do not divide by zero. We now consider this type of elimination.

5.2 Row swaps and permutation matrices

The algorithm in Section 4.1 indicated to swap rows when there are zeros in the pivot positions when we reach them. However, to get tp the desired decomposition $PA = LU$, we need to put all the matrices representing row swaps together - so every time we get to a pivot that doesn't exist (is zero), we swap rows for the original matrix, and start from the beginning.

Example 5.10. Let $A = \begin{bmatrix} 3 & 2 & -1 \\ 6 & 4 & 0 \\ 0 & 4 & 1 \end{bmatrix}$, for which Gaussian elimination begins as:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{E_{21}}\underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 6 & 4 & 0 \\ 0 & 4 & 1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & 4 & 1 \end{bmatrix}}_{A_1}, \qquad \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}_{P_{23}}\underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & 4 & 1 \end{bmatrix}}_{A_1} = \underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 0 & 4 & 1 \\ 0 & 0 & 2 \end{bmatrix}}_{A_2}.$$

The numbers to be used for pivots are highlighted - note the problem in A_1, which we resolve by a row swap. Continuing elimination from here, we would end up with something like $EPE'A = A_n$, where E and E' are elementary matrices and P is the row swap matrix. Rearranging for A is not as nice in this case, so we apply the row swap P_{23} at the very beginning,

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}_{P_{23}}\underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 6 & 4 & 0 \\ 0 & 4 & 1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 0 & 4 & 1 \\ 6 & 4 & 0 \end{bmatrix}}_{PA},$$

and now apply the usual elimination steps.

Inquiry 5.11 (✠1.10): Note that at the end of Gaussian elimination, you have a diagonal matrix on the left side, and you know inverses of diagonal matrices. This inquiry explores the elimination steps from Example 5.10.

1. Continue the elimination algorithm from A_2 until you get a diagonal matrix. Multiply its inverse to get an inverse for A (this will be the product of the elementary matrices).

2. Begin with PA instead of A, and apply the elimination algorithm to it, until you get a diagonal matrix. As before, multiply the diagonal by its inverse to get an inverse for A.

3. Compare the two inverse you got for A - are they the same? Are the elementary matrices involved in construction of the inverse the same? What are the similarities?

Remark 5.12. If swapping rows does not give you enough pivots, it may be that you will get a row of zeros, as described in Example 5.8. In this case elimination will still give you the LU-decomposition, but the difference will be that you have to stop elimination before you get a diagonal matrix.

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{E_{21}} \underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 6 & 4 & 0 \\ -3 & -2 & 2 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 0 & 0 & 2 \\ -3 & -2 & 2 \end{bmatrix}}_{A_1}, \qquad \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}}_{E_{31}} \underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 0 & 0 & 2 \\ -3 & -2 & 2 \end{bmatrix}}_{A_1} = \underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 3 \end{bmatrix}}_{A_2},$$

and swapping the second and third rows gives us $\left[\begin{smallmatrix} 3 & 2 & -1 \\ 0 & 0 & 3 \\ 0 & 0 & 2 \end{smallmatrix}\right]$, which still does not have enough pivots. However, multiplying by the inverses of the elementary matrices we applied still gives an LU-decomposition:

$$\underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 6 & 4 & 0 \\ -3 & -2 & 2 \end{bmatrix}}_{A} = \underbrace{\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}}_{E_{31}^{-1}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{E_{21}^{-1}}}_{L} \underbrace{\underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 3 \end{bmatrix}}_{A_2}}_{U} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 3 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 3 \end{bmatrix}}_{U}.$$

You now have all the tools you need to decompose the matrices A and PA as LU or LDU. We finish off the lecture with some useful types of matrices.

Definition 5.13: Let A be an $n \times m$ matrix.

- 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 $A = -A^T$. Note that if $A \in \mathcal{M}_{n\times n}$ is symmetric and all its pivots exist, then its decomposition into $A = LDU$ has $L = U^T$.

Remark 5.14. The transpose can be thought of as a function $\mathcal{M}_{m\times n} \to \mathcal{M}_{n\times m}$. As noted in Definition 2.11, if plays nicely with the addition, multiplication, and inverse functions. Moreover, the dot product of two vectors from Definition 1.1 can be thought of as matrix multiplication, if we use the transpose:

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

$$\in \mathbf{R}^n \qquad \in \mathbf{R}^n \qquad \in \mathcal{M}_{1\times n} \qquad \in \mathcal{M}_{n\times 1}$$

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 $\mathbf{w} \cdot \mathbf{v}^T$ 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 5.15. 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 5.14, we see some interesting results. For $A \in \mathcal{M}_{m\times n}$ and $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, we have

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

5.3 Row reduction in Python

This section is about using the *SciPy* package (meaning "Scientific Python"), specifically the library imported in `scipy.linalg`. Since Gaussian elimination is just an algorithm, it has been implemented to make life easier for us. One particular implementation (though not the only one) is with the `lu` function from `scipy.linalg`.

```
import numpy as np
import scipy.linalg as lu
A = np.array([[1,2,-1],[10,5,-1],[-1,-3,1]]
P,L,U = la.lu(A)
```

This function takes in a matrix A, and outputs a triple P, L, U, so that $A = PLU$. This is almost the same as the $PA = LU$ decomposition we saw earlier, but the permutation matrix has been inverted and is on the other side of the equation.

```
P
array([[0., 0., 1.],
       [1., 0., 0.],
       [0., 1., 0.]])

L
array([[ 1.  , 0.  , 0.  ],
       [-0.1, 1.  , 0.  ],
       [ 0.1, -0.6, 1.  ]])

U
array([[10.  , 5.  , -1.  ],
       [ 0.  , -2.5 , 0.9 ],
       [ 0.  , 0.  , -0.36]])
```

5.4 Exercises

Exercise 5.1. (✠1.10) Consider the matrix factorization

$$\underbrace{\begin{bmatrix} 6 & 0 & -2 \\ 1 & 3 & 4 \\ -3 & -8 & 2 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 6 & 0 & -2 \\ 0 & 2 & 13/3 \\ 0 & 0 & 113/9 \end{bmatrix}}_{U}.$$

The values a, b, c are determined by the multipliers from row operations to clear the entries below the pivots. What are these values?

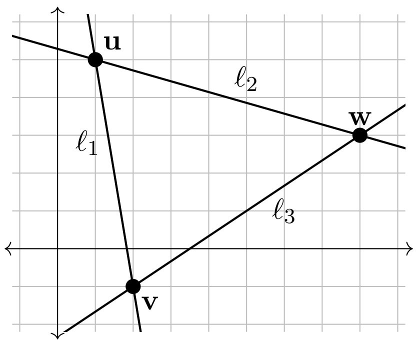
Exercise 5.2. (✠1.10) Decompose $A = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix}$ into $PA = LDU$ factorization.

Exercise 5.3. (✠1.10) Decompose the matrix A from Example 4.1 as $PA = LDU$.

Exercise 5.4. (✠1.10) Suppose that $A \in \mathcal{M}_{n\times n}$ is a product of elementary matrices, that is, $A = E_1 \cdot E_2 \cdots E_k$, where E_i is one of the three types of elementary matrices given in Definition 3.7 . Explain why A is invertible.

Exercise 5.5. (✠1.10) Consider the symmetric matrix $A = \begin{bmatrix} 5 & -6 & -3 \\ -6 & 10 & 2 \\ -3 & 2 & 3 \end{bmatrix}$. Find a 3×3 matrix B for which $B^T B = A$.
Hint: Row recduce A to the form $A = LDU$. How are L and U related to each other?

Exercise 5.6. (✠1.06, 1.11) Consider three points $\mathbf{u} = (1, 5), \mathbf{v} = (2, -1), \mathbf{w} = (8, 3)$ in $\mathbf{R}^2$. Let ℓ_1 be the line through $\mathbf{u}$ and $\mathbf{v}$, ℓ_2 be the line through $\mathbf{u}, \mathbf{w}$, and ℓ_3 be the line through $\mathbf{v}, \mathbf{w}$, as in the diagram below.



1. Give the matrix equation for which the lines in the diagram above are the row picture.

2. Without solving this matrix equation, explain why the the equation has no solutions.

3. Now suppose that $\mathbf{u} = (5, 1)$. Give the new matrix equation (the lines ℓ_1, ℓ_2, ℓ_3 are constructed in the same way), and again, without solving it, explain why it has infinitely many solutions.

Lecture 6: Vector spaces and spans

Chapter 3.1 in Strang's "Linear Algebra"

- Fact 1: A vector space is something like $\mathbf{R}^n$.
- Fact 2: Every vector space may be described as a span of vectors, in many different ways.

✠ Standard 2.01: Determine if something is a vector space or a subspace.

✠ Standard 2.02: Describe a vector space as a span of vectors.

This lecture introduces the very powerful topic of *vector spaces* and focuses on their presentation as a *span* of vectors.

6.1 Conditions to be a vector space

Recall from Lecture 2 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. We now define a set that has new properties. Any field can be used here, but we use $\mathbf{R}$ for simplicity.

Definition 6.1: Let V be a set. The elements of $\mathbf{R}$ are called *scalars*. The set V is a *vector space* if there are two operations

- addition $+\colon V \times V \to V$,
- scalar multiplication $\cdot\ : \mathbf{R} \times V \to V$,

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

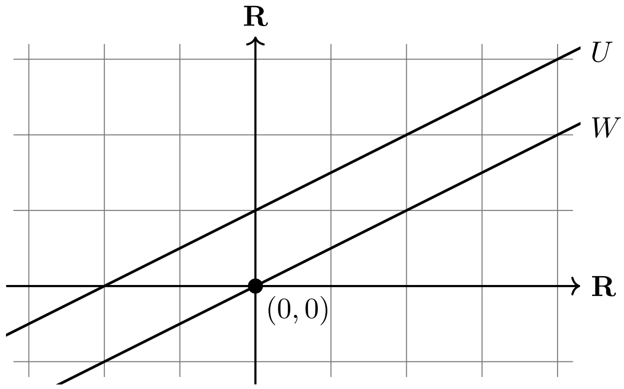
1. addition has an identity element: there exists $0 \in V$ with $0 + \mathbf{v} = \mathbf{v}$
   This is called the *additive identity*.
2. addition has inverse elements: there exists $-\mathbf{v} \in V$ with $\mathbf{v} + (-\mathbf{v}) = 0$
   This is called the *additive inverse*.
3. scalar multiplication has an identity element: there exists $1 \in \mathbf{R}$ with $1\mathbf{v} = \mathbf{v}$
   This is called the *multiplicative identity*.
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 distributive 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 6.2. We consider some basic examples of vector spaces.

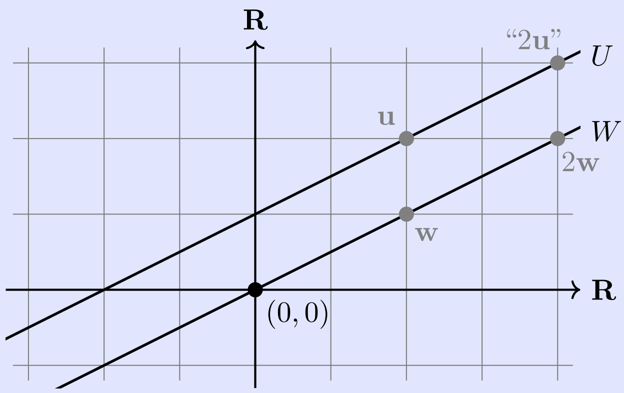
- The empty set $\emptyset$ is not a vector space, because vector space must contain the zero vector.

- 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 with the same rules as W, as $(0,0) \notin U$.



- The set $P = \{ax^2 + bx + c \ : \ a, b, c \in \mathbf{R}\}$ is a vector space, with addition and multiplication defined as expected. This is called a *function space*.

Inquiry 6.3 (✠2.01): This inquiry generalizes the notion of a vector space, continuing with the fourth example in Example 6.2. The set U there looked like it should be a vector space, since we can move along the line of U just like we moved along the line of V.



To formalize this, note that every element of U can be expressed as $\mathbf{u} = c(2,1) + (0,1)$. Define vector addition and scalar multiplication on U by

$$\begin{array}{rcl} U \times U & \to & U, \\ (\mathbf{u}_1, \mathbf{u}_2) & \mapsto & (c_1 + c_2)(2,1) + (0,1), \end{array} \qquad \begin{array}{rcl} \mathbf{R} \times U & \to & U, \\ (a, \mathbf{u}_1) & \mapsto & ac_1(2,1) + (0,1), \end{array}$$

where $\mathbf{u}_1 = c_1(2,1) + (0,1)$ and $\mathbf{u}_2 = c_2(2,1) + (0,1)$.

1. Check that multiplication distributes over addition. That is, check that property 7. is satisfied.
2. Find the additive identity, additive inverse, multiplicative identity on U so that properties 1.-3. are satisfied.
3. Explain why U, with this vector space structure, is not a subspace of $\mathbf{R}^2$.

4. Instead of $(0,1)$ at the beginning, put $(-2,0)$ in its place. What changes? Can any vector be chosen here? Which vector would you choose?

This type of structure is called an *affine space*.

Remark 6.4. 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 6.5. Combining the above remark, Example 6.2, and checking for the existence of an additive identity, multiplicative identity, and additive inverse, we see that:

- $U = \{\text{all upper triangular matrices } \left[\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right]\} \subseteq \mathcal{M}_{2\times 2}$ is a subspace of $\mathcal{M}_{2\times 2}$
- $D = \{\text{all diagonal matrices } \left[\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right]\} \subseteq \mathcal{M}_{2\times 2}$ is also a subspace of $\mathcal{M}_{2\times 2}$, and is a subspace of U

6.2 The span of a set of vectors

Definition 6.6: Let V be any vector space, such as $\mathbf{R}^n$, and $X = \{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subseteq V$ any collection of elements of V. Then the space of all *linear combinations* of elements of X, written

$$\text{span}(X) = \left\{ \sum_{i=1}^{n} c_i \mathbf{v}_i \ : \ c_i \in \mathbf{F} \right\}.$$

This space is called the *span* of the vectors in X.

Due to laziness, sometimes the curly braces $\{\dots\}$ are ommited, and we write $\text{span}(\{\mathbf{v}_1, \mathbf{v}_2\})$ and $\text{span}(\mathbf{v}_1, \mathbf{v}_2)$ to mean the same thing. With the span, we can describe a very large vector space by using a small number of vectors. Finding the smallest number of vectors will play an important role in future lectures.

Proposition 6.7. For V a vector space and $X = \{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subseteq V$, the span of X is a vector space and a subspace of V.

Proof. To see $\text{span}(X)$ is a vector space, note that every element in $\text{span}(X)$ is a vector in V. Adding two elements in $\text{span}(X)$ keeps us in the span:

$$\mathbf{a} + \mathbf{b} = \sum_{i=1}^{n} a_i \mathbf{v}_i + \sum_{i=1}^{n} b_i \mathbf{v}_i = \sum_{i=1}^{n} (a_i + b_i) \mathbf{v}_i \in \text{span}(X).$$

Scalar multiplication works similarly. The identity and inverse elements are the same as in V, and clearly the zero element is in $\text{span}(X)$, by choosing all the coefficients $c_i = 0$. Hence $\text{span}(X) \subseteq V$ is a subspace. □

Note that the above result follows immediately from Example 6.2, which said that all multiples of a single vector is a vector space, and by repeated application of Definition 6.9, which will say that $V + W$ is a vector space, for any vector spaces V, W.

Example 6.8. Two dimensional Euclidean space $\mathbf{R}^2$ can be described in several ways as a span:

- $\mathbf{R}^2 = \text{span}\left(\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} -1 \\ -1 \end{smallmatrix}\right]\right)$ because $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = \frac{x-y}{2}\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right] + \frac{x-y}{2}\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right] - y\left[\begin{smallmatrix} -1 \\ -1 \end{smallmatrix}\right]$
- $\mathbf{R}^2 = \text{span}\left(\left[\begin{smallmatrix} 3 \\ 3 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 4 \end{smallmatrix}\right]\right)$ because $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = \frac{x}{3}\left[\begin{smallmatrix} 3 \\ 3 \end{smallmatrix}\right] + \frac{3y-x}{12}\left[\begin{smallmatrix} 0 \\ 4 \end{smallmatrix}\right]$.

Definition 6.9: 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, $(\mathbf{v}_1, \mathbf{w}_1) + (\mathbf{v}_2, \mathbf{w}_2) = (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_1 + \mathbf{w}_2)$ and $c(\mathbf{v}, \mathbf{w}) = (c\mathbf{v}, c\mathbf{w})$. 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 the set $U \cup V$ (though $U \cup V$ is not necessarily a subspace).

Note that $V \oplus W$ and $V + W$ are vector spaces, 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 6.10. We note some common examples 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 6.4 for more details on why the union of two vector spaces $V \cup W$ is not the same as $+$.

Inquiry 6.11 (✠2.02): Let $V = \mathcal{M}_{2\times 2}$ and $X = \left\{\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right]\right\} \subseteq V$. Let S be the span of X.

1. Show that $S \neq V$ by finding an element in $V \setminus S$.
2. Explain why $X \subseteq S$.
3. Explain why V is as least as big as S (that is, if $M \in S$, then $M \in V$).
4. How would you change X to make $S = V$?

When $S = V$, we say that V is *spanned* by X.

6.3 Exercises

Exercise 6.1. (✠2.01) Check that the subspace $W \subseteq V$ from in the fourth example in Example 6.2 satisfies the conditions of being a vector space from Definition 6.1.

Exercise 6.2. (✠2.02) Let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be three different vectors in a vector space V. Consider the three spans $S_1 = \text{span}(\{\mathbf{u} - \mathbf{v}\})$, $S_2 = \text{span}(\{\mathbf{u}, \mathbf{v}, \mathbf{w}\})$ and $S_3 = \text{span}(\{\mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w}\})$.

1. Show that $S_1 \subseteq S_2$.
2. Show that $S_3 \subseteq S_2$.
3. For $V = \mathbf{R}^3$, given an example of $\mathbf{u}, \mathbf{v}, \mathbf{w}$ for which $S_2 = S_3$.
4. For $V = \mathbf{R}^3$, given an example of $\mathbf{u}, \mathbf{v}, \mathbf{w}$ for which all of S_1, S_2, S_3 are different.

Exercise 6.3. (✠2.01) Consider the set X of all functions $f \colon \mathbf{R} \to \mathbf{R}$.

1. If addition on X is defined as $(f+g)(x) = f(x)+g(x)$ and multiplication is defined as $(cf)(x) = f(cx)$, show that X can not be a vector space.

2. If multiplication is instead defined as $(cf)(x) = cf(x)$, and addition is instead defined as $(f + g)(x) = f(g(x))$ show that X still can not be a vector space.

Hint: Show X is not a vector space with examples!

Exercise 6.4. (✠2.01) Consider the following vector spaces:

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

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 $V \cup W$.

Lecture 7: The column space and the nullspace

Chapter 3.2 in Strang's "Linear Algebra"

- Fact 1: The column space and nullspace of any matrix are vector spaces.
- Fact 2: The nullspace of a matrix A contains all the vectors $\mathbf{x}$ for which $A\mathbf{x} = 0$.

✠ Standard 1.12: Construct the column space and nullspace of a matrix as spans.

✠ Standard 1.13: Describe solutions to $A\mathbf{x} = \mathbf{b}$ using the language of vector spaces.

This lecture provides two concrete examples of vector spaces, which were introudced in the pevious lecture: the *column space*, coming from the columns of a matrix, and the *nullspace*, representing all the "zero solutions" to a matrix equation $A\mathbf{x} = \mathbf{b}$.

7.1 The column space of a matrix

A big 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 7.1: For an $m \times n$ matrix A, the *column space* of A, denoted $\text{col}(A)$, is the set of all vectors $\mathbf{v} \in \mathbf{R}^m$ that are linear combinations of the columns of A. That is,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \qquad \text{col}(A) = \left\{ c_1 \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{m1} \end{bmatrix} + c_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + c_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} \ : \ c_i \in \mathbf{R} \right\}.$$

Since every element in $\text{col}(A)$ is a linear combination of vectors, $\text{col}(A)$ is a subspace of $\mathbf{R}^m$.

Example 7.2. Consider $A = \left[\begin{smallmatrix} 3 & -1 & -2 & 4 \\ 0 & 2 & -2 & 1 \end{smallmatrix}\right]$, for which

$$\text{col}(A) = \left\{ c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ -2 \end{bmatrix} + c_2 \begin{bmatrix} 4 \\ 1 \end{bmatrix} \ : \ c_i \in \mathbf{R} \right\}.$$

Note that $\left[\begin{smallmatrix} 5 \\ 6 \end{smallmatrix}\right] \in \text{col}(A)$, as

$$\begin{bmatrix} 5 \\ 6 \end{bmatrix} = -5 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} -1 \\ 2 \end{bmatrix} - 3 \begin{bmatrix} -2 \\ -2 \end{bmatrix} + 0 \begin{bmatrix} 4 \\ 1 \end{bmatrix}.$$

Inquiry 7.3 (✠1.12, 1.13): Let $A \in \mathcal{M}_{m\times n}$.

1. Show that $\left[\begin{smallmatrix} 3 \\ 4 \end{smallmatrix}\right] \in \text{col}\left(\left[\begin{smallmatrix} 2 & -1 & 1 \\ 0 & 1 & 1 \end{smallmatrix}\right]\right)$.
2. Suppose that $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \in \text{col}(A)$. Explain why $A\mathbf{x} = \mathbf{v}$ has a solution. What is it?
3. Suppose that $\mathbf{x} = (x_1, \ldots, x_n)$ solves the equation $A\mathbf{x} = \mathbf{b}$. Explain why $\mathbf{b} \in \text{col}(A)$.
4. Explain why $0 \in \text{col}(A)$. Hint: What is a solution to $A\mathbf{x} = 0$?

Example 7.4. 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.

7.2 The nullspace of a matrix

Another big reason we are talking about vector spaces is another link to the matrix equation $A\mathbf{x} = \mathbf{b}$, in the special case that $\mathbf{b} = 0$. All the vectors $\mathbf{x}$ satisfying this equation form a vector space. Note that all the vectors satisfying $A\mathbf{x} = \mathbf{b}$ did **not** form a vector space for arbitrary $\mathbf{b}$ - the column space was the space of all vectors $A\mathbf{x}$, not just $\mathbf{x}$.

Definition 7.5: For an $m \times n$ matrix A, the *nullspace* of A is the set

$$\text{null}(A) = \{\mathbf{x} \in \mathbf{R}^n \ : \ A\mathbf{x} = 0\}.$$

The nullspace is a vector space. The nullspace lives inside $\mathbf{R}^n$, but the column space lives in $\mathbf{R}^m$. To find the nullspace of A, we use Gaussian and Gauss–Jordan elimination on A. We may perform row swaps at the beginning or in the middle of elimination, it will not change the result.

Example 7.6. 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 $A\mathbf{x} = 0$. The second row is a multiple of the first (and the second column 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

$$\text{null}\left(\begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix}\right) = \left\{\begin{bmatrix} x_2/2 \\ x_2 \end{bmatrix} \ : \ x_2 \in \mathbf{R}\right\} = \text{span}\left(\begin{bmatrix} 1/2 \\ 1 \end{bmatrix}\right).$$

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

Remark 7.7. 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.

Example 7.8. We describe how to compute the nullspace by way of an example, on $A = \left[\begin{smallmatrix}2 & -2 & 2 & 4 & 8\\1 & 5 & -3 & 0 & 1\\3 & 3 & -1 & -5 & 6\end{smallmatrix}\right]$. We begin with Gaussian elmination to get zeros below the first pivot. The multipliers are given below, and zeros of pivot columns are highlighted:

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

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:

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

We pause this example for a few comments.

Definition 7.9: The form of A in the example above is called the *reduced row echelon form*, or *RREF*, of A. More specifically:

- columns 1,2,4 are the *pivot columns*,
- columns 3,5 are the *free columns*.

In the equation $A\mathbf{x} = 0$, for $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, the variables x_1, x_2, x_4 are the *pivot variables* and x_3, x_5 are the *free variables*.

We continue solving for the nullspace null(A) from Example 7.8. It 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:

$$\begin{bmatrix} 1 & 0 & 1/3 & 0 & 53/18 \\ 0 & 1 & -2/3 & 0 & -7/18 \\ 0 & 0 & 0 & 1 & 1/3 \end{bmatrix} \underbrace{\begin{bmatrix} -1/3 \\ 2/3 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{s}_1} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 1/3 & 0 & 53/18 \\ 0 & 1 & -2/3 & 0 & -7/18 \\ 0 & 0 & 0 & 1 & 1/3 \end{bmatrix} \underbrace{\begin{bmatrix} -53/18 \\ 7/18 \\ 0 \\ -1/3 \\ 1 \end{bmatrix}}_{\mathbf{s}_2} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The two vectors $\mathbf{s}_2, \mathbf{s}_2$ are the *special solutions* for the nullspace of A. Hence the nullspace is

$$\text{null}(A) = \{c_1\mathbf{s}_1 + c_2\mathbf{s}_2 \ : \ c_1, c_2 \in \mathbf{R}\} = \left\{ c_1 \begin{bmatrix} -1/3 \\ 2/3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} -53/18 \\ 7/18 \\ 0 \\ -1/3 \\ 1 \end{bmatrix} \ : \ c_1, c_2 \in \mathbf{R} \right\},$$

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.

Algorithm 3 (Computing the nullspace): To compute the nullspace of $A \in \mathcal{M}_{m\times n}$, we do Gaussian and Gauss–Jordan elimination so that all columns with pivots have 1's as the only entry.

1. Perform Gauss–Jordan elimination on A to clear all entries below the pivots. The matrix is now A'.

2. Perform Gaussian elimination on A' to clear all above below the pivots. The matrix is now A''.

3. Multiply A'' by diagonal matrices to make all the pivots 1's.

4. Suppose columns $c_1, \dots, c_k$ are pivot columns, and columns $f_1, \dots, f_\ell$ are free columns.

   (a) The nullspace will be a span $(\mathbf{v}_1, \dots, \mathbf{v}_\ell)$ of as many vectors as free columns. The vector $\mathbf{v}_i \in \mathbf{R}^{k+\ell}$ has:

   (b) entry 1 in row f_i and entry 0 in all other rows $f_{j\neq i}$

   (c) entry in row c_j the same, but negative, as the entry in column f_i and row j of A''

Remark 7.10. Note that the pivot columns create an identity matrix in RREF of A, which were highlighted in green and yellow in the main example above. Similarly, the free variable rows in the special solutions create an identity matrix.

Inquiry 7.11 (✠1.12): Consider the matrix $A = \begin{bmatrix} 3 & 6 & -1 & 0 & 1 \\ 9 & 18 & -3 & 2 & 0 \\ 0 & 0 & -5 & 1 & 1 \end{bmatrix}$.

1. Compute null(A) as the span of vectors.

2. Construct a matrix B for which null(A) = col(B).

3. Compute col(A) as the span of vectors.

4. Do you have to use all the columns of A? That is, are some columns linear combinations of others? Try to use as few columns of A as possible to express col(A) as a span.

7.3 Exercises

Exercise 7.1. (✠1.12) Consider the matrix $A = \begin{bmatrix} 2 & 9 & 1 & 0 & 0 & 9 \\ 0 & -3 & 0 & 1 & 0 & -3 \\ 8 & -6 & 0 & 0 & 1 & 1 \end{bmatrix}$.

1. Construct the column space of A as a span of three vectors.

2. Construct the nullspace of A as a span of vectors.

Exercise 7.2. (✠1.12) Let V be a vector space.

1. Explain why span(V) = V and span($\{0\}$) = $\{0\}$.

2. For $V = \mathbf{R}^3$, give an example of $A, B \in \mathcal{M}_{3\times 3}$ with col(A) = V and null(B) = V. Explain why A and B can not be the same matrix.

Exercise 7.3. (✠1.13) Let $A \in \mathcal{M}_{2\times 2}$, and let $\mathbf{v} = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathbf{R}^2$ with $a, b \neq 0$. For the following questions, do not choose numbers for a and b, simply leave them as variables.

1. Suppose that $A\mathbf{v} = \mathbf{v}$.

(a) Can $\mathbf{v} \in \text{null}(A)$? Why or why not?

(b) Give two different examples of a 2×2 matrix A that satisfy the given condition.

2. Suppose that $\mathbf{v}$ is the first column of A and that $\text{null}(A) = \text{span}(\mathbf{v})$. Give an example of a 2×2 matrix A that satisfies this setting.

Exercise 7.4. (✠1.12) 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 $\mathbf{R}^2$, not a multiple of I_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 7.5. (✠1.12) 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.

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

Exercise 7.6. (✠1.13) Let X be a set of 2×2 matrices defined in the following way:

- $\left[\begin{smallmatrix}1 & 1\\0 & 1\end{smallmatrix}\right] \in X$
- if $M \in X$, then $MM^T \in X$
- if $M, N \in X$, then $aM + bN \in X$, for any $a, b \in \mathbf{R}$

Using scalar multiplication and matrix addition as in $\mathcal{M}_{2\times 2}$, show that X is a vector subspace of $\mathcal{M}_{2\times 2}$.
Hint: Using the given facts, try to construct the four special matrices that generate $\mathcal{M}_{2\times 2}$.

Lecture 8: Completely solving $A\mathbf{x} = \mathbf{b}$

Chapter 3.3 in Strang's "Linear Algebra"

- Fact 1: The complete solution to $A\mathbf{x} = \mathbf{b}$ consists of the particular solution and linear combinations of the special solutions.
- Fact 2: The rank of a matrix is the number of pivots. It can not be larger than the number of rows or columns.

✠ Standard 1.14: Construct the particular, special, and complete solutions to $A\mathbf{x} = \mathbf{b}$, for any matrix $A \in \mathcal{M}_{m\times n}$.

✠ Standard 1.15: Identify the row rank, column rank, rank of a matrix.

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. In this lecture we generalize to finding soluetions to $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b}$ is not necessarily the zero vector.

8.1 Rank and the particular solution

We begin with the example from the previous lecture,

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

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 8.1: The *rank* of a matrix $A \in \mathcal{M}_{m\times n}$ is denoted rank(A), and is equivalently

- the number of pivots of A, or
- the largest number of columns in A that are not linear combinations of each other.

This number is denoted rank(A). Often *column rank* or *row rank* are used, when specifically referencing the largest number of columns or rows that are not linear combinations of each other.

If the rank of A is equal to the largest number of rows of A that are not linear combinations of each other, then A is said to have *full rank*. Equivalently, $\text{rank}(A) = \min(m, n)$.

Reducing the matrix A to RREF reveals which columns are combinations of others. Since only row operations were performed, any linear relationships among the columns are preserved.

Example 8.2. When a matrix has rank 1, all the columns are multiples of the first one. For example,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

has rank one, and its column space is all the multiples of $(1, 1, 1)$. To find its nullspace, we look at its RREF, 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 the nullspace of A is the span of $\left[\begin{smallmatrix} -2 \\ 1 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} -3 \\ 0 \\ 1 \end{smallmatrix}\right]$.

Remark 8.3. 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 8.4. The identity matrix I has full rank. The zero matrix 0 has rank 0.

Inquiry 8.5 (✠1.15): Consider the matrix $A = \left[\begin{smallmatrix} a & b & c \\ b & c & b \\ c & a & a \end{smallmatrix}\right]$.

1. Find values of a, b, c for which A has rank 0, 1, 2, 3.
2. Suppose another column was added at the end of A to make $\left[\begin{smallmatrix} a & b & c & 0 \\ b & c & b & 0 \\ c & a & a & 0 \end{smallmatrix}\right]$. Explain why your answers to the first part above would not change using this matrix.

Definition 8.6: 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 that can be used to define null(A) as a span. For $A \in \mathcal{M}_{m\times n}$, using the fact that the rank is the number of pivot columns, we immediately get that

$$\text{rank}(A) + \text{nullity}(A) = n, \tag{2}$$

a very powerful equation, more of which we will see later. This is called the *rank-nullity theorem*.

Example 8.7. Recall Example 7.8 from Lecture 7. Suppose that instead of $A\mathbf{x} = 0$, we considered $A\mathbf{x} = \mathbf{b}$, which, after elimination, would become $R\mathbf{x} = \mathbf{d} = [d_1\ d_2\ d_3]^T$. 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:

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

The vector $(d_1, d_2, 0, d_3, 0)$ is called a *particular solution* to $A\mathbf{x} = \mathbf{b}$. This particular solution solves not only $R\mathbf{x} = \mathbf{d}$, but also $A\mathbf{x} = \mathbf{b}$, because if $A = ER$, for some elimination matrix E, then $\mathbf{d} = E\mathbf{b}$.

Remark 8.8. What we have done so far can be summarized as follows:

- The special solutions $\mathbf{x} = \mathbf{s}_1, \mathbf{s}_2$ solve $A\mathbf{x} = 0$
- The particular solution $\mathbf{x} = \mathbf{p}$ solves $A\mathbf{x} = \mathbf{b}$

Finally, the *complete solution* to the system $A\mathbf{x} = \mathbf{b}$ is the sum of the particular and special solutions. That is, $\mathbf{x} = \mathbf{p} + c_1\mathbf{s}_1 + c_2\mathbf{s}_2$ solves the system, for any $c_1, c_2 \in \mathbf{R}$, because

$$A(\mathbf{p} + c_1\mathbf{s}_1 + c_2\mathbf{s}_2) = A\mathbf{p} + c_1A\mathbf{s}_1 + c_2A\mathbf{s}_2 = \mathbf{b} + c_1 \cdot 0 + c_2 \cdot 0 = \mathbf{b}.$$

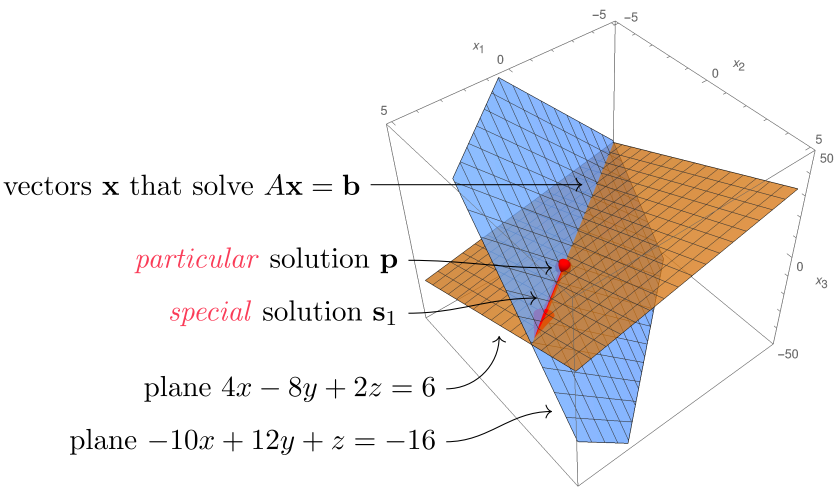
Algorithm 4 (Finding the complete solution): Consider the matrix equation $A\mathbf{x} = \mathbf{b}$.

1. Compute the nullspace of $[A \mid \mathbf{b}]$. That is, find the special solutions $\mathbf{s}_1, \dots, \mathbf{s}_k$ by doing elimination on the augmented matrix $[A \mid \mathbf{b}]$.
2. Elimination on $[A \mid \mathbf{b}]$ produces the matrix $[R \mid \mathbf{d}]$. Construct the particular solution $\mathbf{p}$ from $\mathbf{d}$ as in Example 8.7.
3. The complete solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{p} + c_1\mathbf{s}_1 + \cdots + c_k\mathbf{s}_k$, for all $c_i \in \mathbf{R}$.

Example 8.9. Consider the matrix equation $A\mathbf{x} = \mathbf{b}$, in the form

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

The complete solution to this equation is $\mathbf{x} = \begin{bmatrix} 7/4 \\ 1/8 \\ 0 \end{bmatrix} + c_1 \begin{bmatrix} 1 \\ 3/4 \\ 1 \end{bmatrix}$, for any $c_1 \in \mathbf{R}$.



This equation represents two planes intersecting in space, as in the picture above. The particular solution is a point on the line of intersection and the special solution is a vector in the direction of the line. The line of intersection is all the vectors $\mathbf{x}$ that make $A\mathbf{x} = \mathbf{b}$ true. In other words, it the nullspace shifted by the vector $\mathbf{p}$, hence it is an *affine space*.

In the example above, the two planes are defined by the initial equation. After row reduction, we have two different planes which still have same intersection. Compare this with the 2-dimensional row picture presented in Example 3.6.

Inquiry 8.10 (✠1.14): Consider the nullspace null(A) from Example 8.9.

1. Write the nullspace null(A) as the span of a single vector.
2. Let $\hat{\mathbf{x}}$ be a solution to $A\mathbf{x} = \mathbf{b}$. Explain why $2\hat{\mathbf{x}}$ is not a solution to $A\mathbf{x} = \mathbf{b}$.

3. Using Inquiry 6.3, explain why the collection of all solutions still has some vector space structure (even though "multiplying" vectors in the usual sense does not work, as shown in the previous point). This type of space is an ________ space.

8.2 Different types of complete solutions

Now we consider the implications for the complete solution given the rank of the matrix. Recall from Definition 8.1 that $A \in \mathcal{M}_{m\times n}$ has *full rank* if it has A has $\min(m,n)$ pivots.

Definition 8.11: Let $A \in \mathcal{M}_{m\times n}$.

- If each row of A has a pivot (so A has m pivots), then A has *full row rank*.
- If each column of A has a pivot (so A has n pivots), then A has *full column rank*.

Example 8.12. Consider the following types of common situations for rank, for $A \in \mathcal{M}_{m\times n}$. If A has more rows than columns (so $m > n$) and has full column rank, then in row reduced echelon form A 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:

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

- 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.

Analogously, If A has more columns than rows (so $n > m$) and has full row rank, then in row reduced echelon form A 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:

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

- 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 $\mathbf{b}$

Example 8.13. The equation $A\mathbf{x} = \mathbf{b}$ with A an $m \times 3$ matrix with full column rank represents m planes intersecting in 3-dimensional space $\mathbf{R}^3$. If the planes all intersect in one point, there is a solution to this equation.

- For $1 \leqslant m < 3$ and m randomly chosen planes, it is impossible for them to intersect in one point.
- For $m = 3$ and three randomly chosen planes, they will almost always intersect in one point.
- For $m > 3$ and m randomly chosen planes, they will almost never intersect in one point.

The general theory behind these claims has to do with *general position* of points in $\mathbf{R}^3$, and the fact that three points are necessary to define a plane.

Inquiry 8.14 (✠1.14): Let $A \in \mathcal{M}_{m\times n}$ and $\mathbf{b} \in \mathbf{R}^m$. Find an example of A and $\mathbf{b}$ so that $A\mathbf{x} = \mathbf{b}$ has:

1. exactly one solution, with $m = n = 2$;
2. no solutions, with $m = n = 2$;

3. exactly one solution, with $m = 3$, $n = 2$;
4. no solutions, with $m = 3$, $n = 2$;
5. infinitely many solutions, with $m = 2$, $n = 3$.
6. Explain why $A\mathbf{x} = \mathbf{b}$ can not have exactly one solution if $n > m$. That is, show that if it has one solution, it has infinitely many.

Remark 8.15. 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, depending on what the bottom row(s) of $[A \mid \mathbf{b}]$ look like in RREF.
- $\text{rank}(A) < m$ 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.

8.3 Exercises

Exercise 8.1. (✠1.15) 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 8.2. (✠1.14) Find the complete solution to $A\mathbf{x} = \mathbf{b}$, for

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

Exercise 8.3. (✠1.14) 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

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

for any $c \in \mathbf{R}$.

1. Construct one possible matrix A and vector $\mathbf{b}$ for which this could be the solution.
2. Do the same as above, but make it so that A has no zero entries.

Exercise 8.4. (✠1.15) For the following matrices A, B, find the ranks of A^TA, AA^T, B^TA, BB^T:

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

Exercise 8.5. (✠1.15) Consider the vectors $\mathbf{a} = \begin{bmatrix} a \\ a \\ a \\ a \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} b \\ b \\ b \\ b \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, for $a, b \in \mathbf{R}$.

1. What will be the rank of the following 4×4 matrices:

(a) $\mathbf{a}\mathbf{u}^T$ (b) $\mathbf{b}\mathbf{v}^T$ (c) $\mathbf{a}\mathbf{u}^T + \mathbf{b}\mathbf{v}^T$

Your answers should depend on a and b.

2. Explain why the rank of $\mathbf{x}\mathbf{y}^T$, for any $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, can never be greater than 1.

Lecture 9: Complex numbers

Chapters 9.1 and 9.2 in Strang

- Fact 1: All the math we have done so far can be considered over $\mathbf{C}$ instead of $\mathbf{R}$
- Fact 2: Complex number adition and multiplication have geometric meaning

---

✠ Standard 4.01: Express a complex number in one of four different ways

✠ Standard 4.02: Translate known properties of vectors and matrices to Hermitian vectors and matrices

---

In this lecture we will take some time to introduce fully the topic of coplex numbers. Fortunately, almost all the results we have seen so far with matrices over $\mathbf{R}$ apply to matrices over $\mathbf{C}$ as well.

9.1 The space of complex numbers

Definition 9.1: The *complex numbers* are elements of the set $\mathbf{C} = \{x + iy \ : \ x, y \in \mathbf{R}\}$. The symbol i is the *imaginary number*, having the property that $i^2 = -1$. For every $z = x + iy \in \mathbf{C}$:

- the *standard form* of z is $x + iy$.
- in *Cartesian*, or *rectangular* coordinates, the number z is written (x, y).

The *real part* of z is x and its *imaginary part* is y. If $x = 0$, then z is a *purely imaginary number*.

Let $z = x + iy$ and $w = a + ib$ be complex numbers and $c \in \mathbf{R}$. Complex number addition and multiplication, and real number multiplication are defined in the following way:

$$\begin{aligned} z + w &= (a + x) + i(y + b) \\ zw &= xa + ixb + iya + i^2tb = (xa - yb) + i(xb + ya) \\ cz &= cx + icy \end{aligned}$$

Inquiry 9.2 (✠4.02): The set $\mathbf{C}$ along with complex number addition and scalar multiplication as above form a vector space.

1. Show that the function $f\colon \mathbf{C} \to \mathbf{R}^2$, given by $f(x + iy) = (x, y)$ is a bijection.
2. With the bijection from above, the complex number $z = 1 + i$ could be considered as the vector $\mathbf{v} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right] \in \mathbf{R}^2$. Compute the square $z \cdot z$ and the dot product $\mathbf{v} \bullet \mathbf{v}$. Why do you get two dfferent results?
3. For any $z \in \mathbf{C}$, will $z \cdot z$ always be a real number? Give an example when it is and another example when it isn't.
4. Describe a surjective function $\mathbf{C} \to \mathbf{R}$ that takes in a complex number, and outputs a real number.

Example 9.3. 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 9.4: Let $z = x + yi \in \mathbf{C}$. The (*complex*) *conjugate* of z is $\overline{z} = z^* = x - iy$. The *absolute value*, or *modulus* of z is

$$|z| = \sqrt{z\overline{z}} = \sqrt{(x+iy)(x-iy)} = \sqrt{x^2+y^2}.$$

Taking the conjugate twice returns back the original number: $(z^*)^* = z$.

Proposition 9.5. Let $z = x + iy, w = a + ib \in \mathbf{C}$. 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:

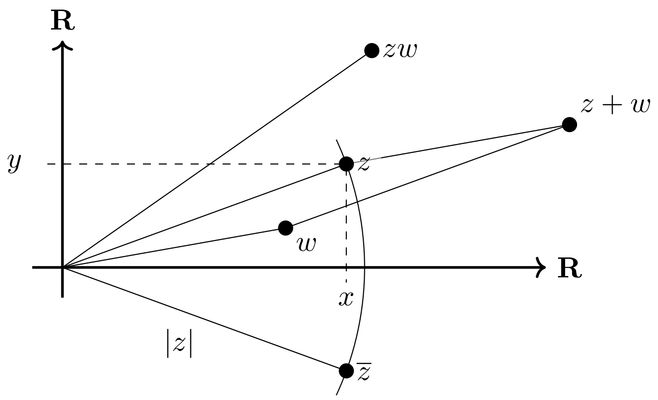
1. $|z| = 0$ iff $z = 0$
2. $|\overline{z}| = |z|$
3. $|zw| = |z||w|$
4. $|z + w| \leqslant |z| + |w|$

Definition 9.6: The third way to express $z = x + iy \in \mathbf{C}$ 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

$$x + iy = r\cos(\theta) + ir\sin(\theta) = re^{i\theta},$$

where the second equality is known as *Euler's formula*. This last expression is in *exponential form*.

Remark 9.7. 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.



$$\begin{aligned}
\overline{z} &= x - iy = r_z\cos(\theta_z) - ir_z\sin(\theta_z) \\
z &= x + iy = r_z\cos(\theta_z) + ir_z\sin(\theta_z) \\
z + w &= r_w\cos(\theta_w) + ir_w\sin(\theta_w) \\
zw &= r_z r_w e^{i(\theta_z+\theta_w)}
\end{aligned}$$

Remark 9.8. Putting complex numbers into polar coordinates makes computations in standard form much easier. For $z = re^{i\theta}$ and $n \in \mathbf{N}$, 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 $k = 0, 1, \ldots, n-1$.

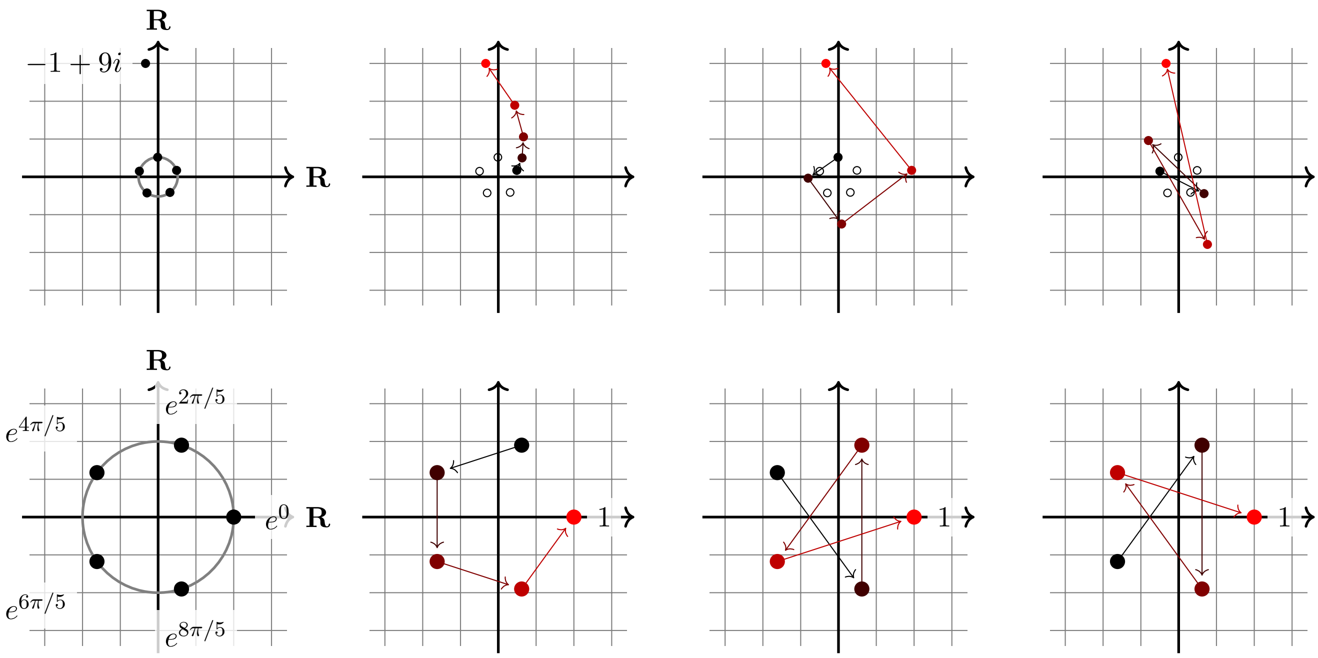
For the second point, when $z = 1 + 0i$, then the kth root of z is called the *kth root of unity*.

Inquiry 9.9 (✠4.01): This inquiry is about the different forms of complex numbers.

1. Express $z = 5\cos(\pi/4) + 5i\sin(\pi/4)$ in standard form.
2. Express $w = -\sqrt{3} - i$ in polar form.

3. Find the 4th roots of $p = 1 + i$ in Cartesian coordinates.
4. Explain why finding nth roots of unity is much easier in polar coordinates than in rectangular coordinates.

Example 9.10. 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 9.11. The space of complex numbers is a 2-dimensional vector space over $\mathbf{R}$ via the identification of Cartesian coordinates. However, it is a 1-dimensional vector space over $\mathbf{C}$.

9.2 Complex vectors and complex matrices

Just like we generalized numbers to vectors, we generalize complex numbers to complex matrices. We now talk about the vector space $\mathbf{C}^n$, of vectors having n components, and the matrix space $\mathcal{M}_{m\times n}(\mathbf{C})$, of $m \times n$ matrices with complex number entries.

Remark 9.12. Multiplication of complex numbers may be viewed as matrix multiplication. Making a correspondence between $z = x + iy \in \mathbf{C}$ and $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \in \mathbf{R}^2$, as in Inquiry 25.2, reveals a correspondence for multiplication:

$$(a + ib)(x + iy) \quad \leftrightarrow \quad \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Definition 9.13: 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 $\mathbf{z} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = x + iy = z$, 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, with $A^* \in \mathcal{M}_{n\times m}(\mathbf{C})$ whenever

$A \in \mathcal{M}_{m\times n}(\mathbf{C})$ defined by $(A^*)_{ij} = \overline{A_{ji}}$.

Definition 9.14: The square matrix $A \in \mathcal{M}_{n\times n}(\mathbf{C})$ is is *Hermitian* if $A = A^*$.

We will see nice properties of Hermitian matrices later in the course. For now we consider some of their properties.

Proposition 9.15. Let $A, B \in \mathcal{M}_{n\times n}(\mathbf{C})$ be Hermitian. Then:

- the entries on the diagonal of A are real numbers
- the identity $(AB)^* = B^*A^*$ holds

Inquiry 9.16 (✠4.02): Let $A \in \mathcal{M}_{n\times n}(\mathbf{C})$ be Hermitian, and let $\mathbf{v} \in \mathbf{C}^n$.

1. Expand the product $(\mathbf{v}^*A\mathbf{v})^*$ to show that it is Hermitian. How many rows and columns does the product have, and in what space must it be?
2. Find the complete solution to $\left[\begin{smallmatrix} 0 & 3+i \\ 3-i & 0 \end{smallmatrix}\right] \mathbf{z} = \left[\begin{smallmatrix} i \\ 2-i \end{smallmatrix}\right]$, for $\mathbf{z} \in \mathbf{C}^2$.

9.3 Exercises

Exercise 9.1. (✠4.02) Show that every complex number $z = x + iy$ 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 9.2. (✠4.02) Prove all the claims of Proposition 9.5, 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 9.3. (✠4.01) This question is about proving Euler's formula $\cos(\theta) + i\sin(\theta) = e^{i\theta}$.

1. Take the derivative of $f(\theta) = (\cos(\theta) + i\sin(\theta))e^{-i\theta}$ with respect to θ.
2. Explain why the result of the previous step means that $f(\theta)$ is constant.
3. Evaluate f at $\theta = 0$ to find this constant from the previous step.
4. Rearrange to get Euler's formula.

Part II
Vector spaces

Lecture 10: Independence, basis, dimension

Chapter 3.4 in Strang's "Linear Algebra"

- Fact 1: A basis of a vector space V is a smalest possible set of vectors that spans V
- Fact 2: Bases of V are not unique. The size of a basis is unique - it is the dimension of V.

✠ Standard 2.03: Identify linearly independent subsets in a given set of vectors.

✠ Standard 2.04: Express the same vector in different bases.

✠ Standard 2.05: Find a basis the dimension of a vector space.

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

10.1 Linear independence

Recall that the rank of a matrix A was the number of pivots A had, or the number of columns of A that are not linear combinations of the other columns. A more precise way to say the second approach is with *linear independence*.

Definition 10.1: Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subseteq \mathbf{R}^m$ be the columns of a matrix $A \in \mathcal{M}_{m\times n}$. These vectors are *linearly independent* if, equivalently,

- the only solution to $A\mathbf{x} = 0$ is $\mathbf{x} = 0$, or
- $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = 0$ implies $x_i = 0$ for all i, or
- the nullspace of A is only the zero vector, that is, $\text{null}(A) = \{0\}$.

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, there is no in-between. 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 10.2. Slight changes in the matrix entries can lead to big differences:

- 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 $\left[\begin{smallmatrix}2\\2.001\end{smallmatrix}\right]$ are linearly independent, because attempting to solve $A\mathbf{x} = 0$ will lead to $\mathbf{x} = 0$.

Inquiry 10.3 (✠2.03): Recall the three different ways to express linear independence.

1. Pick any 3 vectors in $\mathbf{R}^2$. Explain why they must be linearly dependent. *Hint: put them as columns in a matrix and say something about its nullspace.*
2. Does the above work for any $4, 5, \ldots$ vectors in $\mathbf{R}^2$? What about any 2 vectors?
3. Try to generalize the above points into a statement like: "Any set of more than ____ vectors

in ____ will be linearly ____."

Recall the *span* of a collection of vectors from Definition 6.6 and Inquiry 6.11, and the columns of a matrix *spanning* its column space, as well as the vectors from special solutions *spanning* the nullspace.

Definition 10.4: Let V be a vector space and $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subseteq V$. If $V = \text{span}(S)$, then S is called a *spanning set* of V.

Example 10.5. Spanning sets are many and can be easily constructed.

- The vector space $\mathbf{R}^3$ has a spanning set in $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$, as well as $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$.
- The pivot columns of a matrix form a spanning set for its column space.

Some spanning sets are more special than others. In particular, when talking about the size of the spanning set (number of vectors that it has, the number k from Definition 10.4), a *minimal spanning set* of V is one that is never larger than any other spanning set of V.

Definition 10.6: 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, equivalently,

- S is a minimal spanning set for V, or
- S spans V, and S is linearly independent, or
- every $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$, for $a_i \in \mathbf{R}$.

Example 10.7. The *standard basis* for $\mathbf{R}^3$ consists of the vectors $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$. In general, the *standard basis* for $\mathbf{R}^n$ consists of the n column vectors of the $n \times n$ identity matrix, and they are often denoted $\mathbf{e}_1, \dots, \mathbf{e}_n$:

$$\begin{bmatrix}1 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1\end{bmatrix} = \begin{bmatrix}| & | & | & & |\\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \cdots & \mathbf{e}_n\\ | & | & | & & |\end{bmatrix}.$$

The standard basis is not the only basis for $\mathbf{R}^n$, and the columns of every full rank $n \times n$ matrix will give a basis for $\mathbf{R}^n$.

Example 10.8. Let $A \in \mathcal{M}_{m\times n}$.

- A basis for the nullspace $\text{null}(A)$ is the set of special solutions to $A\mathbf{x} = 0$.
- A basis for the column space $\text{col}(A)$ is the pivot columns of A - this is not necessarily all the columns of A.

Algorithm 5 (Find linearly independent vectors in a set): Given a set of vectors in $\mathbf{R}^n$, we can find which of them are linearly independent by either:

- making them columns of a matrix, doing elimination (with row swaps), and taking the positions of the pivot columns, or,
- making them rows of a matrix, doing elimination (without row swaps), and taking the positions of pivots rows.

Inquiry 10.9 (✠2.03): Consider the vectors $\mathbf{u}_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix}0\\1\\1\end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix}-1\\1\\1\end{bmatrix}, \mathbf{v} = \begin{bmatrix}1\\2\\4\end{bmatrix}$ in $\mathbf{R}^3$.

1. Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a linearly independent set.
2. Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{v}\}$ is a linearly dependent set.
3. Find two different pairs of numbers $a, b, c, d \in \mathbf{R}$ with $a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3 + d\mathbf{v} = 0$.

Example 10.10. Consider the following three vectors in $\mathbf{R}^4$:

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

As columns of a matrix, we quickly eliminate entries below the diagonal to identify the the first two as pivot columns and the last as a free column:

$$\begin{bmatrix}3 & 1 & 5\\2 & -1 & 5\\7 & 2 & 12\\1 & 3 & 5\end{bmatrix} \xrightarrow{G.E.} \begin{bmatrix}3 & 1 & 5\\0 & -5/3 & 13/3\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}.$$

So $\mathbf{u}, \mathbf{v}$ are indepdent, and $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$. Alternatively, we can make them rows of a matrix, and the perform Gaussian elimination (without row swaps). That will give us zero rows, which will correspond to linearly dependent vectors:

$$\begin{bmatrix}3 & 2 & 7 & 1\\1 & -1 & 2 & 3\\5 & 5 & 12 & 5\end{bmatrix} \xrightarrow{G.E.} \begin{bmatrix}3 & 2 & 7 & 1\\0 & -5/3 & -1/3 & 8/3\\0 & 0 & 0 & 0\end{bmatrix}.$$

As in the first approach, we get that $\mathbf{w}$ depends on $\mathbf{u}$ and $\mathbf{v}$. Hence $\{\mathbf{u}, \mathbf{v}\}$ is a basis for the vector space $V = \text{span}(\mathbf{u}, \mathbf{v}, \mathbf{w})$.

10.2 Dimension and extending to a basis

The key idea from the first part of this lecture is that the word *basis* is another name for *minimal spanning set*. It is often difficult to consider all possible spanning sets, so we use the word *basis* much more often. Keep in mind three important things:

- bases are not unique,
- every basis of a vector space must have the same number of vectors, and
- every vector space has a basis.

The last conclusion is based on a fundamental (and unproven!) cornerstone of mathematics called the *axiom of choice*. A special case is investigated in Inquiry 10.18.

Definition 10.11: Let V be a vector space. The *dimension* of V is the number of vectors in any basis of V. It is denoted $\dim(V)$.

Example 10.12. 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 is the *nullity* of A.

Inquiry 10.13 (✠2.05): For $\mathbf{v} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right] \in \mathbf{R}^2$, let V be the vector space of vectors in $\mathbf{R}^2$ perpendicular to $\mathbf{v}$.

1. Express V as a set, like $V = \{\ldots \mid \ldots\}$. That is, express V using set builder notation.
2. What is the dimension of V?
3. Express V in another way, as the setof all scalar multiples of a particular vector.

Recall the definition of $U \oplus V$ and $U + V$ from Definiton 6.9. 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 10.14. 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 10.15. Let V be a vector space and $U \subseteq V$. If $\dim(U) = \dim(V)$, then $U = V$. This follows by taking the basis $\mathbf{u}_1, \ldots, u_n$ 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 $\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{v}$ is a linearly independent set of $n+1$ vectors in V, which is impossible.

Definition 10.16: 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 10.17. 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.

Inquiry 10.18 (✠2.03, 2.04): Consider the vector space $\mathbf{R}^4$ with its four standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_4$, as given in Example 10.7. Consider $\mathbf{u} = \left[\begin{smallmatrix}1\\2\\3\\4\end{smallmatrix}\right], \mathbf{v} = \left[\begin{smallmatrix}0\\1\\-1\\2\end{smallmatrix}\right] \in \mathbf{R}^4$

1. Explain why $\{\mathbf{u}, \mathbf{v}\}$ cannot be a basis of $\mathbf{R}^4$. Is it linearly (in)dependent?
2. Explain why $S = \{\mathbf{u}, \mathbf{v}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ cannot be a basis of $\mathbf{R}^4$. Is it a linearly (in)dependent?
3. Find a linearly independent subset of $S = \{\mathbf{u}, \mathbf{v}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ that contains $\mathbf{u}$ and $\mathbf{v}$.

4. For the two vectors of S you did not use in the previous point, express them as a linear combination of the vectors you did use.

The third point is called *extending* a set to a basis. It is revisited again in Remark 15.12

10.3 The change of basis matrix

Example 10.19. As mentioned in Definition 10.6, given a basis for a vector space V, every vector in V can be expressed uniquely as a linear combination of vectors of that basis:

$$\begin{bmatrix}4\\-2\\8\end{bmatrix} = \underbrace{4\begin{bmatrix}1\\0\\0\end{bmatrix} - 2\begin{bmatrix}0\\1\\0\end{bmatrix} + 8\begin{bmatrix}0\\0\\1\end{bmatrix}}_{\text{in the standard basis of } \mathbf{R}^3} = \underbrace{14\begin{bmatrix}1\\0\\0\end{bmatrix} - 10\begin{bmatrix}1\\1\\0\end{bmatrix} + 8\begin{bmatrix}0\\1\\1\end{bmatrix}}_{\text{in the basis of Example 10.5}} = \underbrace{10\begin{bmatrix}1\\0\\1\end{bmatrix} - 8\begin{bmatrix}0\\1\\1\end{bmatrix} + 6\begin{bmatrix}-1\\1\\1\end{bmatrix}}_{\text{in the basis of Inquiry 10.9}}.$$

The coefficients for the basis vectors were found by solving matrix equations by row reduction:

- The solution to $\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\mathbf{x} = \begin{bmatrix}4\\-2\\8\end{bmatrix}$ is $\mathbf{x} = \begin{bmatrix}4\\-2\\8\end{bmatrix}$
- The solution to $\begin{bmatrix}1&1&0\\0&1&0\\0&1&1\end{bmatrix}\mathbf{x} = \begin{bmatrix}4\\-2\\8\end{bmatrix}$ is $\mathbf{x} = \begin{bmatrix}14\\-10\\8\end{bmatrix}$
- The solution to $\begin{bmatrix}1&0&-1\\0&1&1\\1&1&1\end{bmatrix}\mathbf{x} = \begin{bmatrix}4\\-2\\8\end{bmatrix}$ is $\mathbf{x} = \begin{bmatrix}10\\-8\\6\end{bmatrix}$

Knowing the important coefficients $14, -10, 8$ of the second basis, we can get the important coefficients $10, -8, 6$ of the third basis by multiplication:

$$\begin{bmatrix}0&-1&0\\1&3&1\\-1&-2&0\end{bmatrix}\begin{bmatrix}14\\-10\\8\end{bmatrix} = \begin{bmatrix}10\\-8\\6\end{bmatrix}.$$

We now learn how to construct this "change of basis" matrix, which lets us go from one basis to another.

Remark 10.20. Whenever we have a different basis, how can we figure out what the linear combination is in the other basis, without doing the same work all over again? This is where the *change of basis matrix* appears. Suppose that B and B' are bases for V, with

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

Given this $\mathbf{v} \in V$, the coefficients for expressing $\mathbf{v}$ in the basis B are in the solution vector $\mathbf{x}$, and the coefficients for expressing $\mathbf{v}$ in the basis B' are in the solution vector $\mathbf{y}$, for the equations

$$\begin{bmatrix}| & & |\\ \mathbf{u}_1 & \cdots & \mathbf{u}_k\\ | & & |\end{bmatrix}\mathbf{x} = \mathbf{v}, \qquad \begin{bmatrix}| & & |\\ \mathbf{w}_1 & \cdots & \mathbf{w}_k\\ | & & |\end{bmatrix}\mathbf{y} = \mathbf{v}.$$

With the same approach, we can solve for each vector $\mathbf{u}_i$ of the first basis B, and place the solutions

$\mathbf{a}_i$ into a new matrix:

$$\begin{bmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_k \\ | & & | \end{bmatrix} \mathbf{a}_i = \mathbf{u}_i \qquad \text{for all } i\text{, so then} \qquad \underbrace{\begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_k \\ | & & | \end{bmatrix}}_{\substack{\text{change of basis matrix} \\ \text{from } B \text{ to } B'}} \mathbf{x} = \mathbf{y}.$$

Example 10.21. Consider the two bases $B = \left\{ \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right] \right\}$ and $B' = \left\{ \left[\begin{smallmatrix} 3 \\ 7 \end{smallmatrix}\right], \left[\begin{smallmatrix} -2 \\ -4 \end{smallmatrix}\right] \right\}$ of $\mathbf{R}^2$. To construct the change of basis matrix from B to B', we need to express every vector of B as a linear combination of vectors in B'. We do this by sight:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = -\begin{bmatrix} 3 \\ 7 \end{bmatrix} - 2\begin{bmatrix} -2 \\ -4 \end{bmatrix}, \qquad \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -3\begin{bmatrix} 3 \\ 7 \end{bmatrix} - 5\begin{bmatrix} -2 \\ -4 \end{bmatrix},$$

Hence the change of basis matrix from vectors in the basis B to vectors in the basis B' is

$$A = \begin{bmatrix} -1 & -3 \\ -2 & -5 \end{bmatrix}.$$

For example, taking the vector

$$\begin{bmatrix} -1 \\ -13 \end{bmatrix} = -7\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 6\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

in the basis B, we have the coefficient vector $\mathbf{x} = \left[\begin{smallmatrix} -7 \\ 6 \end{smallmatrix}\right]$. The coefficient vector $\mathbf{y}$ in the basis B' is given by computing

$$\mathbf{y} = \begin{bmatrix} -1 & -3 \\ -2 & -5 \end{bmatrix}\begin{bmatrix} -7 \\ 6 \end{bmatrix} = \begin{bmatrix} 7-18 \\ 14-30 \end{bmatrix} = \begin{bmatrix} -11 \\ -16 \end{bmatrix}, \qquad \text{meaning} \qquad \begin{bmatrix} -1 \\ -13 \end{bmatrix} = -11\begin{bmatrix} 3 \\ 7 \end{bmatrix} - 16\begin{bmatrix} -2 \\ -4 \end{bmatrix}.$$

Note that the inverse of A will take us back to coefficients in the basis B.

10.4 Exercises

Exercise 10.1. (✠2.03) Find all sets of size 3 from the vectors below that are linearly independent:

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

Exercise 10.2. (✠2.05) For a 2×2 matrix, linear independence on the columns only depends on if one column is a multiple of the other.

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

⋈ (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 10.3. (✠2.04) 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 10.4. (✠2.04) 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 10.5. (✠2.04) This question is about expressing vectors in different bases.

1. Express the vector $\begin{bmatrix} 3 \\ -2 \\ -8 \end{bmatrix}$ in the basis $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

2. There are two bases $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ and $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ of a vector space V, with the following relations:

$$\mathbf{b}_1 = \mathbf{a}_1 + \mathbf{a}_2, \qquad \mathbf{b}_2 = \mathbf{a}_2 + \mathbf{a}_3, \qquad \mathbf{b}_3 = \mathbf{a}_1 + \mathbf{a}_3.$$

   If you know that $\mathbf{v} = 3\mathbf{a}_1 - 2\mathbf{a}_2 - 8\mathbf{a}_3$, express $\mathbf{v}$ as a linear combination of $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$.

Exercise 10.6. (✠2.05) Prove the claims from Remark 10.14.

Exercise 10.7. (✠2.05) This question is about vector spaces of matrices, where matrix addition and scalar multiplication are defined as usual.

1. Give a basis for the space of diagonal 3×3 matrices and a basis for the space of skew-symmetric 3×3 matrices.

2. For $n \in \mathbf{N}$, what is the dimension of the space of $n \times n$ diagonal matrices and what is the dimension of the space of $n \times n$ skew-symmetric matrices?

3. Show by example that the set of all invertible 2×2 matrices does not form a vector space. Show that all linear combinations of invertible 2×2 matrices describe the set $\mathcal{M}_{2\times 2}$.
   Hint: Construct the basis matrices of $\mathcal{M}_{2\times 2}$ as linear combinatons of invertible matrices.

Lecture 11: The four fundamental subspaces associated to a matrix

Chapter 3.5 in Strang's "Linear Algebra"

- Fact 1: A line in $\mathbf{R}^2$ is given by one equation, in $\mathbf{R}^3$ by two equations.
- Fact 2: Every matrix with m rows splits up $\mathbf{R}^m$ into the column space and the left nullspace.
- Fact 3: Every matrix with n columns splits up $\mathbf{R}^n$ into the row space and the nullspace.

✠ Standard 2.06: Find the intersection of two planes.

✠ Standard 2.07: Describe a hyperplane as a span of vectors.

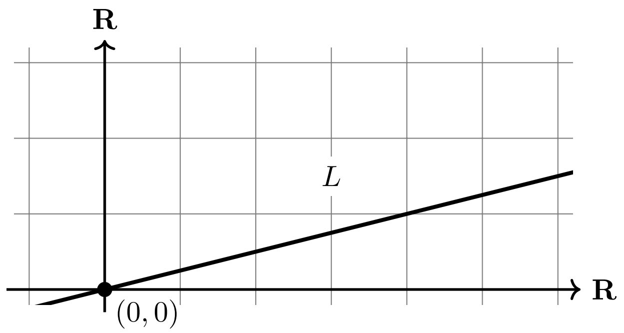
✠ Standard 2.08: Find the bases of the four fundamental subspaces of a matrix.

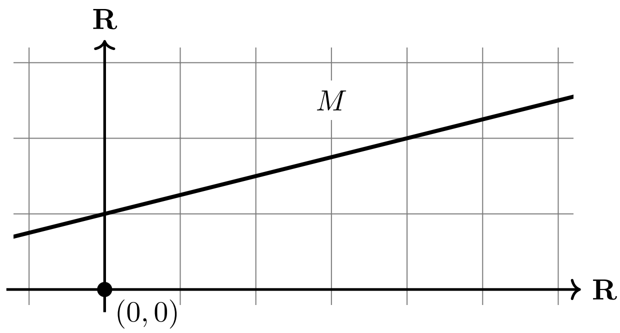
With this lecture we take the column space and nullspace to the transpose matrix, and describe strong relationships among these spaces.

11.1 Lines, planes, and hyperplanes

Since we will be discussing spaces and their relationships with each other in this lecture, we begin with a comparison relating two similar lines.

Example 11.1. Consider the two lines $L, M \subseteq \mathbf{R}^2$ given below.





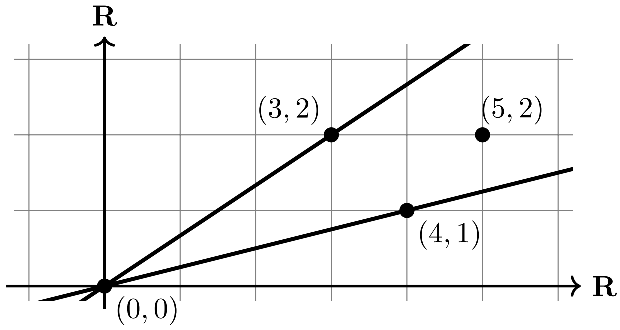
Each of these lines can be considered in similar ways. Each is:

- the line $y = x/4$
- the pairs (x, y) that satisfy $\frac{1}{4}x - y = 0$
- the nullspace of $\begin{bmatrix} 1 & -4 \end{bmatrix}$
- the set of vectors $\mathbf{x}$ for which $\begin{bmatrix} 1 & -4 \end{bmatrix}\mathbf{x} = 0$
- a vector subspace of $\mathbf{R}^2$
- a vector space of dimension 1

- the line $y = x/4 + 1$
- the pairs (x, y) that satisfy $\frac{1}{4}x - y = -1$
- not the nullspace of any matrix
- the set of vectors $\mathbf{x}$ for which $\begin{bmatrix} 1 & -4 \end{bmatrix}\mathbf{x} = -4$
- not a vector subspace of $\mathbf{R}^2$
- an affine space of dimension 1

The line M can be considered as a vector space, using a different addition and multiplication than in $\mathbf{R}^2$. This is the same *affine space* structure seen before in Inquiry 6.3 and Example 8.9.

Note that the vectors in L do not span all of $\mathbf{R}^2$, but if we add another line at a different angle than L, also going through the origin, vectors from both lines together will span $\mathbf{R}^2$.

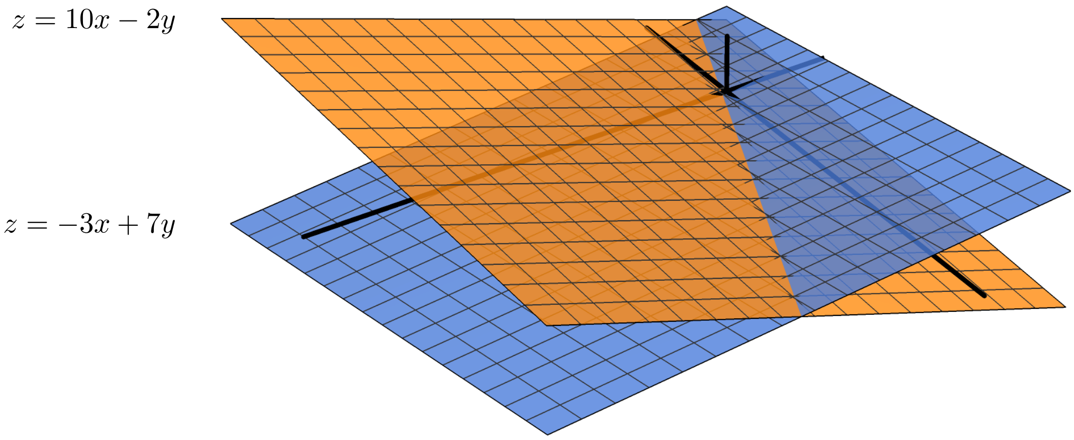


one line: neither of the two lines go through $\left[\begin{smallmatrix}5\\2\end{smallmatrix}\right]$

two lines: there is a unique solution to $\left[\begin{smallmatrix}4&3\\1&2\end{smallmatrix}\right]\left[\begin{smallmatrix}a\\b\end{smallmatrix}\right] = \left[\begin{smallmatrix}5\\2\end{smallmatrix}\right]$

The reason for going through all the different perspectives in Example 11.1 was to connect visual with algebraic intution. We go one step further, into the third dimension, with the following example.

Example 11.2. Consider the two planes in $\mathbf{R}^3$ given below. Note that their intersection is a line.



Both planes go through the origin $(0, 0, 0)$. To find the vector along the line of intersection, we need both equations to be satisfied at the same time. That is, we want to solve the matrix equation

$$\begin{bmatrix}10 & -2 & 1\\ -3 & 7 & 1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}. \qquad \text{Notice } \begin{bmatrix}10 & -2 & 1\\ -3 & 7 & 1\end{bmatrix} \xrightarrow{RREF} \begin{bmatrix}1 & 0 & 9/64\\ 0 & 1 & 13/64\end{bmatrix},$$

so the nullspace of the matrix on the left is the span of the single vector $\left[\begin{smallmatrix}-9\\-13\\64\end{smallmatrix}\right]$. Unlike in $\mathbf{R}^2$, a line in R^3 can not be described by a single equation. We either use two equations (of the two planes), or a single vector. Finally, we observe that to describe a plane as a span of vectors, we also use the nullspace:

$$\text{null}\left(\begin{bmatrix}10 & -2 & 1\end{bmatrix}\right) = \text{null}\left(\begin{bmatrix}1 & -1/5 & 1/10\end{bmatrix}\right) = \text{span}\left(\begin{bmatrix}1/5\\1\\0\end{bmatrix}, \begin{bmatrix}-1/10\\0\\1\end{bmatrix}\right).$$

These are the two vectors whose span is the plane $z = 10x - 2y$.

Inquiry 11.3 (✠2.06): Choose six different nonzero integers a, b, c, d, e, f. These describe two planes $ax + by + cz = 0$ and $dx + ey + fz = 0$ in $\mathbf{R}^3$.

1. Describe each of the two planes as a span of two vectors each. Can you make the vectors

only have integer entries?

2. Find the intersection of these two planes. Describe it as the span of one vector. Can you make the vector only have integer entries?

Definition 11.4: A *hyperplane* in $\mathbf{R}^n$ is the set of points that satisfies a single equation $a_1x_1 + \cdots + a_nx_n = 0$.

- For $n = 1$, a hyperplane in $\mathbf{R}^1$ is a *point*.
- For $n = 2$, a hyperplane in $\mathbf{R}^2$ is a *line*.
- For $n = 3$, a hyperplane in $\mathbf{R}^3$ is a *plane*.

A hyperplane is an $(n-1)$-dimensional (or codimension 1) subspace of $\mathbf{R}^n$.

The intersection of two planes in $\mathbf{R}^3$ is (almost always) a line, and the intersection of three planes in $\mathbf{R}^3$ is (almost always) a point.

Inquiry 11.5 (✠2.06): Consider the vector $\mathbf{v} = \left[\begin{smallmatrix} 1 \\ 2 \\ -3 \end{smallmatrix}\right] \in \mathbf{R}^3$.

1. Construct a 2×3 matrix A so that $\mathbf{v}$ is in the nullspace of A. That is, create an A so that $A\mathbf{v} = 0$. Try to make something other than the zero matrix!
2. Create a vector $\mathbf{w}$ that is perpendicular to $\mathbf{v}$ (that is, $\mathbf{w} \bullet \mathbf{v} = 0$). Is $\mathbf{w} \in \text{null}(A)$?
3. Find two planes in $\mathbf{R}^3$ so that their intersection is $\mathbf{v}$.
4. Construct a 3×3 matrix B so that $\text{col}(B) = \text{span}(\mathbf{v})$.
5. What is the nullspace of B?

11.2 The four fundamental subspaces

Let $A \in \mathcal{M}_{m\times n}$, and let $R \in \mathcal{M}_{m\times n}$ be the result of applying Guassian and Gauss–Jordan elimination to A. We have seen two related vector spaces:

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

We now introduce two other spaces, which are related to the above two by the *transpose* of A.

Definition 11.6: Let $A \in \mathcal{M}_{m\times n}$.

- The *row space*, denoted $\text{row}(A)$, is the span of the rows of A.
- The *left nullspace* is the span of the solutions to $\mathbf{x}^TA = 0$.

Together these four vector spaces are the *four fundamental subspaces*.

Remark 11.7. The left nullspace has no special way to write it. Observing that $(\mathbf{x}^TA)^T = A^T\mathbf{x}$, we see that the left nullspace of A is the vector space $\text{null}(A^T)$. With this, we see several other relations among the four fundamental spaces:

$$\text{row}(A) = \text{col}(A^T), \qquad \text{row}(A^T) = \text{col}(A), \qquad \text{null}(A) = \begin{pmatrix} \text{left null-} \\ \text{space of } A^T \end{pmatrix}, \qquad \text{null}(A^T) = \begin{pmatrix} \text{left null-} \\ \text{space of } A \end{pmatrix}.$$

Remark 11.8. The previous remark makes it clear that the row space and left nullspace are vector spaces. Below we put together all the relationships among these four subspaces, for $A \in \mathcal{M}_{m\times n}$.

1. *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. *dimension* relations:
   - $\dim(\text{col}(A)) = \dim(\text{col}(A^T)) = \text{rank}(A) = \text{rank}(A^T)$
   - $\dim(\text{col}(A)) + \dim(\text{null}(A^T)) = m$
   - $\dim(\text{col}(A^T)) + \dim(\text{null}(A)) = n$
3. *sum* relations:
   - $\text{col}(A) + \text{null}(A^T) = \mathbf{R}^m$
   - $\text{col}(A^T) + \text{null}(A) = \mathbf{R}^n$

The last statement in the second point of Remark 11.8 is called the *rank-nullity theorem*, which we already saw just after Definition 8.6. We now look at more relations among these vector spaces.

Inquiry 11.9 (✠2.08): Consider the matrix $A = \left[\begin{smallmatrix} 2 & 0 & -1 \\ 1 & -2 & 1 \end{smallmatrix}\right]$.

1. Describe the column space of A as the span of two nonzero vectors.
2. Suppose your answer to the above was $\text{col}(A) = \text{span}(\mathbf{u}, \mathbf{v})$. Compute $A^T\mathbf{u}$ and $A^T\mathbf{v}$. Explain why, in general, if $\mathbf{v} \in \text{col}(A)$ is non zero, then $A^T\mathbf{v} \neq 0$.
3. Describe the left nullspace of A. Why does it only contain the zero vector?
4. Construct a 2×3 matrix whose column space is 1-dimensional and whose left nullspace is 1-dimensional.

The statements of the third point in Remark 11.8 claim that *any* vector in $\mathbf{R}^n$ is (for the first statement) a linear combination of the vectors in the column space of A and the left nullspace of A.

Inquiry 11.10 (✠2.08): Let $A = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{smallmatrix}\right]$.

1. Describe the column space of A and the left nullspace of A as a span of vectors. These vectors should all be in $\mathbf{R}^3$.
2. Are the three vectors you found in part 1. linearly indepedent?
3. Are the three standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ in $\text{col}(A)$? Are they in $\text{null}(A^T)$? Explain how to get $\mathbf{e}_1$ from the three vectors you found in part 1.
4. **Bonus:** Find the change of basis matrix from the three vectors of part 2. to the three standard basis vectors of $\mathbf{R}^3$.

Remark 11.11. 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

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

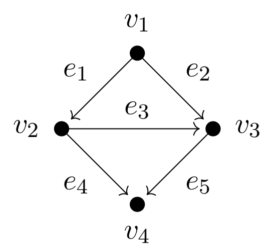
and so every vector in $\text{col}(A)$ is perpendicular to every vector in $\text{null}(A^T)$.

Inquiry 11.12 (✠2.08): Consider the vector $\mathbf{v} = \left[\begin{smallmatrix} 1 \\ 0 \\ -1 \\ 0 \end{smallmatrix}\right] \in \mathbf{R}^4$, and let $A \in \mathcal{M}_{4\times 4}$.

1. Suppose that $\left[\begin{smallmatrix} 1 \\ 0 \\ 0 \\ 0 \end{smallmatrix}\right] \in \text{col}(A)$ and $\left[\begin{smallmatrix} 0 \\ 0 \\ -1 \\ 0 \end{smallmatrix}\right] \in \text{null}(A^T)$. Suppose that $\mathbf{v} \in \text{col}(A)$. Explain why this also means that $\left[\begin{smallmatrix} 0 \\ 0 \\ -1 \\ 0 \end{smallmatrix}\right] \in \text{col}(A)$ as well.
2. Find two vectors in $\mathbf{R}^4$ that are perpendicular to $\mathbf{v}$. Explain why this gives you a 4×2 matrix that contains $\mathbf{v}$ in its left nullspace.

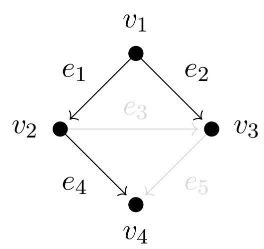
Example 11.13. For a practical application of these spaces, consider the following two matrices, both representations of the directed graph below. In A_{inc}, the rows correspond to edges, and the columns correspond to vertices: each row has a -1 for the vertex where the edge starts and a 1 for the vertex where the edge ends. This is called an *incidence matrix*. In A_{adj}, (i, j)-entry is 1 if there is a directed edge from v_i to v_j, and 0 otherwise. This is called the *adjacency matrix*.

$$A_{inc} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \qquad A_{adj} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$



Bringing the matrix $(A_{inc})^T$ to row reduced echelon form gives information about the row space of A_{inc} and the left nullspace of A_{inc}. The linearly independent rows of A_{inc} 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 of A_{inc}, 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 of A_{inc}, corresponding to edge e_5, because then we have a cycle of four edges.

$$A_{inc}^T = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \xrightarrow{G.E.} \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$



11.3 Exercises

Exercise 11.1. (✠2.07) Consider the plane $P = \{(x, y, z) \in \mathbf{R}^3 \ : \ 2x - 4y - 5z = 0\}$, which is a subspace of $\mathbf{R}^3$.

1. Find a vector $\mathbf{n}$ normal to the plane P. That is, find $\mathbf{n} \in \mathbf{R}^3$ so that $\mathbf{n} \bullet \mathbf{v} = 0$, for $\mathbf{v} = \left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]$ a solution to $2x - 4y - 5z = 0$.
2. Considering the vector $\mathbf{n}$ as a 1×3 matrix A, the nullspace of A is precisely all points in the plane P. Find this nullspace, and express it as a span.
3. What is a basis for P?

Exercise 11.2. (✠2.08) For $a, b, c \in \mathbf{R}$, consider the matrix

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

1. Find a basis for the column space, nullspace, row space, and left nullspace of A.
2. Do the dimensions of the four fundamental spaces change if all of a, b, c are zero?

Exercise 11.3. (✠2.08) Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Describe the four fundamental subspaces of A, $A + I$, and $A + A^2$.

Exercise 11.4. (✠2.08) Let $\mathbf{u} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ and $\mathbf{v} = \left[\begin{smallmatrix} 3 \\ 4 \end{smallmatrix}\right]$.

1. Construct a 2×4 matrix A for which $\text{col}(A) = \text{span}(\{\mathbf{u}, \mathbf{v}\})$.
2. Find a basis for the column space and row space of $\mathbf{u}\mathbf{v}^T + (\mathbf{u}\mathbf{v}^T)^2$.

Exercise 11.5. (✠2.07, 2.08) Consider the following two planes, as subspaces of $\mathbf{R}^3$:

$$P_1 = \{\mathbf{x} = (x_1, x_2, x_3) \in \mathbf{R}^3 \ : \ 3x_1 - 4x_2 + x_3 = 0\},$$
$$P_2 = \{\mathbf{x} = (x_1, x_2, x_3) \in \mathbf{R}^3 \ : \ 5x_1 - 10x_3 = 0\}.$$

1. Find normal vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ to the planes P_1 and P_2, respectively.
2. Find bases B_1 and B_2 for the planes P_1 and P_2, respectively.
   Hint: the basis of a plane is the nullspace of the defining equation.
3. Construct a 2×3 matrix A_1 whose row space is P_1. Show that the nullspace of A_1 is the span of $\mathbf{n}_1$.
4. Construct a 3×2 matrix A_2 whose column space is P_2. Show that the left nullspace of A_2 is the span of $\mathbf{n}_2$.

Lecture 12: Orthogonality

Chapter 4.1 in Strang's "Linear Algebra"

- Fact 1: Two orthogonal subspaces are orthogonal complements if their dimensions sum up to the dimension of the space they are in.
- Fact 2: The column space is the orthogonal complement to the left nullspace, and the nullspace is the orthogonal complement to the row space.

✠ Standard 2.09: Determine if the columns of a matrix are orthogonal.

✠ Standard 2.10: Determine if two subspaces are orthogonal.

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.

12.1 Orthogonal spaces

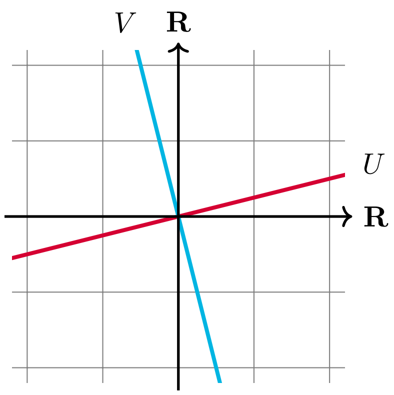
Recall from Lecture 1 that two vectors $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$ are *orthogonal* if $\mathbf{u} \bullet \mathbf{v} = \mathbf{u}^T\mathbf{v} = 0$. Note that in this case we have something that looks like the Pythagorean theorem:

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

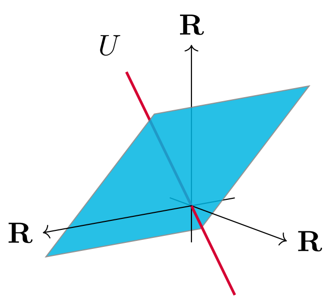
If $\mathbf{u}, \mathbf{v}$ are orthogonal and both have length 1, then they are called *orthonormal*

Definition 12.1: 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. We say that "U is orthogonal to V" and "V is orthogonal to U", which both mean the same thing.

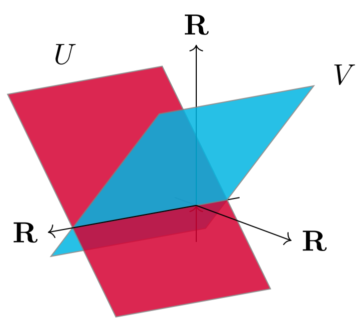
Example 12.2. Consider the following subspaces of Euclidean space.



lines at 90° to each other in $\mathbf{R}^2$ are orthogonal



a line coming out of a plane in $\mathbf{R}^3$ at 90° is orthogonal to the plane



two planes in $\mathbf{R}^3$ that "intersect at 90°" are not orthogonal

Two planes "intersecting at 90°" does not make sense, because any angle can be found between the two planes with vectors in them. The planes intersect in a 1-dimensional vector subspace (the x-axis), and the inner product of $[\,1\ 0\ 0\,]^T$ with itself is not zero.

Inquiry 12.3 (✠2.10): In Example 12.2 above, the third example with two planes looks like it "should" describe a perpendicular intersection.

1. Construct bases for U and for V of two vectors each. Make it so that the bases have a common vector.
2. Take the symmetric difference of the two basis sets. What is the angle between the two vectors?
3. Give a proper description of what the "perpendicular feeling" in the picture is, using bases.

Example 12.4. 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

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

and since the row space $\text{row}(A) = \text{span}(\{\mathbf{r}_1, \dots, \mathbf{r}_m\})$, we see that $\mathbf{v} \bullet \mathbf{x} = 0$ 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).

Inquiry 12.5 (✠2.10): This inquiry uses Python, and follows the Python notebook on the course website.

1. Generate 100 real-valued vectors $\mathbf{R}^2$, with entries in the range $[0, 5]$. How many pairs are orthogonal? How many have inner product very close to zero?
2. Generate 100 integer-valued vectors $\mathbf{R}^2$, with entries in $\{0, 1, 2, 3, 4, 5\}$. How many pairs are orthogonal? How many would you expect to be orthogonal?
3. Generalize the previous point to $\mathbf{R}^3$.

Remark 12.6. 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 12.7. 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

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

Example 12.8. Consider a matrix $A \in \mathcal{M}_{3\times 6}$ as below. It 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:

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

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

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.

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

12.2 Orthogonal relationships

Definition 12.9: 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 12.10. Recall the concept of *codimension* from Definition 10.16. The codimension of a space is equal to the dimension of its orthogonal complement. That is, $\operatorname{codim}(U) = \dim(U^\perp)$.

Remark 12.11. 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 12.11.1. *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

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

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

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

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 $(U+V)^\perp = U^\perp \cap V^\perp$. $\square$

Example 12.12. Combining Example 12.4 and the rank-nullity theorem from Lecture 11, 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 12.11, 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 follows that no row of A can be in the nullspace of A.

Inquiry 12.13 (✠2.10): Let V be a vector space and $U, W \subseteq V$ subpaces, with $U = W^\perp$.

1. If $\dim(V) = n$ and $\dim(U) = \dim(W) = k$, explain what k must be, in terms of n.
2. You are given that $U = \text{span}(\mathbf{u}_1, \ldots, \mathbf{u}_i)$ and $W = \text{span}(\mathbf{w}_1, \ldots, \mathbf{w}_j)$. Explain the relationship between i, j, k.

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

Remark 12.14. 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.

12.3 Exercises

Exercise 12.1. (✠2.09) Let $A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ be a symmetric 4×4 matrix.

1. Find which pairs of columns of A are orthogonal to each other.
2. Give the nullspace of A as a span of the special solutions to $A\mathbf{x} = 0$.
3. Show that the column space of A is orthogonal to the nullspace of A.
4. Explain why for any symmetric matrix (not just the one given), its column space is orthogonal to its nullspace.

Exercise 12.2. (✠2.10) Confirm the observation from Remark 12.6. That is, let $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a basis for a vector space $U \subseteq \mathbf{R}^n$, and let $B' = \{\mathbf{v}_1, \ldots, \mathbf{v}_\ell\}$ 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 $\mathbf{u} \bullet \mathbf{v} = 0$ for arbitrary elements $\mathbf{u} \in U$ and $\mathbf{v} \in V$.

Exercise 12.3. (✠1.02) Check that the claim about the angle between $\mathbf{v}$ and $R_\theta \mathbf{v}$ from Example 12.7 is indeed true.

Exercise 12.4. (✠2.10) Let $A \in \mathcal{M}_{m\times n}$. Show that there is a bijective function $f\colon \mathrm{row}(A) \to \mathrm{col}(A)$.
Hint: use orthogonality and the decomposition of vectors described in Example 9.13.

Exercise 12.5. (✠2.10) Let U, V be subspaces of $\mathbf{R}^n$.

1. Show that $(U^\perp)^\perp = U$.
2. Show that $(U \cap V)^\perp = U^\perp + V^\perp$.
3. Suppose there exist matrices A, B with $U = \mathrm{col}(A)$ and $V = \mathrm{col}(B)$. Find a matrix C for which $\mathrm{null}(C) = (U + V)^\perp$.
   Hint: construct C as a block matrix.

Exercise 12.6. (✠2.10) Let $A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 2 \\ -1 & -2 \\ 1 & 3 \end{bmatrix}$.

1. What are the dimensions of $\mathrm{col}(A)$ and $\mathrm{col}(B)$? Only using dimensions, explain why $\mathrm{col}(A) \neq \mathrm{col}(B)^\perp$.
2. Find a vector that is both in $\mathrm{col}(A)$ and $\mathrm{col}(B)$.
   Hint: Row reduce the block matrix $\begin{bmatrix} A & B \end{bmatrix}$.

Lecture 13: Projections

Chapter 4.2 in Strang's "Linear Algebra"

- Fact 1: A projection of a vector is always a "projection to" somewhere or "projection onto" something.
- Fact 2: A projection of a vector is another vector.
- Fact 3: Projections are often used in problems that require the "best approximations" of something.

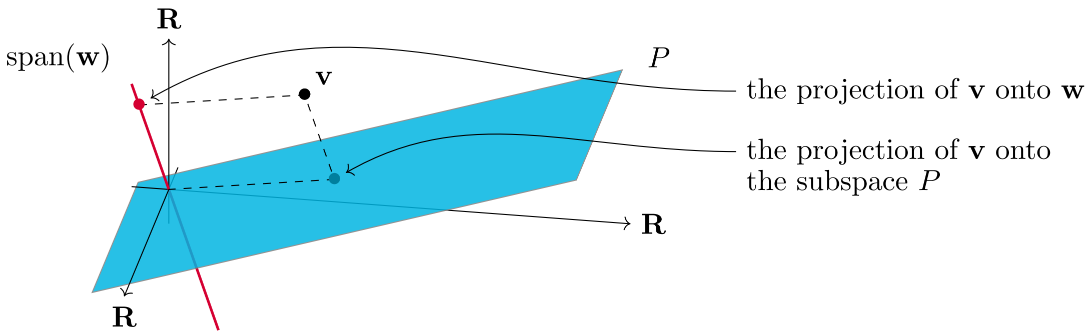
✠ Standard 2.11: Compute the projection of a vector onto another vector.

✠ Standard 2.12: Compute the projection of a vector onto a subspace.

We continue our study of orthogonality by describing how it affects arbitrary vectors, not just ones in the vector subspaces being considered.

13.1 Projecting onto lines

To *project* a vector $\mathbf{v} \in \mathbf{R}^3$ onto some other vector $\mathbf{w} \in \mathbf{R}^3$ (or onto some plane P going through the origin), means to create a new vector that points in the same direction as $\mathbf{w}$ (or lies in P), and is "as close as possible" to the first vector $\mathbf{v}$.



Since both $\mathbf{w}$ and P are subspaces of $\mathbf{R}^3$, projections can be understood in (at least) two ways:

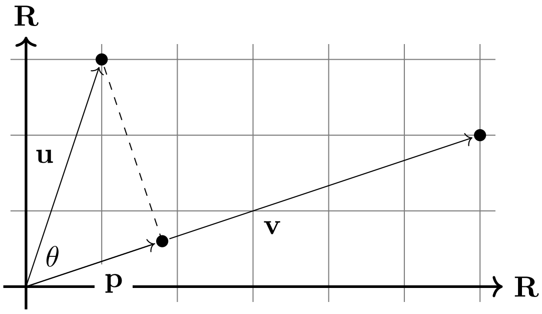
1. the projection of $\mathbf{v}$ is the part of $\mathbf{v}$ that lies in the subspace to which you are projecting
2. the projection of $\mathbf{v}$ produces another vector $\mathbf{v}'$, so projecting is simply multiplying by some appropriate matrix A: $A\mathbf{v} = \mathbf{v}'$

Both of these approaches are correct.

Example 13.1. The act of projecting is often done by a matrix:

- 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 1.19. 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.



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

Since the hypotenuse has length $\|\mathbf{u}\|$, the adjacent, which is $\mathbf{p}$, must have length $\frac{\mathbf{u}\bullet\mathbf{v}}{\|\mathbf{v}\|}$. The vector $\mathbf{v}$ may not have unit length, 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

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

Definition 13.2: The *projection* of $\mathbf{u}$ onto $\mathbf{v}$ is the vector

$$\text{proj}_{\mathbf{v}}(\mathbf{u}) = \frac{\mathbf{u} \bullet \mathbf{v}}{\mathbf{v} \bullet \mathbf{v}}\mathbf{v}. \tag{3}$$

The difference $\mathbf{u} - \text{proj}_{\mathbf{v}}(\mathbf{u})$ is called the *error vector*.

Example 13.3. 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 $\mathbf{u} \bullet \mathbf{v} = 0$ 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 $\frac{\mathbf{u}\bullet\mathbf{v}}{\mathbf{v}\bullet\mathbf{v}}$ 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 (3) becomes

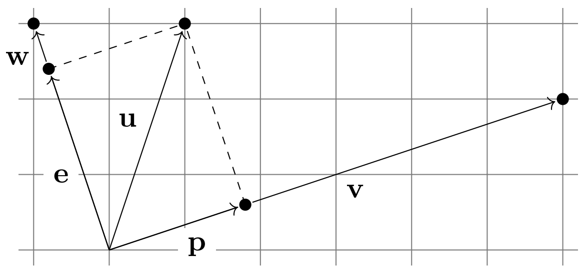
$$\frac{\mathbf{u}^T\mathbf{v}}{\mathbf{v} \bullet \mathbf{v}}\mathbf{v} = \frac{\mathbf{v}^T\mathbf{u}}{\mathbf{v} \bullet \mathbf{v}}\mathbf{v} = \mathbf{v}\frac{\mathbf{v}^T\mathbf{u}}{\mathbf{v} \bullet \mathbf{v}} = \frac{\mathbf{v}\mathbf{v}^T\mathbf{u}}{\mathbf{v} \bullet \mathbf{v}} = \underbrace{\frac{1}{\mathbf{v} \bullet \mathbf{v}}\mathbf{v}\mathbf{v}^T}_{P}\mathbf{u}. \tag{4}$$

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.

Inquiry 13.4 (✠2.11): This inquiry continues the ideas from Example 13.3 above

1. Explain what properties of scalar, vector, or matrix operations are being used for each equality in Equation (4).
2. Explain why the projection matrix is always rank 0 or rank 1, but never rank 2.
3. Let $P = \left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right] \in \mathcal{M}_{2\times 2}$ be a symmetric matrix. Using Equation (4), explain what conditions must be true for $a, b, c \in \mathbf{R}$ for P to represent a projection matrix.

Remark 13.5. The error vector $\mathbf{e} = \mathbf{u} - \mathbf{p}$ from Definition 13.2 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 $I - P$.

13.2 Projecting onto subspaces

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 13.6: 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 13.7. Since $V^\perp$ is the orthogonal complement of V, by Remark 12.11, 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 13.3, 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

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

Inquiry 13.8 (✠2.12): Let $V = \mathbf{R}^2$, choose two perpendicular vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{R}^2$. Let $\mathbf{w} \in \mathbf{R}^2$ be a non-trivial linear combination of both of the vectors.

1. Compute the two 2×2 projection matrices $P_1 = \text{proj}_{\text{span}(\mathbf{v}_1)}$ and $P_2 = \text{proj}_{\text{span}(\mathbf{v}_2)}$.
2. Explain why $P_1P_2\mathbf{w} = P_2P_1\mathbf{w} = 0$. Is $P_1P_2 = P_2P_1 = 0$?
3. Explain why $P_1P_1\mathbf{w} = P_1\mathbf{w}$ and $P_2P_2\mathbf{w} = P_2\mathbf{w}$. Is $P_1P_1 = P_2P_2 = I$?

See Exercise 13.1 for more guidance.

13.3 Exercises

Exercise 13.1. (✠2.11) This question is about repeated projections.

1. Show that projecting twice onto a line is the same as projecting once.

2. Show that projecting twice onto a subspace is the same as projecting once.

Hint: Use the projection matrices P from Equation (4) and Definition 13.6, and show that $P^2 = P$.

3. Let $R_\theta \in \mathcal{M}_{2\times 2}$ be the rotation matrix from Example 12.7. For which $\theta \in [0, 2\pi)$ is R_θ a projection matrix? Justify your answer.

Exercise 13.2. (✠2.11) 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 13.3. (✠2.11, 2.12) Find the projection of $\mathbf{v} = (-3, -1, 6)$ onto the plane $3x+4y-9z = 0$ and its normal vector.

Exercise 13.4. (✠2.12) Let $\mathbf{v} = (x, y, z, w)$.

1. What matrix M projects $\mathbf{v}$ onto the xy-plane to produce $(x, y, 0, 0)$? That is, find M for $M\mathbf{v} = (x, y, 0, 0)$.

2. What matrix N cycles the axes to produce (w, x, y, z)? That is, find N for $N\mathbf{v} = (w, x, y, z)$.

3. Explain why N is not a projection matrix.

Exercise 13.5. (✠2.12) The set $U \subseteq \mathbf{R}^n$ is a subspace with basis $\mathbf{u}_1, \ldots, \mathbf{u}_k$. These basis vectors are the columns of the $n \times k$ matrix A. For any $\mathbf{v} \in \mathbf{R}^n$, define the *reflection* of $\mathbf{v}$ in U to be the vector

$$\text{refl}_U(\mathbf{v}) := \mathbf{v} - 2\text{proj}_{U^\perp}(\mathbf{v}).$$

1. Construct the matrix of refl_U.

2. Show that refl_U preserves length, that is, show that $\|\text{refl}_U(\mathbf{v})\| = \|\mathbf{v}\|$ for all $\mathbf{v} \in \mathbf{R}^n$.

Exercise 13.6. (✠2.12) Let $A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 1 & -1 \\ 0 & -1 & 1 & 0 \\ 2 & 1 & 1 & -2 \\ -1 & 0 & -1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 0 \\ 2 & 1 \\ 1 & -1 \\ 0 & -1 \\ 0 & 2 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$.

1. Compute the projection of $\mathbf{v}$ onto $\text{col}(A)$ and $\text{col}(A)^\perp$. What is the angle between the two projections?

2. Compute the projection of $\text{col}(B)$ onto $\text{col}(A)$.

Lecture 14: The least squares approximation

Chapter 4.3 in Strang's "Linear Algebra"

- Fact 1: The least squares approximation is a vector $\hat{\mathbf{x}}$ that is "the closest solution to" $A\mathbf{x} = \mathbf{b}$ when $\mathbf{b} \notin \text{col}(A)$
- Fact 2: If $A\mathbf{x} = \mathbf{b}$ does not have a solution, then $A^T A\mathbf{x} = \mathbf{b}$ will have a solution, as long as the rows of A are linearly independent.

✠ Standard 2.13: Find the least squares solution to a matrix equation.

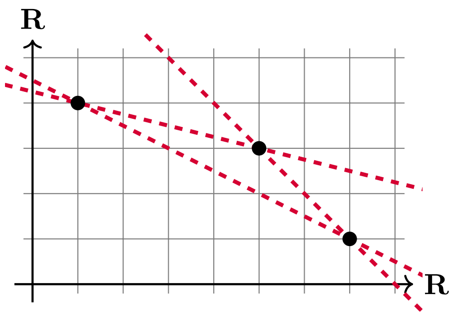
✠ Standard 2.14: Find the degree-d polynomial that approximates a collection of points in $\mathbf{R}^2$.

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

14.1 Least squares for lines

When given points in the plane $\mathbf{R}^2$, it is often assumed there is some underlying relationship among the points. To discover this relationship from the points, some approximation must be made, becaused the points are never arramnged in a neat pattern.

Example 14.1. Consider the points $(1, 4)$, $(7, 1)$, $(5, 3) \in \mathbf{R}^2$. Is there a line $y = ax + b$ goes through all of them? If yes, which one is it? If no, why?



There is no such line, because any two of the points determine a line that does not intersect the third point. We are equivalently asking for a solution to three equations, or to a linear system.

$$\begin{aligned} 4 &= a + b \\ 1 &= 7a + b \\ 3 &= 5a + b \end{aligned} \qquad \underbrace{\begin{bmatrix} 1 & 1 \\ 7 & 1 \\ 5 & 1 \end{bmatrix}}_{A} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} \qquad \left[\begin{array}{cc|c} 1 & 1 & 4 \\ 7 & 1 & 1 \\ 5 & 1 & 3 \end{array}\right] \xrightarrow{G.E.} \left[\begin{array}{cc|c} 1 & 1 & 4 \\ 0 & -6 & -27 \\ 0 & 0 & 1 \end{array}\right]$$

Note that $[\,4\ 1\ 3\,]^T$ is not in the column space of the matrix A, since the auigmented matrix by Gaussian elimination gives the contradictory equation $0 = 1$ in the last row. However, we still want to find a line that is "as close as possible", and projections help us do that.

Remark 14.2. 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 14.3: 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
- is the solution to $A\mathbf{x} = \mathbf{p}$

In practice, we minimize $\|A\mathbf{x} - \mathbf{b}\|^2$ instead of $\|A\mathbf{x} - \mathbf{b}\|$, since square roots are hard to deal with. It does not matter which expresion we minimize, 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 14.4. Using the equation $A\mathbf{x} = \mathbf{b}$ fom Example 14.1, we have

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

which simplifies to

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

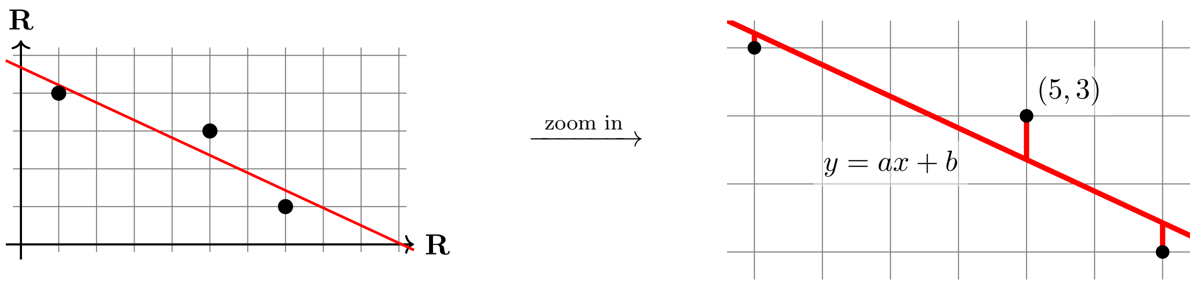
To find its minimum, we take the derivative. Since this is a function in two variables, we have two derivatives to take.

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

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

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

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

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

as small as possible, which is exactly the third term in $M(a, b)$ from Equation (5).

Inquiry 14.5 (✠2.13): Suppose you are given three points above are above each other: $(1, 1)$, $(1, 3)$, and $(1, 4)$.

1. Redo Example 14.4 with these points to see where the example fails.
2. What do you think is the "best" line that goes through these points?
3. Explain why there is no problem if a fourth point is added at a different x-value.

Remark 14.6. 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 24.

The second approach is to observe that for $\mathbf{b} = \mathbf{p} + \mathbf{e}$, the error vector $\mathbf{e}$ is in the left nullspace of A, since the column space and left nullspace are orthogonal complements.

Theorem 14.6.1. *Let $A \in \mathcal{M}_{m\times n}$. If $\boldsymbol{b} \notin \operatorname{col}(A)$, then*

1. *the equation $A\boldsymbol{x} = \boldsymbol{b}$ has no solution, and*
2. *the equation $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$ does have a solution.*

Moreover, the solution to $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$ is the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$.

The justification for the second point of this statement is given in the folowing inquiry.

Inquiry 14.7 (✠2.13): This inquiry explains the reasoning behind Theorem 14.6.1. Let $A\mathbf{x} = \mathbf{b}$ be a matrix equation with $\mathbf{b} \notin \operatorname{col}(A)$, and $\mathbf{b} = \mathbf{p} + \mathbf{e}$, where $\mathbf{p} = \operatorname{proj}_{\operatorname{col}(A)}(\mathbf{b})$.

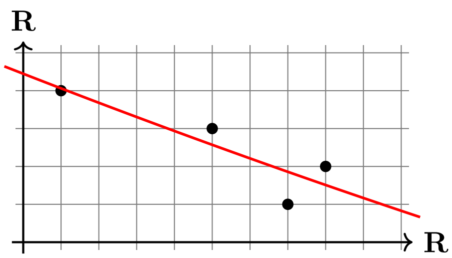
1. Explain why the error vector $\mathbf{e} \in \operatorname{null}(A^T)$.
   Hint: use orthogonal complements.
2. What does $A^T\mathbf{b}$ simplify to, when $\mathbf{b}$ is replaced by $\mathbf{p} + \mathbf{e}$? Use what you showed above.
3. Convince yourself that $A^T\mathbf{p} \in \operatorname{col}(A^T)$. Explain why this means that $A^T\mathbf{p} \notin \operatorname{null}(A)$.
4. Show that if $\mathbf{x} \in \operatorname{null}(A)$, then $\mathbf{x} \in \operatorname{null}(A^TA)$.
   Hint: use orthogonal complements.
5. Show that if $\mathbf{x} \in \operatorname{null}(A^TA)$, then $\mathbf{x} \in \operatorname{null}(A)$.
   Hint: use the positive definiteness of the norm.
6. Put everything together to get that $A^T\mathbf{p} \in \operatorname{col}(A^TA)$. Explain why this means that $A^TA\mathbf{x} = A^T\mathbf{b}$ has a solution.

For points 4 and 5, see Exercise 14.2 for more guidance.

14.2 Least squares for higher degree polynomials

Suppose we want to generalize the previous section, and find a quadratic function that goes through three points in the plane $\mathbf{R}^2$. Quadratics have the form $y = ax^2 + bx + c$, so there are three variables a, b, c that need to be found.

Example 14.8. 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

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

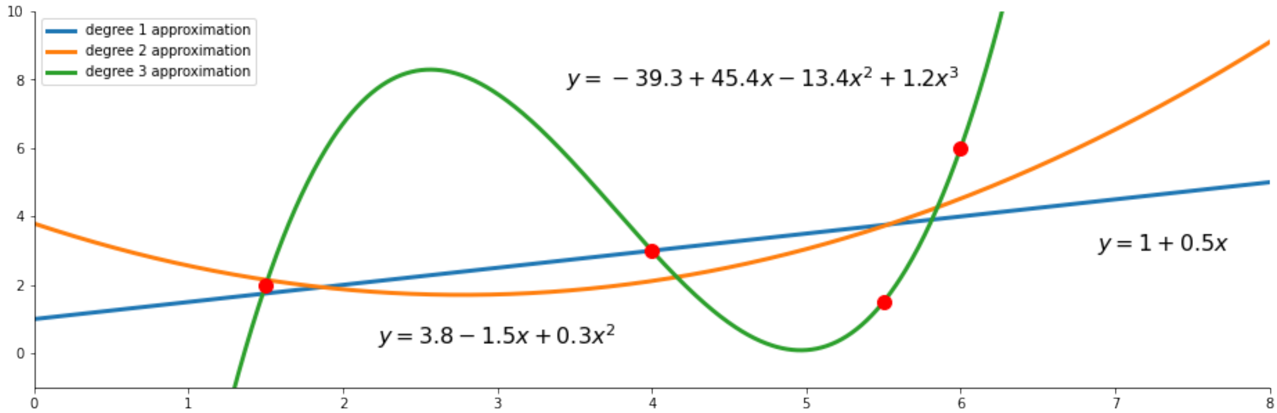
which, when placed into a system, leads to the solutions $a = \frac{1}{372}$, $b = -\frac{241}{620}$, $c = \frac{2068}{465}$, as shown in the plot above.

Definition 14.9: Let $\mathbf{p}_1 = (x_1, y_1), \ldots, \mathbf{p}_n = (x_n, y_n) \in \mathbf{R}^2$. The degree-$d$ polynomial $a_0 + a_1x + a_2x^2 + \cdots + a_dx^d$ that approximates the points $\mathbf{p}_i$ is the least squares solution to the matrix equation

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

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).

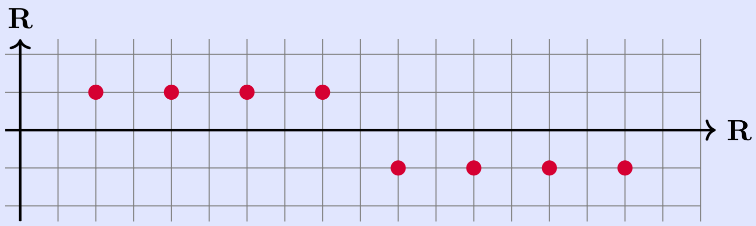
Example 14.10. Suppose that we have four points in the plane. The degree 1, 2, and 3 appoximations to the four points are given below. Note that individually, the points do not get close to the higher degree approximations, but the degree 3 approximation does go through all of them.



Inquiry 14.11 (✠2.14): Let $\mathbf{p}_1, \ldots, \mathbf{p}_n \in \mathbf{R}^2$.

1. Explain why the degree-$(n-1)$ polynomial given by Definition 14.9 will always go through every one of the points perfectly.
   Hint: How many points are needed to define a unique line?

2. Let $\varepsilon > 0$ be a very small value, such as $\frac{1}{1000}$. Let $n = 8$, with $\mathbf{p}_i = (i, \varepsilon)$ for $i = 1, 2, 3, 4$, and $\mathbf{p}_i = (i, -\varepsilon)$ for $i = 5, 6, 7, 8$. Explain which degree-d approximation, for $d = 1, \ldots, 7$, is the "best" approximation for these points. The points are given in the diagram below.



14.3 Exercises

Exercise 14.1. (✠2.13) Using the setup from Example 14.1, finished in Example 14.4, to come to the same conclusion (that is, the same best fit linear equation), but use the projection matrix instead of partial derivatives.

Exercise 14.2. (✠2.13) Let $A \in \mathcal{M}_{m\times n}$.

1. Suppose that $\mathbf{x} \in \text{null}(A)$. Show that $\mathbf{x} \in \text{null}(A^TA)$ as well
2. Suppose that $\mathbf{y} \in \text{null}(A^TA)$. Show that $\mathbf{y} \in \text{null}(A)$ as well.
3. The above two points imply that $\text{null}(A) = \text{null}(A^TA)$. In the case that the columns of A are linearly independent, use this fact to show that A^TA has full rank.

Exercise 14.3. (✠2.13) Consider the set of six points $P = \{p_1, \dots, p_6\} \subseteq \mathbf{R}^2$, with:

$$p_1 = (-1, 3),\ p_2 = (4, 6),\ p_3 = (3, 1),\ p_4 = (-2, -3),\ p_5 = (6, -7),\ p_6 = (-6, 4).$$

1. Either using the projection matrix or partial derivatives, find the line $y = ax + b$ that is the least squares approximation to the points.
2. Find a point $p_7 \in \mathbf{R}^2$ such that the least squares approximation to P is the same as to $P \cup \{p_7\}$. *Hint: Don't redo all your work! Use an observation from partial derivatives.*
3. Let $c \in \mathbf{R}$. Find a point $p_8 \in \mathbf{R}^2$ such that the least squares approximation to $P \cup \{p_8\}$ has slope c.

Exercise 14.4. (✠2.14) ⋈ Write a function in Python that takes two inputs:

- a list of points in $\mathbf{R}^2$,
- a positive integer d,

and returns the degree-d least squares approximation to the input points. You may use the `solve` command from `numpy.linalg` or `scipy.linalg` .

Exercise 14.5. (✠2.13) Consider the following collection of four points $P = \{p_1, p_2, p_3, p_4\} \subseteq \mathbf{R}^3$:

$$p_1 = (1, -2, -4), \quad p_2 = (0, 5, 5), \quad p_3 = (-6, -7, 2), \quad p_4 = (1, 4, -1).$$

1. Generalize the least squares approach and find the closest plane H in $\mathbf{R}^3$ to the points in P (instead of the closest line in $\mathbf{R}^2$).
2. Project the points in P onto the plane H from part 1.
Warning: The plane H will not go through the origin. You need to shift everything first.

Exercise 14.6. (✠2.14) 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 14.7. (✠2.13) Any line in $\mathbf{R}^3$ may be given (not uniquely) by $\ell(t) = (a_1, a_2, a_3)t + (b_1, b_2, b_3)$.

1. Given two such arbitrary lines, find the location of the points on each which minimize the distance between them.

⋈ 2. Take 1000 pairs of such random lines and find the average and standard deviation of the minimum distance between the lines.

Lecture 15: The Gram–Schmidt process

Chapter 4.4 in Strang's "Linear Algebra"

- Fact 1: Every basis can be made into an orthonormal basis.
- Fact 2: The result of the Gram–Schmidt process depends on the order of the vectors input.

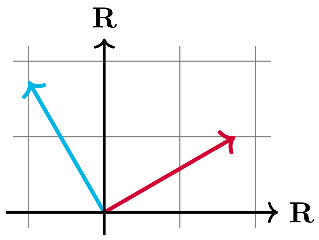
✠ Standard 2.15: Apply the Gram–Schmidt process to a set of vectors.

✠ Standard 2.16: Extend a set of linearly independent vectors to a basis.

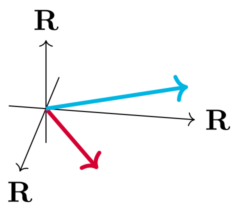
15.1 Orthonormalizing a basis

We previously saw orthogonality and orthonormality in Section 12. We revisit it here from the perspective of bases. Recall that for a set of vectors $B = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ 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 15.1. Placing orthonormal vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbf{R}^n$ as columns in a matrix A will always give $A^T A = I$. For example, taking two orthonormal vectors $\begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}$ and $\begin{bmatrix} -1/2 \\ \sqrt{32} \end{bmatrix}$ in $\mathbf{R}^2$ as columns of a matrix will show this property, as well as when we consider them as lying in the xy-plane of $\mathbf{R}^3$.



$$\underbrace{\begin{bmatrix} \sqrt{3}/2 & 1/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}}_{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\underbrace{\begin{bmatrix} \sqrt{3}/2 & 1/2 & 0 \\ -1/2 & \sqrt{3}/2 & 0 \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \\ 0 & 0 \end{bmatrix}}_{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The key idea here is that even though there are many different pairs of orthonormal vectors, they all have the common property that they multipy with their transpose to the identity matrix.

Example 15.2. We have already seen the *rotation* matrix $R_\theta := \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ from Example 12.7 in Lecture 12 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 15.3. Whenever $A \in \mathcal{M}_{m\times n}$ has orthonormal columns, the lengths of $\mathbf{v}$ and $A\mathbf{v}$ are the same, for any $\mathbf{v} \in \mathbf{R}^n$. This follows directly from Remark 15.1:

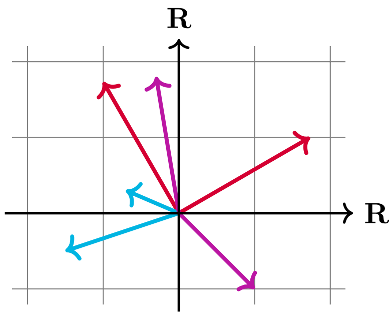
$$\|A\mathbf{v}\|^2 = (A\mathbf{v}) \bullet (A\mathbf{v}) = (A\mathbf{v})^T(A\mathbf{v}) = (\mathbf{v}^T A^T)A\mathbf{v} = \mathbf{v}^T(A^T A)\mathbf{v} = \mathbf{v}^T I\mathbf{v} = \mathbf{v}^T\mathbf{v} = \mathbf{v} \bullet \mathbf{v} = \|\mathbf{v}\|^2.$$

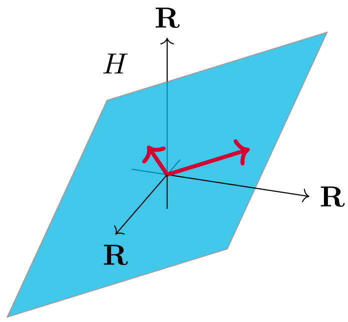
Inquiry 15.4 (✠2.15): Let V be a vector space, and $U \subseteq V$ a subspace with basis vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$. Suppose that the basis vectors of U are orthonormal.

1. Using Definition 13.6, write the projection matrix P for projecting to U.
2. Let $\mathbf{v} \in V$. Express the projection $\text{proj}_U(\mathbf{v})$ as a linear combination of the basis vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ of U.
   Hint: keep track of the basis vectors $\mathbf{u}_i$ in the projection matrix P.

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 15.5. In the plane $\mathbf{R}^2$, every pair of vectors that do not lie on the same line form a basis for the plane. However, some pairs of vectors $\mathbf{u}, \mathbf{v}$ are more special than others - those which lie at a 90° angle to each other. Equivalently, it is those pairs $\mathbf{u}, \mathbf{v}$ for which $\text{proj}_{\mathbf{u}}(\mathbf{v}) = \text{proj}_{\mathbf{v}}(\mathbf{u}) = 0$.





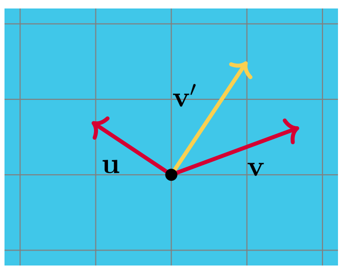
Vectors perpendicular to each other are much easier to deal with, so we try to only work with those. This is the case also for subspaces of vector spaces, for example the plane H defined by $2x+3y-2z = 0$ in $\mathbf{R}^3$. To find the two basis vectors of this plane, we compute a nullspace:

$$H = \text{null}\left(\begin{bmatrix} 2 & 3 & -2 \end{bmatrix}\right) = \text{null}\left(\begin{bmatrix} 1 & \frac{3}{2} & -1 \end{bmatrix}\right) = \text{span}\left(\begin{bmatrix} -\frac{3}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right).$$

These two vectors in the span are not orthogonal to each other, as

$$\underbrace{\begin{bmatrix} -\frac{3}{2} \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{u}} \bullet \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}} = -\frac{3}{2} \neq 0.$$

In the plane H:



We would like to make them orthogonal to get a nicer basis. All that we need is to make $\mathbf{v}$ orthogonal to $\mathbf{u}$, and recalling that everything in the orthogonal complement of $\text{span}(\mathbf{u})$ will fulfill this criteria, we simply project $\mathbf{v}$ onto $\text{span}(\mathbf{u})^\perp$. Following the formula in Remark 13.5, this new vector is

$$\mathbf{v}' = \text{proj}_{\text{span}(\mathbf{u})^\perp}(\mathbf{v}) = (I-P)\mathbf{v} = \mathbf{v} - \frac{\mathbf{v} \bullet \mathbf{u}}{\mathbf{u} \bullet \mathbf{u}}\mathbf{u} = \begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{-3/2}{9/4+1}\begin{bmatrix}-\frac{3}{2}\\1\\0\end{bmatrix} = \begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{-6}{13}\begin{bmatrix}-\frac{3}{2}\\1\\0\end{bmatrix} = \begin{bmatrix}1-\frac{9}{13}\\\frac{6}{13}\\1\end{bmatrix} = \begin{bmatrix}\frac{4}{13}\\\frac{6}{13}\\1\end{bmatrix}.$$

The two vectors $\mathbf{u}, \mathbf{v}'$ still span H, but now we have the added benefit of orthogonality:

$$\mathbf{u} \bullet \mathbf{v}' = \begin{bmatrix}-\frac{3}{2}\\1\\0\end{bmatrix} \bullet \begin{bmatrix}\frac{4}{13}\\\frac{6}{13}\\1\end{bmatrix} = -\frac{6}{13} + \frac{6}{13} = 0.$$

Algorithm 6 (The Gram–Schmidt Process): Suppose you have a set $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ of linearly independent vectors. The Gram–Schmidt pocess will first create a set of orthogonal vectors $\mathbf{w}_1, \ldots, \mathbf{w}_n \in V$, and then a set of orthonormal vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n \in V$. They will have all the same span: $\text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \text{span}(\mathbf{w}_1, \ldots, \mathbf{w}_n) = \text{span}(\mathbf{q}_1, \ldots, \mathbf{q}_n)$.

- Let $\mathbf{w}_1 = \mathbf{v}_1$
- For each $i = 2, \ldots, n$:
  - Let $\mathbf{w}_i = \mathbf{v}_i - \Big(\text{proj}_{\mathbf{w}_{i-1}}(\mathbf{v}_i) + \cdots + \text{proj}_{\mathbf{w}_1}(\mathbf{v}_i)\Big)$.
- The orthonormal set of vectors is $\mathbf{q}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}$.

Inquiry 15.6 (✠2.15): Let V be a vector space and $\mathbf{u}, \mathbf{v} \in V$ be linearly independent vectors.

1. What will be the output of the Gram–Schmidt process when it is run on $\mathbf{v}, 2\mathbf{v}, 3\mathbf{v}$?
2. What will be the output of the Gram–Schmidt process when it is run on $\mathbf{v}, \mathbf{u}, \mathbf{v}+\mathbf{u}$?
3. Explain why running the Gram–Schmidt process on the two sets $\mathbf{v}, \mathbf{u}, \mathbf{v}+\mathbf{u}$ and $\mathbf{v}, \mathbf{u}+\mathbf{v}, \mathbf{u}$ in that order will give the same result.

Example 15.7. 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

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

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:

$$\underbrace{\begin{bmatrix}1&2&0&1\\2&0&2&0\\0&1&1&1\\2&1&1&2\end{bmatrix}}_{A^T}\underbrace{\begin{bmatrix}1&2&0&2\\2&0&1&1\\0&2&1&1\\1&0&1&2\end{bmatrix}}_{A} = \begin{bmatrix}6&2&3&4\\2&8&2&4\\3&2&3&4\\4&4&4&10\end{bmatrix} \neq I_4.$$

Exercise 15.2 works through the Gram–Schmidt process on these vectors.

Inquiry 15.8 (✠2.15): The three vectors $\mathbf{u} = \begin{bmatrix}1\\0\\1\end{bmatrix}, \mathbf{v} = \begin{bmatrix}0\\-1\\1\end{bmatrix}, \mathbf{w} = \begin{bmatrix}1\\-2\\0\end{bmatrix}$ span all of $\mathbf{R}^3$.

1. Run the Gram–Schmidt process on $\mathbf{u}, \mathbf{v}, \mathbf{w}$, in that order, and then on the different order $\mathbf{u}, \mathbf{w}, \mathbf{v}$. You may use a computer.
2. Why are the results different? Is the span of the resulting vectors different?
3. How do you think the two results are related?

If possible, visualize the locations of the vectors on a computer.

15.2 Factorizing and extending

As now is very common, we consider vectors as columns of matrices. Given some vectors $\mathbf{v}_i$ as columns in A, and the resulting orthonormal vectors $\mathbf{q}_i$ as columns in Q, a natural question arises: How are A and Q related?

Proposition 15.9. 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 $\mathbf{q}_i \bullet \mathbf{v}_j = 0$ whenever $j < i$. Indeed, we first note that $\mathbf{q}_i \bullet \mathbf{w}_j = 0$ whenever $i \neq j$, since the $\mathbf{q}_i$ point in the same direction as the $\mathbf{w}_i$. So for example,

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

Remark 15.10. Recall that to find the least squares solution to $A\mathbf{x} = \mathbf{b}$, we projected $\mathbf{b}$ onto $\mathrm{col}(A)$ 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

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

does, least squares was about solving $A^TA\mathbf{x} = A^T\mathbf{b}$. Using the result from Proposition 15.9, this equation becomes

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

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

Inquiry 15.11 (✠2.15): Consider the vector space of all functions $[0,1] \to [0,1]$, similar to Exercise 6.3, with the "dot product" defined by $f \bullet g = \int_0^1 f(x)g(x)\,dx$.

1. Are the two functions x, x^2 linearly independent? Are they orthogonal?
2. Run the Gram–Schmidt process on x, x^2 to get an orthonormal set of functions.
3. Changing the space to set of all functions $[0,2\pi] \to [0,\pi]$, check that $\sin(x), \cos(x)$ are orthogonal.

Remark 15.12. The Gram–Schmidt process is useful for *extending* a basis, a concept previously visited in Inquiry 10.18. That is, given an orthonormal basis for $U \subseteq V$, we can extend the basis to a basis for all of V by simply running the Gram–Schmidt process on the vectors in the given basis, and add as many vectors from V as necessary. For example, given

$$V = \mathbf{R}^4 = \operatorname{span}\left(\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right) \qquad \text{and} \qquad U = \operatorname{span}\left(\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}, \begin{bmatrix}2\\2\\2\\1\end{bmatrix}\right) \subseteq V,$$

we can extend the two vectors in the basis of U to a basis of V. Since the given basis vectors of U are orthogonal (but not orthonormal), the first part of Gram–Schmidt process will not affect them. Since $V = \mathbf{R}^4$ is 4-dimensional, we know two facts:

- two vectors are not enough for a basis of V, so $\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}, \begin{bmatrix}2\\2\\2\\1\end{bmatrix}$ is too small to be a basis, and
- six vectors are too many for a basis of V, so $\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}, \begin{bmatrix}2\\2\\2\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$ is too big to be a basis.

To find an orthonormal basis of V that contains the two basis vectors of U, simply run the Gram–Schmidt process on all six vectors, beginning with the two from the basis of U.

15.3 Exercises

Exercise 15.1. (✠2.15) Check that the columns of the 2×2 rotation matrix (introduced in Lecture 12.1) and of the 3×3 permutation matrices (introduced in Lecture 3) are all orthogonal. Are they orthonormal?

Exercise 15.2. (✠2.15) Apply the Gram-Schmidt process to the vectors $\begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \begin{bmatrix}2\\0\\2\\0\end{bmatrix}, \begin{bmatrix}2\\1\\1\\2\end{bmatrix}, \begin{bmatrix}2\\0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\2\\0\\1\end{bmatrix}$.

Exercise 15.3. (✠2.15) Consider the 2-dimensional subspace $H \subseteq \mathbf{R}^4$ defined by

$$H = \left\{ (x,y,z,w) \in \mathbf{R}^4 \ : \ \begin{array}{rcl} 2x + 3y - w & = & 0, \\ y - z + 2w & = & 0. \end{array} \right\}$$

1. Express H as a span of two vectors.
2. Apply the Gram–Schmidt process to the two vectors from above to get H as a span of two orthonormal vectors.
3. The space $\mathbf{R}^4$ has the xy-plane as a 2-dimensional subspace, with basis $\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$. Give the change of basis matrix from the two vectors in part 2. to these two vectors.

Exercise 15.4. (✠2.16) Let V be the vector space of polynomials of degree at most 3 with domain $[0, 1]$. The "dot product" on V is defined by $f \bullet g = \int_0^1 f(x)g(x)\, dx$, which helps to define length and angle. You may assume that $\{1, x, x^2, x^3\}$ is a basis for V.

1. Is the given basis orthogonal? Find the lengths of the elements in the basis.
2. Are the two functions $2x, x^2 - 1$ linearly independent in V? Are they orthogonal?
3. Extend $\{2x, x^2 - 1\}$ to an orthonormal basis of V.

Lecture 16: Generalized distances

Chapter IV.10 in Strang's "Learning from Data"

- Fact 1: The inner product generalizes the concept a distance for other spaces.
- Fact 2: (Relative) positions of points can be recovered knowing just the distances between them.

✠ Standard 2.17: Determine if something is or is not an inner product space.

✠ Standard 2.18: Compute the length, angle, and projections of vectors in arbitrary inner product spaces.

We now take a small detour from Strang's *Linear Algebra* and work with the material from Strang's *Learning from Data*. The topic follows the topics of the previous lectures, expanding on the idea of orthogonality and unit length in different vector spaces.

16.1 Functions on spaces

Definition 16.1: Let V be a vector space. An *inner product* on V is a function $\langle \cdot\, , \,\cdot \rangle : V^2 \to \mathbf{R}$ 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) $\langle c\mathbf{v}, \mathbf{u} \rangle = c\langle \mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{v}, c\mathbf{u} \rangle$
- (bilinear) $\langle \mathbf{v} + \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$

A vector space V that has an inner product is called an *inner product space*. Given any two vectors $\mathbf{u}, \mathbf{v}$ in an inner product space V,

- they are *orthogonal* if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$,
- the *angle* $\theta \in [0, 2\pi)$ between them is given by $\cos(\theta) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|}$.

Recall that the only required operations for a vector space were scalar multiplication and vector addition (a dot product was not required).

We have already seen an example of the inner product in the *dot product* of two vectors. 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

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v} \bullet \mathbf{v}}.$$

Example 16.2. 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 $\langle A, B \rangle := \text{trace}(A^T B)$. 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.

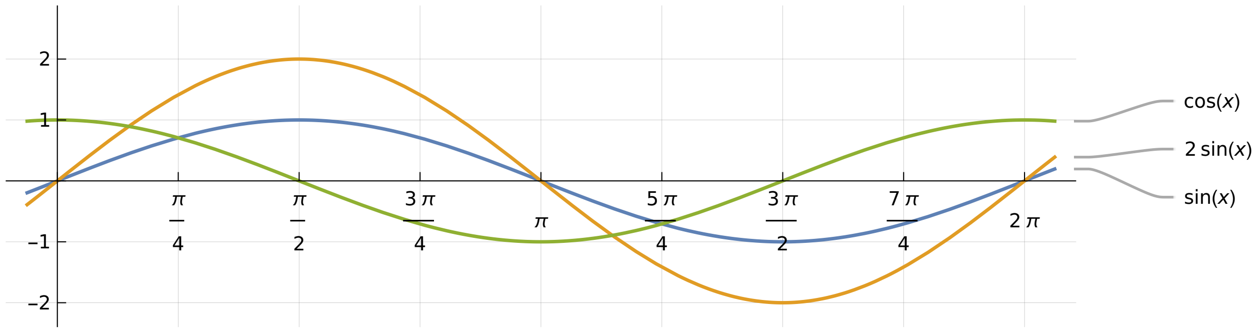
Inquiry 16.3 (✠2.18): This inquiry is about the properties of an inner product space V given in Definition 16.1. Using them, show that:

1. $\langle 0, \mathbf{v}\rangle = \langle \mathbf{v}, 0\rangle = 0$ for all $\mathbf{v} \in V$
2. the only vector in V that is orthogonal to itself is 0
3. the *parallelogram equality* holds: $\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2)$

Theorem 16.3.1. *The inner product* $\langle\,\cdot\,,\,\cdot\,\rangle$ *in any inner product space* $V \ni \boldsymbol{v}, \boldsymbol{w}$ *satisfies:*

- *the Cauchy–Schwarz inequaity:* $|\langle \boldsymbol{v}, \boldsymbol{w}\rangle| \leqslant \|\boldsymbol{v}\|\|\boldsymbol{w}\|$ *with equality iff* $\boldsymbol{v}$ *and* $\boldsymbol{w}$ *are linearly dependent*
- *the triangle inequality:* $\|\boldsymbol{v}+\boldsymbol{w}\| \leqslant \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$

Example 16.4. Using the first point of Theorem 16.3.1, we can show that the functions $\sin(x)$ and $\cos(x)$ are linearly independent in $C[0, 2\pi]$, and that $\sin(x)$ and $2\sin(x)$ are linearly dependent.



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 $0 \neq \sqrt{\pi}\cdot\sqrt{\pi} = \pi$, the functions $\sin(x)$ and $\cos(x)$ are linearly independent, but since $2\pi = \sqrt{\pi}\cdot\sqrt{4\pi}$, the functions $\sin(x)$ and $2\sin(x)$ are linearly dependent. 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 16.5. 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

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

Remark 16.6. The Gram–Schmidt process in Lecture 15 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:

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

Inquiry 16.7 (✠2.18): Consider the inner product spaces $C[a,b]$ and $\mathcal{M}_{2\times 2}$.

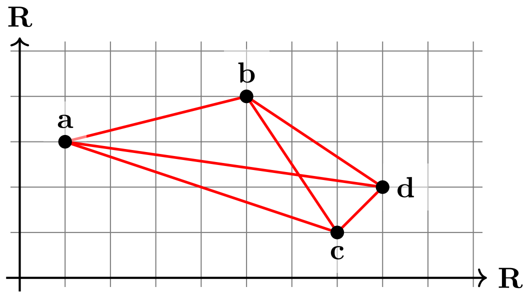
1. What is the angle between $x+1$ and x^2+1 in $C[0,1]$?
2. Compute the projection of $\cos(x)$ onto $\sin(x)$ in $C[0,2\pi]$.
3. Compute the projection of the rotation matrix R_θ onto $\left[\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right]$. For what angles θ are these matrices orthogonal to each other?

16.2 Distance matrices

Recall the points from Exercise 14.8 in Lecture 15, which were used in the motivating least squares example. If the points were located elsewhere but their *relative* position to each other was the same, we can 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 16.8. Consider the distances among the four points, slightly adapted from Exercise 14.8.



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

The matrix X is called the *position matrix*. We can easily compute the symmetric *distance matrix*

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

which contains the squares of the distance among the points. A method to recover X knowing only D is not so clear, however.

Proposition 16.9. Let $D \in \mathcal{M}_{k\times k}$ be the matrix containg squares of distances among k points $\mathbf{v}_1,\dots,\mathbf{v}_k \in \mathbf{R}^n$. The relationship between D and the position matrix X is given by

$$X^TX = \frac{1}{2}\left(\mathbf{s}\begin{bmatrix}1&1&\cdots&1\end{bmatrix} + \begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}\mathbf{s}^T - D\right), \qquad \text{for} \qquad \mathbf{s} = \begin{bmatrix}\|\mathbf{v}_1-\mathbf{v}_1\|^2\\ \|\mathbf{v}_1-\mathbf{v}_2\|^2\\ \vdots\\ \|\mathbf{v}_1-\mathbf{v}_k\|^2\end{bmatrix}$$

the transpose of the first row of the matrix D.

Example 16.10. Continuing Example 16.8, we fix one of the points as a reference point. Without loss of generality, we simply say

$$\mathbf{a} = 0.$$

That is, we subtract $\mathbf{a}$ from all the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ to get new ones (which we call the same). Any other vector $\mathbf{b}, \mathbf{c}, \mathbf{d}$ could have been chose. Now the first line of D becomes (squares of) the lengths $\|\cdot\|$ of all the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$, and the lengths also appear on the diagonal of $X^T X$:

$$D = \begin{bmatrix} 0 & \|\mathbf{b}\|^2 & \|\mathbf{c}\|^2 & \|\mathbf{d}\|^2 \\ \|\mathbf{b}\|^2 & 0 & \|\mathbf{b}-\mathbf{c}\|^2 & \|\mathbf{b}-\mathbf{d}\|^2 \\ \|\mathbf{c}\|^2 & \|\mathbf{c}-\mathbf{b}\|^2 & 0 & \|\mathbf{c}-\mathbf{d}\|^2 \\ \|\mathbf{d}\|^2 & \|\mathbf{d}-\mathbf{b}\|^2 & \|\mathbf{d}-\mathbf{c}\|^2 & 0 \end{bmatrix}, \qquad X^T X = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathbf{b}\bullet\mathbf{b} & \mathbf{b}\bullet\mathbf{c} & \mathbf{b}\bullet\mathbf{d} \\ 0 & \mathbf{c}\bullet\mathbf{b} & \mathbf{c}\bullet\mathbf{c} & \mathbf{c}\bullet\mathbf{d} \\ 0 & \mathbf{d}\bullet\mathbf{b} & \mathbf{d}\bullet\mathbf{c} & \mathbf{d}\bullet\mathbf{d} \end{bmatrix}.$$

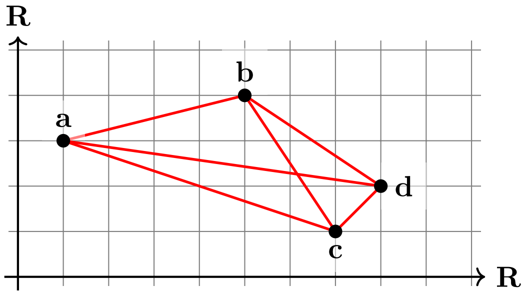
Applying the result of Proposition 16.9, we construct the position vector $\mathbf{s}^T = \begin{bmatrix} 0 & \|\mathbf{b}\|^2 & \|\mathbf{c}\|^2 & \|\mathbf{d}\|^2 \end{bmatrix} = \begin{bmatrix} 0 & 17 & 40 & 50 \end{bmatrix}$, and compute

$$\begin{aligned} X^T X &= \frac{1}{2}\left(\mathbf{s}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mathbf{s}^T - D \right) \\ &= \frac{1}{2}\left(\begin{bmatrix} 0 \\ 17 \\ 40 \\ 50 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 17 & 40 & 40 \end{bmatrix} - \begin{bmatrix} 0 & 17 & 40 & 50 \\ 17 & 0 & 13 & 13 \\ 40 & 13 & 0 & 2 \\ 50 & 13 & 2 & 0 \end{bmatrix} \right) \\ &= \frac{1}{2}\left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 17 & 17 & 17 & 17 \\ 40 & 40 & 40 & 40 \\ 50 & 50 & 50 & 50 \end{bmatrix} + \begin{bmatrix} 0 & 17 & 40 & 50 \\ 0 & 17 & 40 & 50 \\ 0 & 17 & 40 & 50 \\ 0 & 17 & 40 & 50 \end{bmatrix} - \begin{bmatrix} 0 & 17 & 40 & 50 \\ 17 & 0 & 13 & 13 \\ 40 & 13 & 0 & 2 \\ 50 & 13 & 2 & 0 \end{bmatrix} \right) \\ &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 17 & 22 & 27 \\ 0 & 22 & 40 & 44 \\ 0 & 27 & 44 & 50 \end{bmatrix}. \end{aligned}$$

Note that $X^T X$ is symmetric (see Lecture 22 for more on this topic) with rank 2. Hence, doing row reduction to get the LDU-decomposition of $X^T X$ will produce symmetric matrices, that is, $LDU = (L\sqrt{D})(\sqrt{D}U)$, with $L = U^T$. This will recover X, up to a shift and potentially a rotation and a reflection. Note that the matrix "D" here is the diagonal matrix from the LDU-decomposition, and is different from the distance matrix "D" used above.
For this example, we have

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 17 & 22 & 27 \\ 0 & 22 & 40 & 44 \\ 0 & 27 & 44 & 50 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{22}{17} & 1 & 0 \\ 0 & \frac{27}{17} & \frac{11}{14} & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 17 & 22 & 27 \\ 0 & 0 & \frac{196}{17} & \frac{154}{17} \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{U} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{22}{17} & 1 & 0 \\ 0 & \frac{27}{17} & \frac{11}{14} & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 17 & 0 & 0 \\ 0 & 0 & \frac{196}{17} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{22}{17} & \frac{27}{17} \\ 0 & 0 & 1 & \frac{11}{14} \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{U},$$

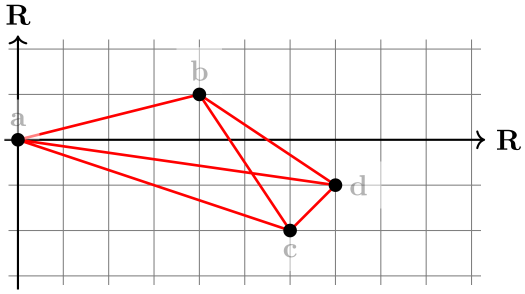
where we note that L is not precisely the transpose of U, due to zero rows. However, considering the highlighted 2×4 subatrix X of U, we do indeed see it and its transpose in the decomposition. From this we recover the points $\mathbf{a}', \mathbf{b}', \mathbf{c}', \mathbf{d}'$ as the middle two rows of $\sqrt{D}U$. These are not exactly the original points, but there are “transformations” that get us to them from the original points.



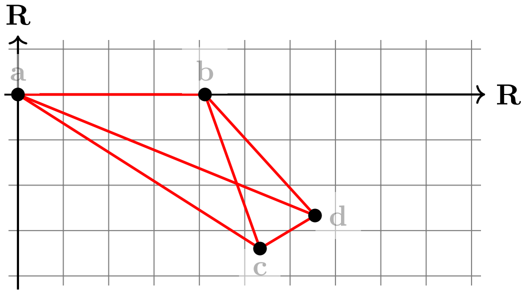
original points $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$

$\mathbf{a} = (1, 3)$ $\mathbf{c} = (7, 1)$

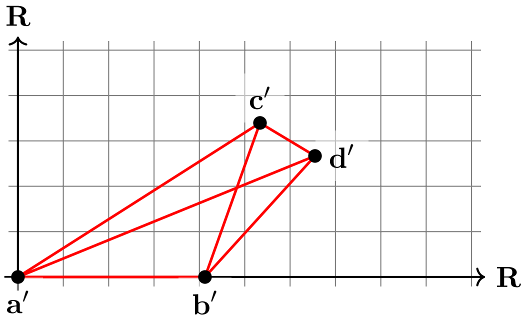
$\mathbf{b} = (5, 4)$ $\mathbf{d} = (8, 2)$



shifted so that $\mathbf{a}$ is at the origin



rotated so that $\mathbf{b}$ is on the x-axis



reflected on the y-axis to get points $\mathbf{a}', \mathbf{b}', \mathbf{c}', \mathbf{d}'$

$\mathbf{a}' = (0, 0)$ $\mathbf{c}' = \left(\frac{22}{\sqrt{17}}, \frac{14}{\sqrt{17}}\right)$

$\mathbf{b}' = (\sqrt{17}, 0)$ $\mathbf{d}' = \left(\frac{27}{\sqrt{17}}, \frac{11}{\sqrt{17}}\right)$

Inquiry 16.11 (✠2.18): Consider vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ in $\mathbf{R}^2$. If we know only the matrix D of distances between them, the recovery method presented in Example 16.8 computes the positions of $\mathbf{a} - \mathbf{a}, \mathbf{b} - \mathbf{a}, \mathbf{c} - \mathbf{a}, \mathbf{d} - \mathbf{a}$.

1. Suppose instead $\mathbf{b}$ was subtracted from all the vectors. What is the relationship between the vectors recovered in this way to those recovered by subtracting $\mathbf{a}$?
2. Suppose you have 4 new vectors, which are just $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ rotated by 90° clockwise. After applying the recovery method to get X, how are the recovered vectors related to the vectors recovered by the first method?

Remark 16.12. If instead we have a set of vectors $\mathbf{v}_i, \ldots, \mathbf{v}_k$, 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 16.13. 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

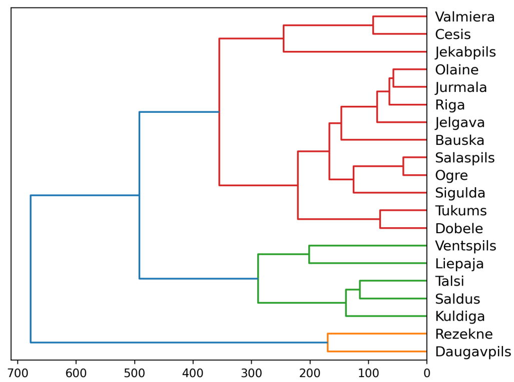
$$\begin{aligned} 3^2 &= \|\mathbf{b} - \mathbf{c}\|^2 \\ &= \langle \mathbf{b} - \mathbf{c}, \mathbf{b} - \mathbf{c} \rangle \\ &= \langle \mathbf{b}, \mathbf{b} - \mathbf{c} \rangle - \langle \mathbf{c}, \mathbf{b} - \mathbf{c} \rangle \\ &= \langle \mathbf{b}, \mathbf{b} \rangle - \langle \mathbf{b},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

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 16.14. 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

$$(\text{distance between } c_i \text{ and } c_j) = \frac{1}{|c_i||c_j|} \sum_{\mathbf{v}_i \in c_i} \sum_{\mathbf{v}_j \in c_j} \|\mathbf{v}_i - \mathbf{v}_j\|. \tag{6}$$

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.

Inquiry 16.15 (✠2.18): This question is about coding in Python.

1. Generate a collection of 333 random points in $\mathbf{R}^2$, with:
   - 300 of them randomly selected from $[0,1] \times [0,1]$,
   - 30 of them randomly selected from $[5,6] \times [5,6]$,
   - 3 of them randomly selected from $[1,2] \times [8,9]$.
2. Construct the 333×333 distance matrix between them.
3. Construct the dendrogram from this matrix. Does it reflect the clusters as you created them?

16.3 Exercises

Exercise 16.1. (✠2.17) For each of the following "definitions", show that each cannot be an inner product.

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

Exercise 16.2. (✠2.17, 2.18) Check the conditions for the space of $m \times n$ matrices over $\mathbf{R}$ from Example 16.2 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 16.3. (✠2.18) 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 $\text{span}\{A, B, C\}$. Use the inner product on matrices given in Example 16.2.

Exercise 16.4. (✠2.18) Let $P(\mathbf{R})$ be the vector space of all polynomials $\mathbf{R} \to \mathbf{R}$, with scalar multiplication and polynomial addition defined as you would expect. You may assume that the following is an inner product on $P(\mathbf{R})$:

$$\langle p(x), q(x)\rangle = \int_0^\infty p(x)q(x)e^{-x}\, dx.$$

1. Check that $p(x) = 2x - 1$ and $q(x) = x + 3$ are not orthogonal to each other.
2. Using the Gram–Schmidt process on $p(x)$ and $q(x)$ as in part 1., find a polynomial $r(x) \in P(\mathbf{R})$ that is orthogonal to $p(x)$. Give your answer as $r(x) = ax + b$, for $a, b \in \mathbf{Z}$.

Exercise 16.5. (✠2.18) Given the distance D matrix below, construct the dendrogram using the same average distance method as in Example 16.14. 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 17: Linear transformations

Chapters 8.1, 8.2 in Strang's "Linear Algebra"

- Fact 1: A linear transformation is the same thing as a matrix.
- Fact 2: A linear transformation is injective iff it is surjective.

✠ Standard 2.19: Determine whether or not a function is a linear transformation.

✠ Standard 2.20: Construct a matrix for a linear transformation, knowing what it does to a basis.

✠ Standard 2.21: Construct the image and kernel of a linear transformation, as vector spaces.

This lecture focuses on a generalization: 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 12. By the end of this lecture, we will see that every such function comes from a matrix.

17.1 Types of linear transformations

Definition 17.1: Let V, W be vector spaces. A *linear transformation*, or *linear map*, is a function $f\colon V \to W$ that satisfies

$$f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) \qquad \text{and} \qquad f(c\mathbf{x}) = cf(\mathbf{x}) \tag{7}$$

for every $\mathbf{x}, \mathbf{y} \in V$ and every $c \in \mathbf{R}$. These are conditions for *linearity*.

Example 17.2. 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 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

$$\left\| \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\| = \left\| \begin{bmatrix} -1 \\ -2 \end{bmatrix} \right\| = \sqrt{3}, \qquad \text{but} \qquad \left\| \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ -2 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\| = 0 \neq 2\sqrt{3}.$$

Inquiry 17.3 (✠2.19): Each of the functions below are linear. For each, show that the two conditions for linearity are satisfied.

1. the dot product of a vector $\mathbf{v} \in \mathbf{R}^3$ with $\left[\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right] \in \mathbf{R}^3$, as a function $\mathbf{R}^3 \to \mathbf{R}$
2. projection of a vector $\mathbf{v} \in \mathbf{R}^3$ to the x-axis, considered as the span of $\left[\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right]$
3. differentiation and integration on the space $C[\mathbf{R}]$ of continuous functions

Each of the functions below is not linear. For each, show which of the linearity conditions are violated.

4. addition of the vector $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$: $f\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$
5. squaring of every component: $f\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} x^2 \\ y^2 \end{smallmatrix}\right]$

Linearity is good because it gives a complete picture with a small amount of information.

Proposition 17.4. 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 $\mathbf{v}_1, \dots, \mathbf{v}_n$ of V and taking any (not necessarily linearly independent!) vectors $\mathbf{w}_1, \dots, \mathbf{w}_n \in W$, there is only one unique linear map $f\colon V \to W$ for which $f(\mathbf{v}_i) = \mathbf{w}_i$, for all i.

Inquiry 17.5 (✠2.20): This inquiry is about the vector space $\mathbf{R}^3$.

1. Come up with two different bases $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ for $\mathbf{R}^3$.
2. Let A be the 2×3 matrix with $A\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $A\mathbf{v}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $A\mathbf{v}_3 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$. What is A?
3. Let B be the 4×3 matrix with $B\mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$, $B\mathbf{v}_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \end{bmatrix}$ and $B\mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 2 \end{bmatrix}$. What is B?
4. What are the ranks of A and B? How are these numbers related to the dimension of $\mathbf{R}^3$?

Every linear transformation $V \to W$ creates new subspaces of V and of W.

Definition 17.6: 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 $\operatorname{im}(f) \subseteq W$ is a subspace of W.

Example 17.7. 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,

$$\ker(f) = \operatorname{null}(A), \qquad \operatorname{im}(f) = \operatorname{col}(A).$$

Recall that a function $f\colon X \to Y$ 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 17.8. Let $f\colon V \to W$ be linear.

- f is injective iff $\ker(f) = \{0\}$
- if $\dim(W) = \dim(\operatorname{im}(f))$, then f is surjective.

Inquiry 17.9 (✠2.21): This inquiry describes the justification for Proposition 17.8.

1. Suppose that $\ker(f) = \{0\}$. Show that assuming $f(\mathbf{x}) = f(\mathbf{y})$ implies that $\mathbf{x} = \mathbf{y}$.
2. Suppose that f is injective. Use the second linearity condition with $c = 0$ to show that assuming a nonzero vector is in the kernel of f implies a contradiction.
3. Revisit Remark 10.15 and explain why it justifies the second point of the proposition.

Combining injective and surjective linear transformations gives us a very special transformation.

Definition 17.10: A linear transformation $f\colon V \to W$ 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 17.11. We have already seen examples of isomorphisms:

- The map $f\colon \mathbf{R}^n \to \mathbf{R}^n$ with $f(\mathbf{x}) = 2\mathbf{x}$ is an isomoprhism.
- The change of basis matrix from Lecture 10 is an isomorphism
- The dot product of any vector in $\mathbf{R}^2$ with $(-1,2)$ is not an isomoprhism, as it fails injectivity: $(3,4)\cdot(-1,2) = (-5,0)\cdot(-1,2)$.

17.2 The matrix of a linear transformation

Theorem 17.11.1. *Let $f\colon \mathbf{R}^n \to \mathbf{R}^m$ 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 $\mathbf{e}_1,\dots,\mathbf{e}_n$ for $\mathbf{R}^n$ and $\mathbf{e}_1,\dots,\mathbf{e}_m$ for $\mathbf{R}^m$. By Proposition 17.4, 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

$$\begin{aligned} f(\mathbf{x}) &= f(b_1\mathbf{e}_1 + \cdots + b_n\mathbf{e}_n) \\ &= b_1 f(\mathbf{e}_1) + \cdots + b_n f(\mathbf{e}_n) \\ &= b_1(a_{11}\mathbf{e}_1 + \cdots + a_{m1}\mathbf{e}_m) + \cdots + b_n(a_{1n}\mathbf{e}_1 + \cdots + a_{mn}\mathbf{e}_m) \\ &= (b_1a_{11} + \cdots + b_na_{1n})\mathbf{e}_1 + \cdots + (b_1a_{m1} + \cdots b_na_{mn})\mathbf{e}_m. \end{aligned}$$

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 $\mathbf{v}_1,\dots,\mathbf{v}_n$ is some basis for $\mathbf{R}^n$ and $\mathbf{w}_1,\dots,\mathbf{w}_m$ 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 $C_W^{-1}AC_V$. □

Inquiry 17.12 (✠2.21): This inquiry connects linear transformations with matrices. Recall that A^T is the transpose of A.

1. Considering the transpose as a "function" $\mathcal{M}_{3\times 2} \to \mathcal{M}_{2\times 3}$, explain why this cannot be linear.
   Hint: How would this work as a matrix multiplication?

2. Explain why the function $f\colon \mathcal{M}_{3\times 2} \to \mathbf{R}^6$, for which

$$f\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\right) = \begin{bmatrix} a_{11} & a_{12} & a_{21} & a_{22} & a_{31} & a_{32} \end{bmatrix}^T$$

is linear. What is its inverse, and is it linear as well?

3. Describe a function $g\colon \mathbf{R}^6 \to \mathbf{R}^6$ so that

$$(f^{-1} \circ g \circ f)\colon \mathcal{M}_{3\times 2} \to \mathcal{M}_{2\times 3}$$

produces the transpose of a matrix. Is it linear?

This Theorem above has several implications. Combining the rank-nullity theorem from Lecture 11 along with observations above, we immediately get the following.

Corollary 17.13. 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 17.11.1 in Example 17.7 to describe every linear map as a matrix.

The second point follows immediately from the first point and Proposition 17.8. □

Remark 17.14. 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

$$(g \circ f)(\mathbf{x}) = g(f(\mathbf{x})) = g(A\mathbf{x}) = B(A\mathbf{x}) = (BA)\mathbf{x}.$$

17.3 Exercises

Exercise 17.1. (✠2.19) For this question, the vector $T_i(\mathbf{x})$ is simply written $T_i\mathbf{x}$, to both ease notation and as a reminder that linear transformatons are simply matrices. You are given 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.
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 17.2. (✠2.21) Prove the claim from Definition 17.6 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 17.3. (✠2.21) 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 17.4. (✠2.20) Consider the three orthogonal vectors

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \qquad \mathbf{z} = \begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix}.$$

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 17.5. (✠2.20, 2.21) 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 transformation 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 17.6. (✠2.21) Prove the claim from Remark 17.14 that the composition of two linear maps is linear.

Part III
Eigentheory

Lecture 18: Defining the determinant

Chapters 5.1,5.2 in Strang's "Linear Algebra"

- Fact 1: The determinant may be computed either recursively or combinatorially, only for a square matrix.
- Fact 2: The determinant is related to the pivots and invertibility of a matrix.

✠ Standard 3.01: Compute the determinant using both the recursive and combinatorial definitions.

✠ Standard 3.02: Use the multilinearity and alternating properites to infer results for special matrices.

We now begin a new part of this course, on everything to do with *eigenvectors* and *eigenvalues*. The first step is the *determinant* of a matrix, which is a rough estimate of the eigenvalues of the matrix. As we will see later, the determinant is the product of all the eigenvalues.

18.1 The recursive definition

The *determinant* is a function $\det\colon \mathcal{M}_{n\times n} \to \mathbf{R}$, and denoted as either $\det(A)$ or with vertical bars $|A|$. Before we get to definitions and new ideas, we consider some concepts you have already seen, in this and previous courses.

Example 18.1. The determinant is often associated with *invertibility* of a matrix.

- (Definition 18.5) The determinant of a 1×1 matrix $\left[\, a \,\right]$ is a. The matrix is not invertible is $a = 0$.
- (Definition 18.5) 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.

Example 18.2. The determinant is also associated with the *shape change* of a square. In $\mathbf{R}^2$:

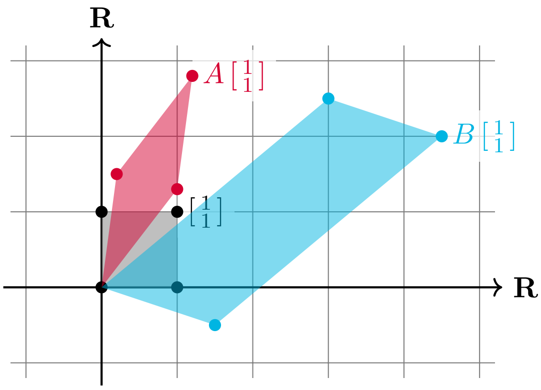
- the vectors $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ enclose a square with area 1,
- the vectors $A\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right], A\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], A\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], A\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ enclose a parallelogram with area $|\det(A)|$,

for any 2×2 matrix A. This extends to $\mathbf{R}^n$, with a generalization of "square" and "area."

> **Definition 18.3:** Let $n \in \mathbf{N}$. The *unit n-cube* in $\mathbf{R}^n$ is the set of points $(x_1, \dots, x_n)$ with $0 \leqslant x_i \leqslant 1$ for all i. The unit n-cube has *n-dimensional volume*, or simply *n-volume*, equal to 1. The n-volume of any other shape in $\mathbf{R}^n$ is given by the number of (fractions of) unit n-cubes in the shape.

This way to define n-dimensional volume is a rough estimate of the more accurate way, which would be to take an n-fold integral.

Example 18.4. Consider $A \in \mathcal{M}_{n\times n}$ as a function $\mathbf{R}^n \to \mathbf{R}^n$. The absolute value of the determinant of A is the n-dimensional volume of the shape that the unit n-cube becomes, after multiplying each of its corners by A.



$$A = \begin{bmatrix} 1 & \frac{1}{5} \\ \frac{13}{10} & \frac{3}{2} \end{bmatrix} \qquad \det(A) = 1.24$$

$$B = \begin{bmatrix} \frac{3}{2} & 3 \\ -\frac{1}{2} & \frac{5}{2} \end{bmatrix} \qquad \det(B) = 5.25$$

The red and blue images are called *parallelograms*. In general, the image of the corners of the unit n-cube, when multiplied by an $n \times n$ matrix, is called a *parallelotope*.

Outr first definition of the determinant is a recursive definition, which justifies the first two examples in Example 18.1.

Definition 18.5 (Recursive definition): 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. In this setting,

- the number $\det(A^{ij})$ is called the ij-*minor* of A,
- the number $(-1)^{i+j}\det(A^{ij})$ is called the ij-*cofactor* of A.

The $n \times n$ matrix with ij-entry the ij-cofactor is called the *cofactor matrix* $\mathrm{cofac}(A)$ of A.

Example 18.6. Following Definition 18.5, we compute the determinant of a matrix A, using $i = 1$:

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

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

Inquiry 18.7 (✠3.01): Let $A \in \mathcal{M}_{n\times n}$. Using the recursive Definition 18.5 of a determininant, show that the following statements are true, for any $n \in \mathbf{N}$.

1. The determinant of the $n \times n$ identity matrix is 1. That is, $\det(I_n) = 1$.
2. The determinant of an upper (or lower) triangular matrix is the product of the diagonal entries.

Hint: use induction for both!

Next we describe some general properties of the determinant.

Proposition 18.8. Let $A \in \mathcal{M}_{n\times n}$. As a function of the rows of A, the determinant is:

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

Proof. The first point follows by induction on n, and by using the recursive definition (Definition 18.5) to expand along row i. The statement is immediately true for a 1×1 matrix. For the inductive step, notice that

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

and A^{ij} is the same in both cases.

The second point follows by using the combinatorial definition of the determinant (Definition 18.13). Fix two different indices $i, j \in \{1, 2, \dots, n\}$. For every permutation σ on a set of size n, let σ' be the permutation given by

$$\sigma'(k) = \begin{cases} \sigma(k) & k \neq i, j,\\ \sigma(j) & k = i,\\ \sigma(i) & k = j.\end{cases}$$

That is, σ' is the same as σ, except it swaps the images of i and j. Note that $\mathrm{sgn}(\sigma') = -\mathrm{sgn}(\sigma)$, since σ' has one row swap that σ does not have. Now suppose that for a matrix A, the matrix A' is the same, except with rows i and j swapped. Then

$$\begin{aligned}\det(A') &= \sum_{\text{permutations } \sigma} \mathrm{sgn}(\sigma){A'}_{1\sigma(1)}{A'}_{2\sigma(2)}\cdots{A'}_{i\sigma(i)}\cdots{A'}_{j\sigma(j)}\cdots{A'}_{n\sigma(n)} && \text{(definition of det)}\\ &= \sum_{\text{permutations } \sigma} \mathrm{sgn}(\sigma)A_{1\sigma(1)}A_{2\sigma(2)}\cdots{A'}_{i\sigma(i)}\cdots{A'}_{j\sigma(j)}\cdots A_{n\sigma(n)} && \text{(definition of } A')\\ &= \sum_{\text{permutations } \sigma} \mathrm{sgn}(\sigma)A_{1\sigma(1)}A_{2\sigma(2)}\cdots A_{i\sigma(j)}\cdots A_{j\sigma(i)}\cdots A_{n\sigma(n)} && (A_i = {A'}_j \text{ and } A_j = {A'}_i)\\ &= \sum_{\text{permutations } \sigma'} \mathrm{sgn}(\sigma)A_{1\sigma'(1)}A_{2\sigma'(2)}\cdots A_{i\sigma'(i)}\cdots A_{j\sigma'(j)}\cdots A_{n\sigma'(n)} && \text{(definition of } \sigma')\\ &= -\sum_{\text{permutations } \sigma'} \mathrm{sgn}(\sigma')A_{1\sigma'(1)}A_{2\sigma'(2)}\cdots A_{i\sigma'(i)}\cdots A_{j\sigma'(j)}\cdots A_{n\sigma'(n)} && \text{(property of } \sigma')\\ &= -\det(A). && \text{(definition of det)}\end{aligned}$$

□

Example 18.9. Consider the following example of multilinearity (on the left) and the alternating property (on the right):

$$\begin{aligned} -46 &= 11 \cdot (-2) - 8 \cdot 3 \\ &= \begin{vmatrix} 11 & 8 \\ 3 & -2 \end{vmatrix} \\ &= \begin{vmatrix} 6+5 & 9-1 \\ 3 & -2 \end{vmatrix} \\ &= 3\begin{vmatrix} 2 & 3 \\ 3 & -2 \end{vmatrix} + \begin{vmatrix} 5 & -1 \\ 3 & -2 \end{vmatrix} \\ &= 3\,(2 \cdot (-2) - 3 \cdot 3) + (5 \cdot (-2) - (-1) \cdot 3) \\ &= -39 - 7 \\ &= -46. \end{aligned}$$

$$\begin{aligned} \begin{vmatrix} -5 & 6 \\ -5 & 6 \end{vmatrix} &= (-5) \cdot 6 - (-5) \cdot 6 \\ &= -30 + 30 \\ &= 0 \end{aligned}$$

Inquiry 18.10 (3.02): There are two immediate consequences of Proposition 18.8. Show why they are both true using the proposition, for any $n \times n$ matrix.

1. A matrix with a zero row has determinant zero.
2. A matrix with two equal rows has determinant zero.

Hint: consider the determinant as a function of the rows, as in the proposition.

18.2 A combinatorial definition

We now consider the determinant in a combinatorial context, that is, as it relates to all permutations of the rows and columns of a matrix.

Definition 18.11: Let $S = (a_1, \ldots, a_n)$ be an ordered set. A *permutation* of S is equivalently

- 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.

Example 18.12. A permutation can be denoted in several different ways:

$$(1\ 2)(4\ 6\ 5) \qquad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 6 & 4 & 5 \end{pmatrix} \qquad \begin{array}{ccc} 1 & \mapsto & 2 \\ 2 & \mapsto & 1 \\ 3 & \mapsto & 3 \\ 4 & \mapsto & 6 \\ 5 & \mapsto & 4 \\ 6 & \mapsto & 5 \end{array}$$

all describe the same permutation. Moreover, $(1\ 2)(4\ 6\ 5)$ is the same as $(1\ 2)(6\ 5)(4\ 6)$, if we apply the 2-element permutations from right to left:

$$
\begin{array}{ccccc}
 & (4\ 5\ 6) & & (1\ 2) & \\
1 & \mapsto & 1 & \mapsto & 2 \\
2 & \mapsto & 2 & \mapsto & 1 \\
3 & \mapsto & 3 & \mapsto & 3 \\
4 & \mapsto & 6 & \mapsto & 6 \\
5 & \mapsto & 4 & \mapsto & 4 \\
6 & \mapsto & 5 & \mapsto & 5
\end{array}
\qquad\qquad
\begin{array}{ccccccc}
 & (4\ 6) & & (6\ 5) & & (1\ 2) & \\
1 & \mapsto & 1 & \mapsto & 1 & \mapsto & 2 \\
2 & \mapsto & 2 & \mapsto & 2 & \mapsto & 1 \\
3 & \mapsto & 3 & \mapsto & 3 & \mapsto & 3 \\
4 & \mapsto & 6 & \mapsto & 6 & \mapsto & 6 \\
5 & \mapsto & 5 & \mapsto & 4 & \mapsto & 4 \\
6 & \mapsto & 4 & \mapsto & 5 & \mapsto & 5
\end{array}
$$

On a set of size n there are $n!$ permutations and $n(n-1)/2$ transpositions. They are related to each other, but in a difficult to prove way.

Theorem 18.12.1. *Every permutation on a set of n elements may be uniquely (up to rearrangement) described as a composition of transpositions.*

This is a nontrivial fact and we do not prove it here.

Definition 18.13 (Combinatorial definition): Let $A \in \mathcal{M}_{n\times n}$, and let σ be a permutation on a set of size n.

- The *parity* of σ is "odd" or "even," depeding on if the number of transpositions necessary to represent it is odd or even, respectively.
- The *sign* of σ is $+1$ if the parity of σ is even, and -1 is the parity of σ is odd. This number is denoted by $\operatorname{sgn}(\sigma)$.
- The *determinant* of a matrix A is a sum over all permutations of the columns of A:

$$\det(A) = \sum_{\text{permutations } \sigma} \operatorname{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)} = \sum_{\text{permutations } \sigma} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i\sigma(i)}. \tag{8}$$

This definition may seem much more technical than the recursive Definition 18.5, but it very useful in cases where direct construction is important.

Example 18.14. We use the combinatorial definition to compute the determinant of a 3×3 matrix:

$$
\begin{array}{ccccccccccc}
() & & (1\ 2) & & (1\ 3) & & (2\ 3) & & (1\ 2\ 3) & & (1\ 3\ 2) \\
\begin{bmatrix} \mathbf{2} & 5 & -1 \\ 3 & \mathbf{-2} & 6 \\ 1 & 0 & \mathbf{2} \end{bmatrix} & &
\begin{bmatrix} 2 & \mathbf{5} & -1 \\ \mathbf{3} & -2 & 6 \\ 1 & 0 & \mathbf{2} \end{bmatrix} & &
\begin{bmatrix} 2 & 5 & \mathbf{-1} \\ 3 & \mathbf{-2} & 6 \\ \mathbf{1} & 0 & 2 \end{bmatrix} & &
\begin{bmatrix} \mathbf{2} & 5 & -1 \\ 3 & -2 & \mathbf{6} \\ 1 & \mathbf{0} & 2 \end{bmatrix} & &
\begin{bmatrix} 2 & \mathbf{5} & -1 \\ 3 & -2 & \mathbf{6} \\ \mathbf{1} & 0 & 2 \end{bmatrix} & &
\begin{bmatrix} 2 & 5 & \mathbf{-1} \\ \mathbf{3} & -2 & 6 \\ 1 & \mathbf{0} & 2 \end{bmatrix} \\
-8 & - & 30 & - & 2 & - & 0 & + & 30 & + & 0
\end{array}
$$

Hence $\det(A) = -10$.

Inquiry 18.15 (3.02): Recall from Definition 3.9 that elementary matrices are either *permutation* (swaps rows), *elimination* (adds multiples of rows), or *diagonal* (multiplies rows by a number) matrices.

1. What is the determinant of any elimination matrix?
2. What is the determinant of any diagonal matrix?
3. What is the determinant of a permutation matrix that swaps two rows? What about three, four rows? Start with some small examples to see what happens.

Convince yourself that permutation matrices with an *odd* number of row swaps have determinant -1, and permutation matrices with an *even* number of row swaps have determinant 1. This is the concept of *parity*.

Remark 18.16. Every term of the combinatorial definition of the determinant of $A \in \mathcal{M}_{n\times n}$ has

- exactly one factor in every row of A, and
- exactly one factor in every column of A.

All the different ways to choose n elements from A respecting both of these conditions gives all the differnt terms in the determinant.

Example 18.17. 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 element 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 (8) will have at least one factor of 0, so can be safely ignored.

18.3 Exercises

Exercise 18.1. (✠3.01) Show with a counterexample 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 18.2. (✠3.01) Recall the definition of an inverse of a square matrix A, which was a matrix B such that $AB = BA = I$. Show that the statement $AB = I$ implies $BA = I$.
Hint: Use the fact that a matrix having an inverse is the same as the matrix having nonzero determinant.

Exercise 18.3. (✠3.01) Let A be a 3×3 matrix. Suppose that $\det(A) = k$.

1. Use the multilinearity property of the determinant to compute $\det(A + A)$.
2. Use the multilinearity property of the determinant to compute $\det(-A)$.
   Hint: Use the fact that $-A = A - 2A$.
3. Explain how the result for part (b) would be different if A was a 2×2 matrix.

Exercise 18.4. (✠3.02) 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 $A\mathbf{x} = 0$.

Exercise 18.5. (✠3.01) 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 18.6. (✠3.01) 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 $A = \begin{bmatrix} 7 & 0 & -1 & 0 \\ 3 & 0 & 2 & 0 \\ 0 & -2 & 6 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Exercise 18.7. (✠3.01) Use the combinatorial definition of the determinant for this question. Recall from Example 19.15 that every term in the combinatorial definition uses exactly one entry from each row and one entry from each column of the matrix.

1. Compute the determinant of
$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}.$$
2. The $n \times n$ identity matrix has $n^2 - n$ zeroes and exactly one nonzero term in the combinatorial definition. What is the smallest number of zeroes an $n \times n$ matrix can have so that the combinatorial definition has only one nonzero term?

Lecture 19: Properties of the determinant

Chapters 5.2,5.3 in Strang's "Linear Algebra"

- Fact 1: The determinant of a product is the product of determinants: $\det(AB) = \det(B)\det(A)$
- Fact 2: The determinant is the product of the pivots, up to a sign change.
- Fact 3: The determinant is nonzero iff the matrix is invertible.

✠ Standard 3.03: Compute determinants of products, inverses, transposes of matrices.

✠ Standard 3.04: Prove simple properties of the determinant.

This lecture explores some properties of the determinant.

19.1 Splitting the determinant

We begin by showing that the determinant is *multiplicative*, that is, that $\det(AB) = \det(A)\det(B)$ for any $n \times n$ matrices A, B. First we need to revisit elementary matrices in Definition 3.9.

Lemma 19.1. Let $A \in \mathcal{M}_{n\times n}$ be an invertible matrix. That is, A^{-1} exists.

- If P is a permutation matrix of a single row swap, then $\det(PA) = \det(P)\det(A) = -\det(A)$.
- If E is an elimination matrix, then $\det(EA) = \det(E)\det(A) = \det(A)$.
- If D is a diagonal matrix, then $\det(DA) = \det(D)\det(A)$.

Proof. The first point follows from the alternating property from Proposition 18.8 and the third point of Inquiry 18.15.

The second point follows by multilinearity from Proposition 18.8 and the first point ofInquiry 18.15, which gives that $\det(E) = 1$. Elimination matrices are row operations, so in terms of A and the rows $\mathbf{r}_1, \dots, \mathbf{r}_n$ of A,

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

The third point follows by the second point of Inquiry 18.7, which says that $\det(D)$ is the product of its diagonal entries, and by the recursive definition of the determinant. If D has all ones on the diagonal except on row i, then

$$\begin{aligned}\det(DA) &= \sum_{j=1}^{n} (-1)^{i+j} (DA)_{ij} \det((DA)^{ij}) \\ &= D_{ii} \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det((DA)^{ij}) \\ &= D_{ii} \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det(A^{ij}) \\ &= D_{ii} \det(A) \\ &= \det(D)\det(A).\end{aligned}$$

If D has more than one diagonal entry that is not 1, repeat this step for every such row. □

Inquiry 19.2 (✠3.03, 3.04): Let $A \in \mathcal{M}_{3\times 3}$. Use Lemma 19.1 for these tasks.

1. Suppose that row operations turn A into $\begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$. Explain why A has determinant zero.
2. Suppose that A has 3 pivots. Explain why A has a nonzero determinant.

Let A be the elimination matrix $\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

3. Compute the inverse B of A.
4. What is $\det(A)$ and what is $\det(B)$?

The above inquiry sets the scene for the following inquiry and the proposition afterward.

Inquiry 19.3 (✠3.04): Let $A, B \in \mathcal{M}_{n\times n}$, not necessarily related matrices.

1. Suppose that $\det(A) = 0$. Show by contradiction that $\det(AB)$ must also be 0.
2. Suppose that $\det(A) \neq 0$ and $\det(B) \neq 0$. Show that $\det(AB) = \det(A)\det(B)$.

For the second point, convince yourself that A having n pivots means A can be expressed as a product of elementary matrices.

We conclude this section with a strong relationship among some big concepts we have seen so far: pivots, invertibility, and the determinant.

Proposition 19.4. Let $A \in \mathcal{M}_{n\times n}$.

- The determinant of A is the product of the pivots of A, up to a sign change.
- The determinant of A is nonzero if and only if A has n pivots.
- The determinant is zero if and only if A is not invertible.

Proof. The first point follows from the second point of Inquiry 19.3. The second point is a direct consequence of the first point. The third point follow from both points of Inquiry 19.3. □

19.2 Inverses and transposes

Now we take a look at how the determinant works with transposes and inverses.

Proposition 19.5. Let $A \in \mathcal{M}_{n\times n}$ be invertible (that is, have nonzero determinant).

- The determinant of the tranpose is the same as the determinant: $\det(A^T) = \det(A)$
- The determinant of the inverse is the reciprocal of the determinant: $\det(A^{-1}) = \det(A)^{-1}$

Proof. The first statement follows by using the fact that if A is invertible, then it may be expressed as the product of elementary matrices. Using the properties of the transpose (after Definition2.11, the transpose of the product is the (reversed) product of the individual factors. Finally by applying multiplicativity of the determinant, we get back the original matrix A.

The second statement follows from Proposition 19.1 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 from Definition 18.5 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 19.6. 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 and $\mathbf{c}_i$ 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. Inquiry 18.10 told us that a matrix with two equal rows has determinant zero. Hence

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

or $AC^T = \det(A)I$. ◻

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 19.7. 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 10.7 in Lecture 10.

Definition 19.8: 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 19.9. 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 19.10. 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*: $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$

Inquiry 19.11 (✠3.03): This inquiry is about the cross product.

1. Compute the cross product of the two basis vectors for the plane defined by $z = 10x - 2y$ (see Example 11.2).
2. Compare your answer above with a normal vector to this plane. Are the two vectors the same? Are they similar?
3. You should have four vectors from the two points above. Explain why their span can be expressed using at most three vectors.

19.3 Exercises

Exercise 19.1. (✠3.03) Let $a, b, c, d \in \mathbf{R}$. Using elementary matrices (permutation, elimination, diagonal) to bring these matrices to triangular form, compute their determinants.

$$A = \begin{bmatrix} 0 & a & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & c \\ d & 0 & 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} a & b & a \\ a & c & a \\ a & d & a \end{bmatrix} \qquad C = \begin{bmatrix} a & b & c \\ b & 0 & b \\ c & b & a \end{bmatrix}$$

Exercise 19.2. (✠3.04) 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 19.3. (✠3.04) Let A be an $n \times n$ matrix, for some $n \in \mathbf{N}$.

1. Explain why $\det(kA) = k^n \det(A)$, for any real number k.
2. If A is skew-symmetric, explain why randomly choosing n in the range $[1, 100]$ means $\det(A) = 0$ exactly half of the time.

3. Suppose that A is a projection matrix, projecting from $\mathbf{R}^n$ to an $(n-1)$-dimensional subspace of $\mathbf{R}^n$. Explain why $\det(A) = 0$.

Lecture 20: Defining eigenvalues and eigenvectors

Chapter 6.1 in Strang's "Linear Algebra"

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

✠ Standard 3.05: Find eigenvectors and eigenvalues of a matrix

✠ Standard 3.06: Given only eigenvalues and eigenvectors of A, compute $A\mathbf{x}$ for any $\mathbf{x}$

✠ Standard 3.07: Given only eigenvalues and eigenvectors, construct a matrix with these eigenvalues and eigenvectors

---

This lecture gets to the heart of the current topic of *eigensystems*. Eigenvalues are important to undrstand what a matrix does to vectors. Eigenvectors are unique in that their direction does not change when multiplied by a matrix A (though their length may change).

20.1 Words beginning with "eigen"

It is important to remember that eigenvalues are unique, but eigenvectors are not, as they can be multiplied by any real number.

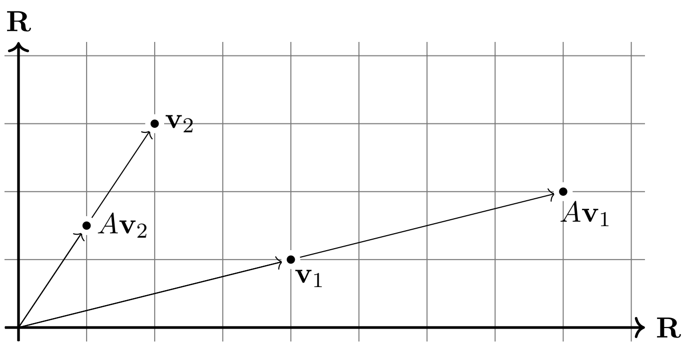
Definition 20.1: Let $A \in \mathcal{M}_{n\times n}$. For every vector $\mathbf{v}$ with $A\mathbf{v} = \lambda\mathbf{v}$, where $\lambda \in \mathbf{R}$,

- the vector $\mathbf{v}$ is called an *eigenvector*,
- the value λ is called the *eigenvalue*,
- the pair $(\mathbf{v}, \lambda)$ is called an *eigenpair*.

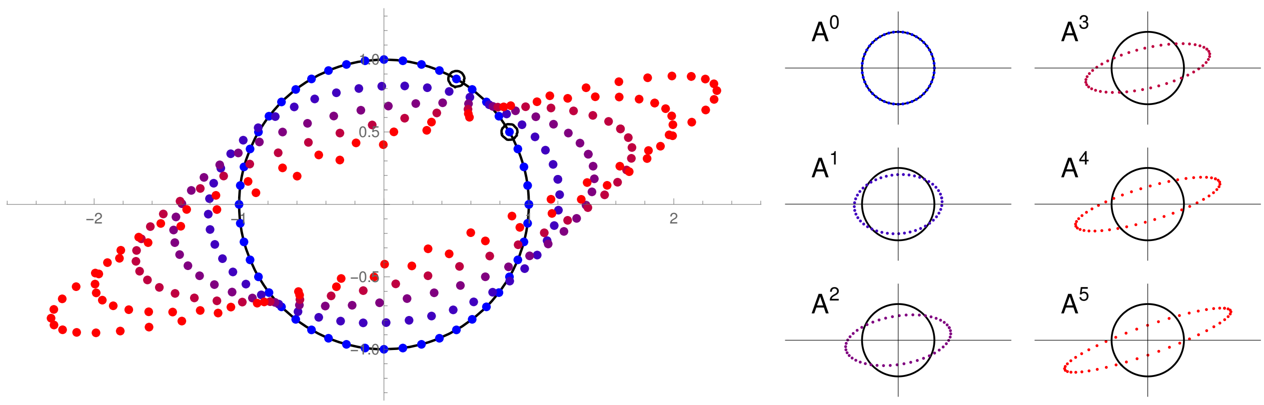
The set of all eigenvalues of A is called the *spectrum* of A. The set of all eigenpairs whose eigenvectors are linearly independent is called the *eigensystem* of A. Eigensystems are unique up to vector scaling.

Eigenvectors describe the direction in which a matrix changes $\mathbf{R}^n$, and the eigenvalues describe the stretching that is done in that direction. Although the zero vector $\mathbf{v} = 0$ satisfies the above definition for any number λ, it is not considered an eigenvector, since it does not give any information about the matrix.

Example 20.2. In $\mathbf{R}^2$, the matrix $A = \begin{bmatrix} 23/10 & -6/5 \\ 9/20 & 1/5 \end{bmatrix}$ has eigenvector $\mathbf{v}_1 = \left[\begin{smallmatrix} 4 \\ 1 \end{smallmatrix}\right]$ with eigenvalue 2, and eigenvector $\mathbf{v}_2 = \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$ 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 20.8 we see what happens to vectors that are not exactly an eigenvector.

Example 20.3. Consider the following examples of eigenvectors and eigenvalues.

- The matrix $A = \left[\begin{smallmatrix} 4 & -2 \\ 0 & 2 \end{smallmatrix}\right]$ has eigenvector $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ with eigenvalue 2. But A also has eigenvalue $\left[\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right]$ with eigenvalue 2.
- The matrix $B = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$ 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 13) 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.

Inquiry 20.4 (✠3.05): Let $\mathbf{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right], \mathbf{w} = \left[\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right] \in \mathbf{R}^2$ be fixed.

1. If $A \in \mathcal{M}_{2\times 2}$ with $A\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] = \mathbf{v}$ and $A\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right] = \mathbf{w}$, what is the determinant of A?
2. Suppose that there is $B \in \mathcal{M}_{2\times 2}$ with $\mathbf{v}, \mathbf{w}$ as eigenvectors. In what cases will the vector $\mathbf{v} + \mathbf{w}$ be an eigenvector for B?
3. Must there always exist a 2×2 matrix with $\mathbf{v}$ and $\mathbf{w}$ as eigenvectors? That is, knowing only $\mathbf{v}$ and $\mathbf{w}$, can you construct a 2×2 matrix with these as eigenvectors?

20.2 The characteristic polynomial

So far we have seen just examples of eigenvalues and eigenvectors, but not yet a procedure for finding them. We describe this procedure now.

Definition 20.5: Let $A \in \mathcal{M}_{n\times n}$. The *characteristic polynomial* of A is

$$\chi(t) = \det(A - tI). \tag{9}$$

The roots λ_i of the characteristic polynomial are the eigenvalues of A. The multiplicity of each root λ_i is its *algebraic multiplicity*.

Once the roots $\lambda_1, \dots, \lambda_k$ of χ are found, then $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ can be solved in each coordinate to find the corresponding eigenvector $\mathbf{v}_i$.

Example 20.6. 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:

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

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 $\left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$.

Inquiry 20.7 (✠3.07): Consider the vectors $\mathbf{v} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right], \mathbf{w} = \left[\begin{smallmatrix} -1 \\ 3 \end{smallmatrix}\right] \in \mathbf{R}^2$.

1. Construct a 2×2 matrix A that has $\mathbf{v}$ as an eigenvector.
2. What is the determinant of A? What does that say about its other eigenvalue?
3. Construct a 2×2 matrix B that has $\mathbf{v}$ as an eigenvector with eigenvalue 2 and $\mathbf{w}$ as an eigenvector with eigenvalue 3.
4. Compute the determinant and trace of B.

Remark 20.8. 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 20.6 we found two eigenvalues and two eigenvectors. Then for any other vector we have

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

Definition 20.9: Let $A \in \mathcal{M}_{n\times n}$. For every eigenvalue λ,

- the number of linearly independent eigenvectors with λ as their eigenvalue is the *geometric multiplicity* of λ,
- the span of these linearly independent eigenvectors is the *eigenspace* of λ.

In other words, if A has k eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ in its eigensystem, then the number of $\mathbf{v}_i$ with eigenvalue λ is the geometric multiplicity of λ.

Inquiry 20.10 (✠3.05): Consider the matrix $A = \begin{bmatrix} 0&1&0\\ 1&0&0\\ 0&1&1 \end{bmatrix}$.

1. Compute the eigensystem of A.
2. What are the eiegenspaces of A?
3. Explain the relationship between the dimension of an eigenspace and its geometric multiplicity.

20.3 Exercises

Exercise 20.1. (✠3.05) Consider the matrix $A = \begin{bmatrix} 6&-5\\ 5&-2 \end{bmatrix}$.

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 20.2. (✠3.05) Let $\mathbf{v} \in \mathbf{R}^n$ be a unit vector, and let $A = \mathbf{v}\mathbf{v}^T$.

1. Show that A is a projection matrix.
2. Show that $\mathbf{v}$ is an eigenvector of A and find its eigenvalue.
3. Show that if $\mathbf{u} \perp \mathbf{v}$, then $A\mathbf{u} = 0$.
4. How many independent eigenvectors does A have with eigenvalue 0?

Exercise 20.3. (✠3.07) 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 20.4. (✠3.05) 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

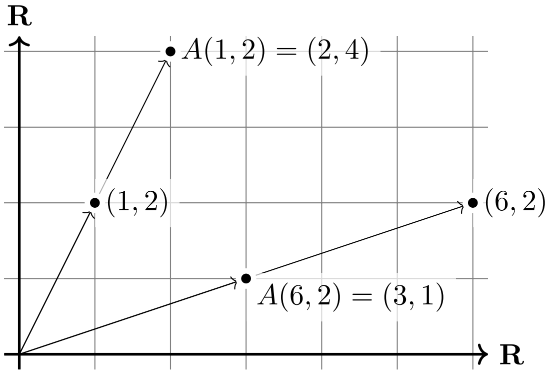
$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & -2 & 0 & 0 \\ -2 & 2 & 0 & 0 & 0 & 7 \\ 8 & 0 & 3 & 0 & 8 & 0 \\ 0 & 0 & 0 & 9 & -9 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -5 & 2 & 1 \end{bmatrix}.$$

Exercise 20.5. (✠3.07) Let λ, μ be real numbers, and $\mathbf{u} = \begin{bmatrix} x\\ y \end{bmatrix}, \mathbf{v} = \begin{bmatrix} z\\ w \end{bmatrix} \in \mathbf{R}^2$ be two vectors.

1. Construct a 2×2 matrix with eigenpairs $(\mathbf{u}, \lambda)$ and $(\mathbf{v}, \mu)$.

2. What assumptions did you make in the first part to reach a conclusion?

Exercise 20.6. (✠3.06, 3.07) Let $A\colon \mathbf{R}^2 \to \mathbf{R}^2$ be the 2×2 matrix for which $A\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right] = \left[\begin{smallmatrix}2\\4\end{smallmatrix}\right]$ and $A\left[\begin{smallmatrix}6\\2\end{smallmatrix}\right] = \left[\begin{smallmatrix}3\\1\end{smallmatrix}\right]$. This is described in the picture below.



1. What is the eigensystem of A?
2. Express $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ as linear combinations of the eigenvectors of A.
3. Compute $A\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $A\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$. Use this to construct the matrix of A.
4. Using eigenvalues, explain why A is invertible.

Lecture 21: Properties of eigenvalues and eigenvectors

Chapters 6.1 and 6.2 in Strang's "Linear Algebra"

- Fact 1: The sum of the eigenvalues is the trace of the matrix.
- Fact 2: The product of the eigenvalues is the determinant of the matrix.
- Fact 3: An $n \times n$ matrix has exactly n eigenpairs, counting multiplicity.
- Fact 4: Eigenvalues may be zero. Eigenvectors cannot be the zero vector.

✠ Standard 3.08: Compute trace, determinant, eigensystems of special matrices.

✠ Standard 3.09: Diagonalize a matrix with linearly independent eigenvectors, and identify when it is not possible.

We continue understading the key properties of eigenvalues and eigenvectors.

21.1 Properties of eigensystems

Recall that the key idea of the eigensystem of a matrix $A \in \mathcal{M}_{n\times n}$ was that it explains how $\mathbf{R}^n$ is transformed, when A multiplies any vector in $\mathbf{R}^n$.

Definition 21.1: Let $A \in \mathcal{M}_{n\times n}$. If there are vectors $\mathbf{v}, \mathbf{w} \in \mathbf{R}^n$ for which there exists $\lambda, \mu \in \mathbf{R}$, such

- $A\mathbf{v} = \lambda\mathbf{v}$, then $\mathbf{v}$ is called an *eigenvector*, or *right eigenvector* of A,
- $\mathbf{w}^T A = \mu\mathbf{w}^T$, then $\mathbf{w}$ is called a *left eigenvector* of A.

Note that a right eigenvector of A is a left eigenvector of A^T.

If no adjective "right" or "left" is used, then "right" is assumed. The relationship between left and right eigenpairs is not immediate.

Inquiry 21.2 (✠3.08): Let $A \in \mathcal{M}_{n\times n}$.

1. Suppose that A is symmetric. If $(\mathbf{v}, \lambda)$ is an eigenpair for A, show that $\mathbf{v}$ is an eigenvector for $A^T A$. What is its eigenvalue?
2. Suppose that there are n distinct eigenpairs $(\mathbf{v}_i, \lambda_i)$ for A, with each eigenvector being both a right and a left eigenvector. Show that $AA^T = A^T A$.

Remark 21.3. Let $A \in \mathcal{M}_{n\times n}$ have eigenvalues $\lambda_1, \ldots, \lambda_n$ (not all necessarily distinct). The characteristic polynomial can then be expressed as

$$\chi(t) = (-1)^n (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n).$$

This follows from the definition of the characteristic polynomial and the recursive definition of the determinant. The coefficient $(-1)^n$ comes from the fact that $-t$ is multiplied by itself n times, and so the leading term must be $(-1)^n t^n$.

Proposition 21.4. Let $A \in \mathcal{M}_{n\times n}$.

- The eigenvalues of A and A^T are the same, but not necessarily their eigenvectors.

- If A is upper or lower triangular, its eigenvalues are on its diagonal.
- If the rank of A is less than n, then A has an eigenvalue 0 for a non-trivial eigenvector.
- If A has an eigenpair $(\mathbf{v}, \lambda)$, then A^n has an eigenpair $(\mathbf{v}, \lambda^n)$.

Proof. The first point follows by distributing transposes in a sum (see Remark 5.14) 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. □

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.

Inquiry 21.5 (✠3.09): Recall that the characteristic polynomial of $A \in \mathcal{M}_{n\times n}$ is $\chi(t) = \det(A - tI)$. For this question, let $A = \left[\begin{smallmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{smallmatrix}\right]$.

1. The recursive definition of $\det(A - tI)$ has 6 terms in its expansion. Write all of these out, without expanding the $(a_{ij} - t)$ factors.
2. When the $(a_{ij} - t)$ factors are all expanded,
   - what is the coefficient of t^3?
   - what is the coefficient of t^2?
   - what is the coefficient of t?
   - what is the constant term?
3. Among the parts above, find the trace and the determinant.
4. Express the characteristic polynomial using the trace and the determinant.

How do you think this generalizes to higher $n \in \mathbf{N}$?

Sometimes we come across matrices (as in Example 20.3) that do not seem to have eigenvalues, such as $A = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$. Its characteristic polynomial is $\chi(t) = t^2 + 1$. This polynomial has no real solutions, but does have *complex* solutions.

Definition 21.6: 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:

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

Remark 21.7. Here are some key properties of the complex numbers .

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

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

Inquiry 21.8 (✠3.08): Let $A \in \mathcal{M}_{n\times n}$ be a skew-symmetric matrix.

1. Compute the eigensystem for $A = \begin{bmatrix} 0 & -5 \\ 5 & 0 \end{bmatrix}$. How many complex and how many real eigenvalues does A have?
2. Compute the eigensystem for $A = \begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & 3 \\ 2 & -3 & 0 \end{bmatrix}$. How many complex and how many real eigenvalues does A have? You may use a computer.
3. If n is an odd number, explain why A has at least one real root. *Hint: use limits.*
4. How many real and how many complex values will a skew-symmetric $n \times n$ matrix have? Begin by showing that $\|A\mathbf{v}\|^2 = -\lambda^2\|\mathbf{v}\|^2$ for any eigenpair $(\mathbf{v}, \lambda)$ of A.

More about complex numbers is discussed in Lecture 25.

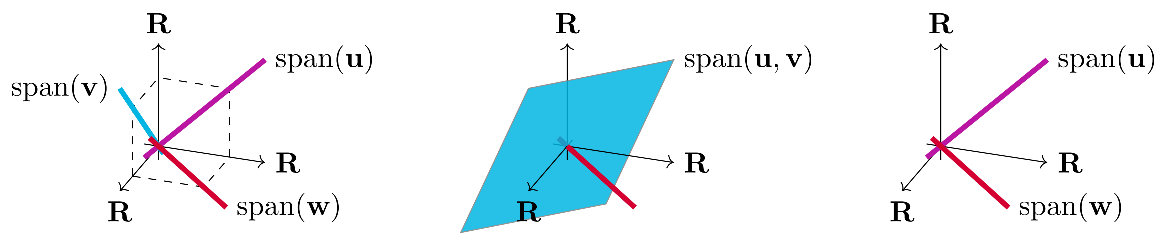
Proposition 21.9. Let $A, B \in \mathcal{M}_{n\times n}$.

- 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, $AB = BA$).

21.2 Multiplicity and diagonalization

We begin with considering several different possibilites of eigenpairs for a 3×3 matrix.

Example 21.10. Consider the three linearly independent vectors $\mathbf{u} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$



eigenspaces of A eigenspaces of B eigenspaces of C

These three vectors may appear in six different ways as eigenvectors of a 3×3 matrix.

- $A = \begin{bmatrix} 1 & 5 & -5 \\ 2 & 4 & -2 \\ -3 & 3 & -1 \end{bmatrix}$ has 3 different eigenvalues, 3 different eigenvectors: $(2, \mathbf{u}), (-4, \mathbf{v}), (6, \mathbf{w})$
- $B = \begin{bmatrix} 1 & 5 & -5 \\ 5 & 1 & -5 \\ 0 & 0 & -4 \end{bmatrix}$ has 2 different eigenvalues, 3 different eigenvectors: $(-4, \mathbf{u}), (-4, \mathbf{v}), (6, \mathbf{w})$
- $C = \begin{bmatrix} 1 & 5 & -5 \\ 6 & 0 & -4 \\ 1 & -1 & -3 \end{bmatrix}$ has 2 different eigenvalues, 2 different eigenvectors: $(-4, \mathbf{u}), (6, \mathbf{w})$
- $D = \begin{bmatrix} -4 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -4 \end{bmatrix}$ has 1 eigenvalue, 3 different eigenvectors: $(-4, \mathbf{u}), (-4, \mathbf{v}), (-4, \mathbf{w})$
- $E = \begin{bmatrix} -5 & 1 & 1 \\ 0 & -4 & 0 \\ -1 & 1 & -3 \end{bmatrix}$ has 1 eigenvalue, 2 different eigenvectors: $(-4, \mathbf{v}), (-4, \mathbf{w})$
- $F = \begin{bmatrix} -3 & 1 & -1 \\ 2 & -4 & 0 \\ 3 & 1 & -5 \end{bmatrix}$ has 1 eigenvalue, 1 eigenvector: $(-4, \mathbf{u})$

For B and C, $\lambda = -4$ has algebraic multiplicity 2. For D, E, F, it has algebraic multiplicity 3.

Inquiry 21.11 (✠3.09): Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^3$ be distinct vectors. Explain why the following situations each cannot happen. Justify your reasoning with the matrix equation $A\mathbf{v} = \lambda\mathbf{v}$, for an eigenpair $(\lambda, \mathbf{v})$.

1. A is a 2×3 matrix with eigensystem $\{(1, \mathbf{u}), (2, \mathbf{v})\}$.
2. B is a 3×3 matrix with determinant zero and eigensystem $\{(1, \mathbf{u}), (2, \mathbf{v})\}$.
3. C is a 3×3 matrix with trace zero and eigensystem $\{(1, \mathbf{u}), (2, \mathbf{v})\}$.
4. D is a 3×3 matrix with eigensystem $\{(1, \mathbf{u}), (2, \mathbf{u})\}$.
5. E is a 3×3 matrix with eigensystem $\{(1, \mathbf{u}), (2, \mathbf{v}), (3, \mathbf{u} + \mathbf{v})\}$.
6. F is a 3×3 matrix with eigensystem $\{(0, \mathbf{u}), (0, \mathbf{v}), (1, \mathbf{w})\}$ and a 2-dimensional column space.

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

Example 21.12. Let $\mathbf{u} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$, $\mathbf{v} = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] \in \mathbf{R}^2$, which are linearly indepedent vectors. Let $A = \left[\begin{smallmatrix}a & b\\c & d\end{smallmatrix}\right]$ be a matrix with these two as eigenvectors, and corresponding eigenvalues $2, 3$, respectively. What are the entries a, b, c, d of A? We know that

$$\begin{bmatrix}a & b\\c & d\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}2\\2\end{bmatrix}, \quad \begin{bmatrix}a & b\\c & d\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\3\end{bmatrix}, \quad \Longrightarrow \quad \begin{aligned}a+b &= 2\\ c+d &= 2\\ b &= 0\\ d &= 3\end{aligned} \quad \Longleftrightarrow \quad \begin{bmatrix}a & b\\c & d\end{bmatrix} = \begin{bmatrix}2 & 0\\-1 & 3\end{bmatrix}.$$

The equations $A\mathbf{u} = 2\mathbf{u}$ and $A\mathbf{v} = 3\mathbf{v}$ on the left, which can be combined into a single equation

$$A\underbrace{\begin{bmatrix}| & |\\ \mathbf{u} & \mathbf{v}\\ | & |\end{bmatrix}}_{X} = \begin{bmatrix}| & |\\ 2\mathbf{u} & 3\mathbf{v}\\ | & |\end{bmatrix} = \begin{bmatrix}| & |\\ \mathbf{u} & \mathbf{v}\\ | & |\end{bmatrix}\underbrace{\begin{bmatrix}2 & 0\\0 & 3\end{bmatrix}}_{\Lambda} \quad \Longrightarrow \quad A = \underbrace{\begin{bmatrix}| & |\\ \mathbf{u} & \mathbf{v}\\ | & |\end{bmatrix}}_{X}\underbrace{\begin{bmatrix}2 & 0\\0 & 3\end{bmatrix}}_{\Lambda}\underbrace{\begin{bmatrix}| & |\\ \mathbf{u} & \mathbf{v}\\ | & |\end{bmatrix}^{-1}}_{X^{-1}}$$

The inverse of X can be constructed because $\mathbf{u}, \mathbf{v}$ are linearly independent, so X has rank 2

Definition 21.13: A matrix $A \in \mathcal{M}_{n\times n}$ is *diagonalizable* if it has n linearly independent eigenvectors. If A is diagonalizable, then the *diagonalization* of A is the decomposition of A as the product

$$A = X\Lambda X^{-1}, \tag{10}$$

for Λ a diagonal matrix and $(\Lambda_{ii}, \mathbf{x}_i)$ an eigenpair of A, for every $i = 1, \dots, n$. The vector $\mathbf{x}_i$ is the ith column of X.

Remark 21.14. The matrix X is not unique, as its columns (the eigenvectors of A) may be scaled by any real number. 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 21.12, we could have the eigenvectors $5\mathbf{u}$ and $-7\mathbf{v}$ instead of just $\mathbf{u}$ and $\mathbf{v}$. In that case,

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

and the decomposition in that case is

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

$$= \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix} \Lambda \begin{bmatrix} 5 & 0 \\ 0 & -7 \end{bmatrix} \begin{bmatrix} 1/5 & 0 \\ 0 & -1/7 \end{bmatrix} \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix} \Lambda \begin{bmatrix} | & | \\ \mathbf{u} & \mathbf{v} \\ | & | \end{bmatrix}^{-1},$$

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 21.15. 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 *algebraic 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 $[\,x\ y\ z\ w\,]^T$ 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.

21.3 Exercises

Exercise 21.1. (✠3.08) Let $A \in \mathcal{M}_{n\times n}$ and let $\chi(t)$ be its characteristic polynomial.

1. Show that $\chi(0) = (-1)^n \det(A)$. That is, show that the constant term in $\chi(t)$ is $(-1)^n$ times the determinant of A.
2. Show that the coefficient of t^{n-1} in $\chi(t)$ is $-\text{trace}(A)$.

Exercise 21.2. (✠3.08) Let A, B, C be any 3×3 matrices, with C diagonalizable.

1. Show that $\text{trace}(AB) = \text{trace}(BA)$.
2. Use that above to show that $\text{trace}(C)$ is the sum of the three eigenvalues of C.
Hint: Split up the diagonalization of C into two matrices.
3. Suppose that the eigenvalues of C are $1, \frac{1}{2}, \frac{1}{3}$. Show why the limit $\lim_{n\to\infty} C^n$ exists, and why it has rank 1.

Exercise 21.3. (✠3.09) Decompose 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 21.4. (✠3.09) Let $A \in \mathcal{M}_{3\times 3}$ with the eigenvectors $\left[\begin{smallmatrix} 1 \\ 2 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} -1 \\ -1 \\ 0 \end{smallmatrix}\right]$ 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 21.5. (✠3.09) 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.

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

Lecture 22: Diagonalizability and special matrices

Chapters 6.2, 6.4, 6.5 in Strang's "Linear Algebra"

- Fact 1: A matrix being invertible and diagonalizable are not the same. These properties are preserved under similarity.
- Fact 1: Symmetric matrices have orthonormal eigenvectors.
- Fact 2: Positive definiteness can be expressed in terms of pivots, eigenvalues, determinants, and matix or vector multiplications.

✠ Standard 3.10: Given a matrix A, construct and identify matrices similar to A.

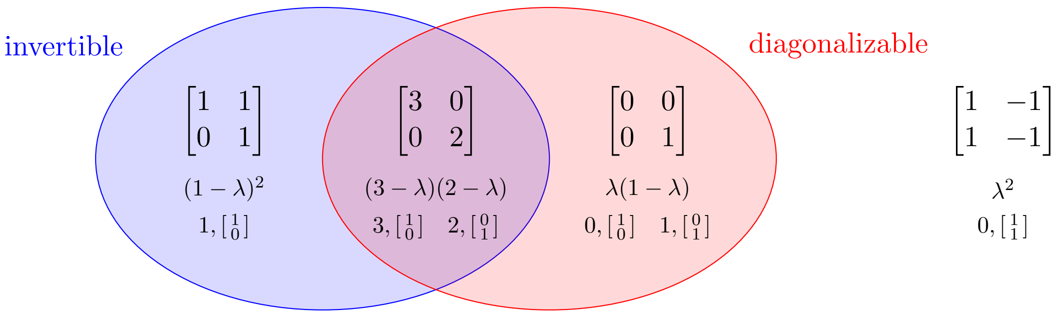
✠ Standard 3.11: Identify symmetric and positive definite matrices, directly and indirectly.

✠ Standard 3.12: Express a symmetric matrix as a sum of rank one matrices.

The goal of this section is to show how diagonalizability works with *symmetric* and *positive definite* matrices. The diagonal matrix Λ is easier to deal with, because it acts like a number rather than a matrix. We will see that for any matrix $A \in \mathcal{M}_{m\times n}$, the matrices $A^TA \in \mathcal{M}_{n\times n}$ and $AA^T \in \mathcal{M}_{m\times m}$ are both symmetric and positive definite.

22.1 Invertibility and similarity

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.

> **Inquiry 22.1 (✠3.10):** Let $A \in \mathcal{M}_{3\times 3}$, and suppose that A has 3 different eigenvalues.
>
> 1. Explain why A must have 3 linearly independent eigenvectors.
>   *Hint: Show this by contradiciton, assuming that two eigenvectors are linearly independent, and the third is a linear combination of the first two.*
> 2. If none of the eigenvalues are zero, explain why A is invertible. What happens if one of the eigenvalues is zero?
>   *Hint: Use the diagonalization equation.*
> 3. Convince yourself that the statement generalizes to any $n \in \mathbf{N}$.

Remark 22.2. 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.
- 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 $\lim_{k\to\infty} A^k\mathbf{x} = 0$ 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* $\lambda^* = a - bi$:

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

Example 22.3. 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

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

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

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

Definition 22.4: Let $A, B, C \in \mathcal{M}_{n\times n}$ with C invertible. The matrices A and B are *similar* if $A = CBC^{-1}$.

As mentioned in Remark 22.2, similar matrices have the same eigenvalues with the same algebraic multiplicity, but may have different eigenvectors.

22.2 Symmetric matrices

Recall from Definition 5.2 in Lecture 3 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 22.5 (The Spectral Theorem)**.** Let $A \in \mathcal{M}_{n\times n}$. If A is symmetric, then A has n real eigenvalues and n orthogonal eigenvectors.

This implies that a symmetric matrix can always be diagonalized. Symmetric matrices will often be written "S".

Inquiry 22.6 (✠3.10): Consider the symmetric matrix $S = \left[\begin{smallmatrix}1&2&0\\2&1&2\\0&2&1\end{smallmatrix}\right]$.

1. Compute the matrices X, Λ for the diagonalization of S.
2. Find the matrix B for which BX has orthonormal columns.
3. Consider the matrix X' which is the same as X, but with the first two columns swapped. Explain why $X'\Lambda'(X')^{-1}$ is still equal to S. As with X, here Λ' is the same as Λ, but with the first two columns swapped.
4. Does column swapping as in the point above work for any symmetric matrix, or only for this particular S?

Example 22.7. 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

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

These vectors are orthogonal as $\mathbf{v}_1 \bullet \mathbf{v}_2 = 0$. 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 22.8. 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 Lecture 23.

We finish off the first part of this lecture with another comment about the relationship between pivots and eigenvalues.

Remark 22.9. 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.

Inquiry 22.10 (✠3.11): Consider the symmetric matrix $S = \left[\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{smallmatrix}\right]$. For this inquiry, you may use a computer.

1. Compute the eigensystem of S.
2. As given, S is not invertible. What is the smallest number of entries in S that should be changed to keep it symmetric and to make it invertible? Give an example.
3. The eiegnsystem of S contains only integers. Change at least one entry of S to keep it symmetric and to keep the eigensystem with only integers.
   Hint: This can be done by changing entries to 0 or -1.

22.3 Positive definite matrices

The second part of this lecture focuses on special types of symmetric matrices.

Definition 22.11: Let $S \in \mathcal{M}_{n\times n}$ be symmetric. The matrix S is *positive definite* if, equivalently,

- all eigenvalues of S are positive
- $\mathbf{v}^T S\mathbf{v} > 0$ for any nonzero $\mathbf{v} \in \mathbf{R}$.

Weakening the conditions to $\lambda \geqslant 0$ and $\mathbf{v}^T S\mathbf{v} \geqslant 0$ means S is (*positive*) *semidefinite*.

Finding eigenvalues is computationally intensive for large matrices, so we use the relationship with pivots from Remark 22.9 to determine when eigenvalues are positive. This gives several quick ways to determine when a matrix is positive definite.

Example 22.12. 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 22.13. To see why the two definitions from Definition 22.11 are equivalent, consider an $n \times n$ positive definite matrix S with eigenvector $\mathbf{v}$ and positive eigenvalue λ. Then

$$S\mathbf{v} = \lambda\mathbf{v} \quad \Longrightarrow \quad \mathbf{v}^T S\mathbf{v} = \lambda\mathbf{v}^T\mathbf{v} = \lambda(v_1^2 + \cdots v_n^2) > 0.$$

Conversely, any $\mathbf{x} \in \mathbf{R}^n$ can be expressed as a linear combination $a_1\mathbf{v}_1 + \cdots a_n\mathbf{v}_n$ of the orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of S. Then by orthonormality of the eigenvectors,

$$\mathbf{x}^T S\mathbf{x} = (a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n)^T(a_1\lambda_1\mathbf{v}_1 + \cdots + a_n\lambda_n\mathbf{v}_n) = a_1^2\lambda_1\|\mathbf{v}_1\|^2 + \cdots + a_n^2\lambda_n\|\mathbf{v}_n\|^2 > 0.$$

Proposition 22.14. 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 $A^T A$ is positive definite.

Proof. The first point follows from distributing

$$\mathbf{x}^T(S + T)\mathbf{x} = \mathbf{x}^T S\mathbf{x} + \mathbf{x}^T T\mathbf{x}.$$

The second point comes from rewriting

$$\mathbf{x}^T(A^T A)\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = \|A\mathbf{x}\|^2 > 0.$$

□

The proof of the second claim implies that $A^T A$ (and also AA^T) is always positive semidefinite.

Inquiry 22.15 (✠3.11): This inquiry uses Definition 16.1 of an *inner product* from Lecture 16.

1. Let $S \in \mathcal{M}_{n\times n}$ be positive definite. Show that $\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^T S\mathbf{v}$ satisfies all the properties of an inner product on $\mathbf{R}^n$.
2. Let $A \in \mathcal{M}_{n\times n}$ be a matrix and $\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^T A\mathbf{v}$ an inner product. Show that A must be a positive definite matrix.
   Hint: To show A must be symmetric, use the symmetric property of the inner product with the standard basis vectors. To show A must be positive definite, use the positive definite property of the inner product.

Proposition 22.16. Let $S \in \mathcal{M}_{n\times n}$ be symmetric. Then, equivalently,

- S is positive definite
- S has all positive pivots
- S has all positive eigenvalues
- Every top-left submatrix of S has positive determinant
- $\mathbf{x}^T S\mathbf{x} > 0$ for any nonzero $\mathbf{x} \in \mathbf{R}^n$
- There exists $A \in \mathcal{M}_{m\times n}$ with independent columns and $S = A^T A$

Example 22.17. 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:

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \quad \to \quad \begin{bmatrix} 2 & -1 & 0 \\ 0 & 3/2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \quad \to \quad \begin{bmatrix} 2 & -1 & 0 \\ 0 & 3/2 & -1 \\ 0 & 0 & 4/3 \end{bmatrix}$$

The pivots are $2, 3/2, 4/3$, which are all positive. The eigenvalues are the roots of

$$\det(S - \lambda I) =$$

22.4 Hermitian and unitary matrices

Recall from a previous lecture about complex numbers.

Definition 22.18: 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 22.19. 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 22.20. 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.

Inquiry 22.21 (✠3.12): Consider the *Fourier matrix* $F = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{2\pi i/3} & e^{4\pi i/3} \\ 1 & e^{4\pi i/3} & e^{2\pi i/3} \end{bmatrix}$.

1. Note that F is symmetric. Is it Hermitian? Which entries must change so that it is Hermitian?

2. Show that F is unitary.

3. Compute the third power F^3 of F.

This matrix is very useful for real-world applications of linear algebra.

22.5 Exercises

Exercise 22.1. (✠3.10) The three vectors $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ are linearly independent.

1. Construct a matrix A with eigensystem $\{(\mathbf{u}, 2), (\mathbf{v}, -1), (\mathbf{w}, 3)\}$.

2. Give examples of two matrices B, C that are similar to A.

Exercise 22.2. (✠3.11) 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 22.3. (✠3.11) Let $A \in \mathcal{M}_{m \times n}$. Show that AA^T and A^TA are both symmetric matrices.

Exercise 22.4. (✠3.11) 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 22.5. (✠3.11) 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.

Exercise 22.6. (✠3.08, 3.12) Consider the symmetric matrix $A = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 0 & 1 \\ -2 & 1 & 2 \end{bmatrix}$.

1. Find the trace and determinant of A. Do not use a calculator, show your work.
2. Diagonalize A as $Q\Lambda Q^T$.
3. Express A as a sum of rank one matrices using the part above.

Lecture 23: Singular values

Chapters 7.1,7.2 in Strang's "Linear Algebra"

- Fact 1: No matter what size A has, AA^T and A^TA are both positive semidefinite and have the same nonzero eigenvalues.
- Fact 2: The SVD contains orthonormal bases of the four fundamental subspaces.

✠ Standard 3.13: Compute the rank r approximation to a matrix A

✠ Standard 3.14: Decompose a non-square matrix A by the SVD

This lecture continues with generalizing diagonalizibility. Instead to having some $X\Lambda X^{-1}$ decomposition for a square matrix, as in the previous lecture, we get a decomposition for a matrix of any rectangular size.

23.1 Eigenvalues of symmetric matrices

The word *singular* so far has been used when talking about matrices. A square matrix was seen to be singular if its determinant is zero, and non-singular otherwise. Before we begin with the new concept of *singular*, we make two observations.

Remark 23.1. Let any $A \in \mathcal{M}_{m\times n}$. Then $AA^T \in \mathcal{M}_{m\times m}$ and $A^TA \in \mathcal{M}_{n\times n}$

- both have the same nonzero eigenvalues, not counting algebraic multiplicity;
- both are positive semidefinite.

The first points follows by using the usual eigenvalue-eigenvector equations. Suppose that $(\lambda, \mathbf{u})$ is an eigenpair for AA^T, and the $(\mu, \mathbf{v})$ is an eigenpair for A^TA. Then

$$AA^T\mathbf{u} = \lambda\mathbf{u} \quad \Longrightarrow \quad A^TA(A^T\mathbf{u}) = \lambda(A^T\mathbf{u}), \tag{11}$$

$$A^TA\mathbf{v} = \mu\mathbf{v} \quad \Longrightarrow \quad AA^T(A\mathbf{v}) = \mu(A\mathbf{v}). \tag{12}$$

In other words, we the immediately get that $(\lambda, A^T\mathbf{u})$ is an eigenpair for A^TA and $(\mu, A\mathbf{v})$ is an eigenpair for AA^T. However, we only get this conclusion if $A\mathbf{u}$ and $A^T\mathbf{v}$ are not the zero vector! Recall that an eigenvector cannot be the zero vector. This situation is explored more in Inquiry 23.2.

The second point follows by obsevervation. Let $\mathbf{x} \in \mathbf{R}^m$ and $\mathbf{y} \in \mathbf{R}^n$. Then

$$\mathbf{x}^TAA^T\mathbf{x} = (\mathbf{x}^TA)\cdot(A^T\mathbf{x}) = (A^T\mathbf{x})^T\cdot(A^T\mathbf{x}) = (A^T\mathbf{x}) \bullet (A^T\mathbf{x}) = \|A^T\mathbf{x}\|^2 \geqslant 0,$$
$$\mathbf{y}^TA^TA\mathbf{y} = (\mathbf{y}^TA^T)\cdot(A\mathbf{y}) = (A\mathbf{y})^T\cdot(A\mathbf{y}) = (A\mathbf{y}) \bullet (A\mathbf{y}) = \|A\mathbf{y}\|^2 \geqslant 0,$$

where the last inequality follows from the nonnegativity of the norm $\|\cdot\|$. Note that $\mathbf{x}^TAA^T\mathbf{x}$ may be equal to zero even when $\mathbf{x} \neq 0$. Indeed, if $A^T\mathbf{x} = 0$, it simply means there is linear dependence among the columns of A^T (equivalently, among the rows of A).

Inquiry 23.2 (✠3.14): Consider the matrix $A = \begin{bmatrix} 1&1&1&2&3\\ 2&2&2&2&3\\ 3&3&3&3&3 \end{bmatrix} \in \mathcal{M}_{3\times 5}$. You may use a computer for this Inquiry.

1. Compute the 3 eigenvalues of AA^T and the 5 eigenvalues of A^TA.
2. Compute the eigenvectors for the zero eigenvalues of A^TA are zero.
3. Attempt to use Equation (11) to get the associated eigenvectors for AA^T. What is happening?

Definition 23.3: Let $A \in \mathcal{M}_{m\times n}$. The *singular values* of A are the square roots of the eigenvalues that AA^T and A^TA have in common.

Example 23.4. Continuing with the matrix A from Inquiry 23.2, we can apply the decomposition from Remark 22.8. For AA^T, suppose that it has eigensystem $\{(\lambda_1, \mathbf{u}_1), (\lambda_2, \mathbf{u}_2), (\lambda_3, \mathbf{u}_3)\}$, with $\lambda_1 > \lambda_2 > \lambda_3$. Then, using decimals,

$$AA^T \approx \underbrace{83.38}_{\lambda_1} \underbrace{\begin{bmatrix} 0.17 & 0.23 & 0.3 \\ 0.23 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.53 \end{bmatrix}}_{\mathbf{u}_1\mathbf{u}_1^T} + \underbrace{2.49}_{\lambda_2} \underbrace{\begin{bmatrix} 0.69 & 0.08 & -0.45 \\ 0.08 & 0.01 & -0.05 \\ -0.45 & -0.05 & 0.3 \end{bmatrix}}_{\mathbf{u}_2\mathbf{u}_2^T} + \underbrace{0.13}_{\lambda_3} \underbrace{\begin{bmatrix} 0.14 & -0.31 & 0.15 \\ -0.31 & 0.69 & -0.34 \\ 0.15 & -0.34 & 0.17 \end{bmatrix}}_{\mathbf{u}_3\mathbf{u}_3^T}.$$

Notice the very large eigenvalue and the two smaller ones. This decomposition will be useful when we ignore the smaller eigenvalues.

Inquiry 23.2 above showed that if AA^T has more eigenvalues than A^TA, or vice versa, then the extra eigenvalues are zero. However, this does not imply that AA^T and A^TA have the same number of independent eigenvectors!

Remark 23.5. Let $A \in \mathcal{M}_{m\times n}$. Let $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbf{R}^m$ be the eigenvectors of AA^T and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbf{R}^n$ be the eigenvectors of A^TA, where both are repeated depending on algebraic multiplicity. Without loss of generality, we assume that $n \geqslant m$, so $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$ are all eigenvectors for the zero eigenvalue. Let $\sigma_1, \ldots, \sigma_m \in \mathbf{R}$ 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 $i = 1, \ldots, m$. 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 Greek. 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). Combining all the equations $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$ into a big equation, and assuming that the $\mathbf{u}_i$ are orthonormal, and the $\mathbf{v}_i$ are orthonormal as well ,we get the following decomposition:

$$\begin{aligned}
A\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix} &= \begin{bmatrix} | & | & & | \\ \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \cdots & \sigma_m\mathbf{u}_m \\ | & | & & | \end{bmatrix} \\
A\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix} &= \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \end{bmatrix} \\
A\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_1 & - \\ - & \mathbf{v}_2 & - \\ & \vdots & \\ - & \mathbf{v}_m & - \end{bmatrix} &= \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_1 & - \\ - & \mathbf{v}_2 & - \\ & \vdots & \\ - & \mathbf{v}_m & - \end{bmatrix} \\
A &= \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_1 & - \\ - & \mathbf{v}_2 & - \\ & \vdots & \\ - & \mathbf{v}_m & - \end{bmatrix} \\
&= \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_m\mathbf{u}_m\mathbf{v}_m^T.
\end{aligned}$$

Definition 23.6: 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 singular values of A on the diagonal of its upper left $\text{rank}(A)\times\text{rank}(A)$ 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 singular values in Σ. The vectors $\mathbf{u}_i$ are called the *left singular vectors* and the $\mathbf{v}_i$ are called the *right singular vectors* of A.

Singular value decomposition allows us to have an eigenvalue-eigenvecftor type decomposition for non-square matrices. This is very powerful, as most data in real life is not square.

Inquiry 23.7 (✠3.13): Consider the two flags below (of Lithuania and Benin), given as matrices.

$$L = \begin{bmatrix} y & y & y & y & y & y & y & y & y \\ y & y & y & y & y & y & y & y & y \\ g & g & g & g & g & g & g & g & g \\ g & g & g & g & g & g & g & g & g \\ r & r & r & r & r & r & r & r & r \\ r & r & r & r & r & r & r & r & r \end{bmatrix} \qquad B = \begin{bmatrix} g & g & g & y & y & y & y \\ g & g & g & y & y & y & y \\ g & g & g & y & y & y & y \\ g & g & g & r & r & r & r \\ g & g & g & r & r & r & r \\ g & g & g & r & r & r & r \end{bmatrix}$$

1. How many singular values do these two matrices have?
2. Express both matrices as sums of rank one matrices.

You may use a computer for this task, and a Python function such as `svd` from the package `scipy.linalg`. You will need to convert colors to numbers (the choice of number does not matter, but distinct colors should have distinct numbers).

Very often we do not need the whole decomposition, only a part of it.

Definition 23.8: Let $A \in \mathcal{M}_{m\times n}$, and let $\sigma_1, \sigma_2, \dots$ be the singular values of A 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 $1 \leqslant r \leqslant \text{rank}(A)$. If $r = \text{rank}(A)$, then the rank r approximation of A is equal to A.

These rank r approximations help is massively reduce the amount of "information" in a matrix. For example, given a 100×100 matrix, which has $100^2 = 10\,000$ numbers, we could just consider the rank 5 approximation, which has $5 + 5\cdot(100+100) = 1005$ numbers, an approximately 90% reduction in size.

Inquiry 23.9 (✠3.13): This inquiry explores how "similar" the rank r approximations are to the input, continuing on the example given in class.

1. Open up the Google Colab notebook (link here) and execute the cells in your Python IDE.
2. For each of the two new images, find the r for which the rank r approximation "essentially looks like" the input image. What percentage reduction in information size did this achieve?

3. Find some images on your own, and perform the same steps as above. Without using single color images, try to find the images that have the highest reduction in size.

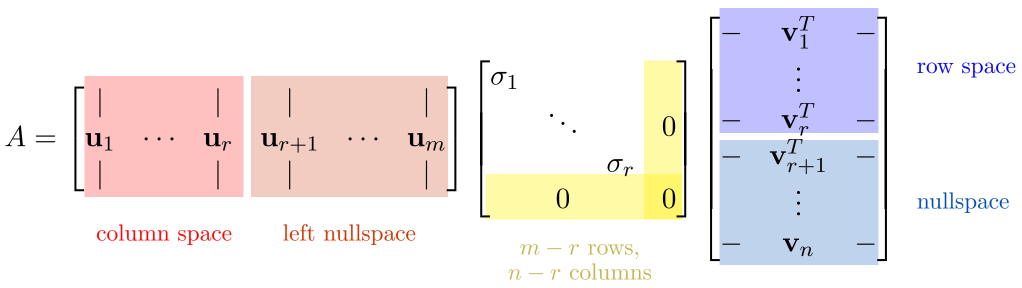
23.2 Bases in the decomposition

For this section, let $r = \text{rank}(A) \leqslant \min\{m, n\}$, for $A \in \mathcal{M}_{m\times n}$. We have alrerady seen the decomposition of A into three matrices, using eigenvalues and eigenvectors of AA^T and A^TA:

$$A = \underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \\ | & | & & | \end{bmatrix}}_{\text{eigenvectors of } AA^T} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}}_{\text{singular values}} \underbrace{\begin{bmatrix} - & \mathbf{v}_1^T & - \\ - & \mathbf{v}_2^T & - \\ & \vdots & \\ - & \mathbf{v}_r^T & - \end{bmatrix}}_{\text{eigenvectors of } A^TA}. \tag{13}$$

Hiding in this equation are the bases for the *four fundamental subspaces* that we have already seen in Lecture 11.

Remark 23.10. The rank(A)-approximation of A contains orthonormal basis vectors of other subspaces:



Example 23.11. 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 23.12. 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.

23.3 Exercises

Exercise 23.1. (✠3.13) Consider the two "matrices" below.

$$L = \begin{bmatrix} r & r & r & r & r & r & r & r & r & r \\ r & r & r & r & r & r & r & r & r & r \\ w & w & w & w & w & w & w & w & w & w \\ r & r & r & r & r & r & r & r & r & r \\ r & r & r & r & r & r & r & r & r & r \end{bmatrix} \qquad B = \begin{bmatrix} w & w & w & r & w & r & w & w & w \\ w & w & w & r & w & r & w & w & w \\ r & r & r & r & w & r & r & r & r \\ w & w & w & w & w & w & w & w & w \\ r & r & r & r & w & r & r & r & r \\ w & w & w & r & w & r & w & w & w \\ w & w & w & r & w & r & w & w & w \end{bmatrix}$$

1. Express L, the flag of Latvia, as a rank one product of two vectors.
2. Express B, the flag of Latvian battleships, as a sum of two rank one matrices. That is, decompose B as $B = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T$.

Exercise 23.2. (✠3.13) 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.

$\bowtie$ 1. Find the singular values of the image. How many of them are less than $1/100$ of the largest singular value?

$\bowtie$ 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 23.3. (✠3.14) 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 23.4. (✠3.13, 3.14)

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 24: Principal component analysis

Chapter 7.3 in Strang's "Linear Algebra"

- 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

✠ Standard 3.15: Normalize and center a matrix of n samples on its mean

✠ Standard 3.16: Identify the principal components of $A \in \mathcal{M}_{m\times n}$, in terms of the total covariance of A

✠ Standard 3.17: Solve the perpendicular least squares problem using SVD

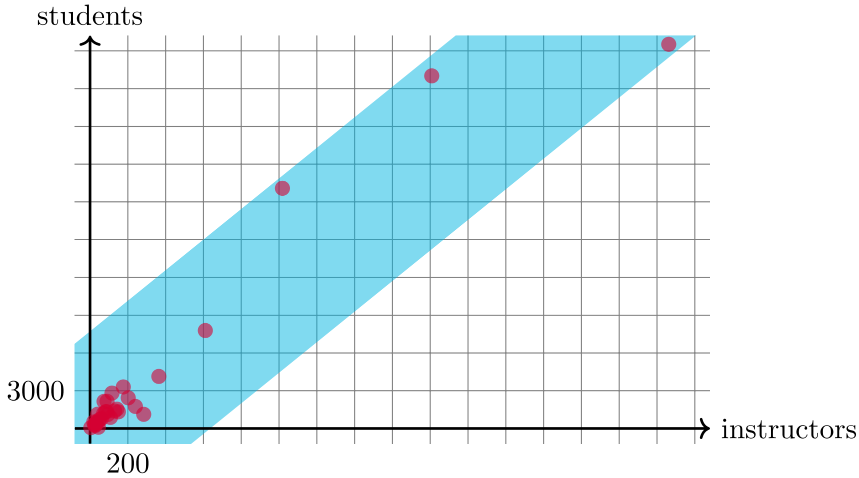
In the previous lecture, we saw how to simplify images, thought of as a matrix A, for compressed communication, using the eigenvectors of AA^T and A^TA, which appear in the singular value decomposition of A. in this lecture we will apply SVD, but to a different problem: *dimensionality reduction*.

24.1 The first significant direction of data

Data used in this lecture is available at `jlazovskis.com/teaching/linearalgebra/spring2022`.

Example 24.1. 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.

$$\{(1531, 15260), \ldots, (2, 33)\}$$



There seems to be a general trend! In Lecture 14 we saw how to approximate this data with a least squares line of best fit. We do something similar now, but slightly differently, and as motivation for higher dimensions. Each pair in this data set is a *sample*, so we can costruct a *sample matrix* $A \in \mathcal{M}_{2\times 32}$.

Definition 24.2: Let $A \in \mathcal{M}_{m\times n}$ and consider each of the n columns of A as a sample. There are two matrices associated to A:

$$M_{ij} = A_{ij} - \underbrace{\frac{1}{n}\sum_{k=1}^{n} A_{ik}}_{\text{mean of row } i}, \qquad S = \frac{MM^T}{n-1}.$$

A has a *mean-centered* matrix $M \in \mathcal{M}_{m\times n}$ and a *sample covariance* matrix $S \in \mathcal{M}_{m\times m}$.

By definition, S is symmetric.

Inquiry 24.3 (✠3.15): Consider the matrix $A = \begin{bmatrix} 2 & 3 & -1 & 6 & 1 \\ 0 & -3 & -4 & 1 & -1 \end{bmatrix}$.

1. Compute the mean-centered matrix M.
2. Suppose you add one column (sample) to M. Will M still be mean-centered? Why or why not?
3. Suppose you add two columns to M. What must be true about the two columns for the new M to still be mean-centered?

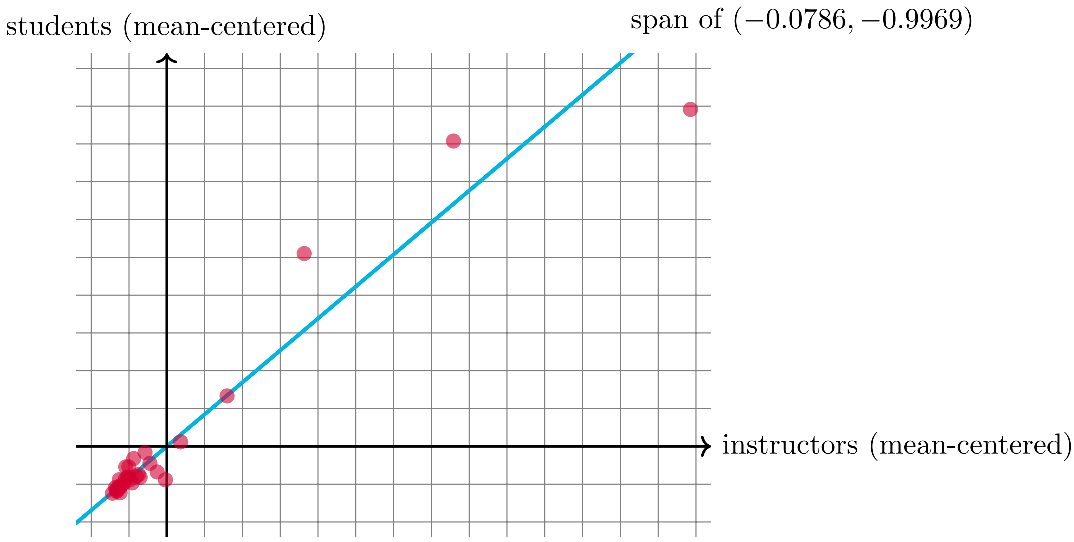
Continuing with Example 24.1, we find the means and center the matrix accordingly:

$$\begin{aligned} \text{mean of row 1 (students):} &\quad 145.6 \\ \text{mean of row 2 (instructors):} &\quad 1890.6 \end{aligned}$$

This lets us create the mean-centered 2×32 matrix M and the sample covariance 2×2 matrix S for the data. The key lies in the singular value decomposition of

$$S = \begin{bmatrix} 73909.14 & 864786.84 \\ 864786.84 & 10971745.39 \end{bmatrix} = \underbrace{\begin{bmatrix} -0.0786 & -0.9969 \\ -0.9969 & 0.0786 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 11039942.91 & 0 \\ 0 & 5711.62 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} -0.0786 & -0.9969 \\ -0.9969 & 0.0786 \end{bmatrix}}_{V^T}.$$

Since S is symmetric, the matrices U and V are the same. The singular vector with the largest eigenvalue identifies the *principal component* of the mean-centered data. This can be thought of 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. 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.



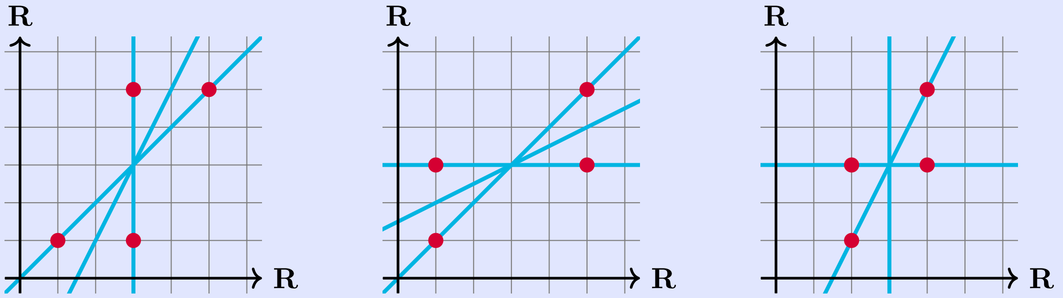
The line here is $y = \frac{0.9969}{0.0786}x$, which best approximates the mean-centered data. The line that best approximates the original data is this line, but shifted back by the mean:

$$y = \frac{0.9969}{0.0786}(x - 145.6) + 1890.6.$$

Definition 24.4: Let $A \in \mathcal{M}_{m\times n}$. The *(first) principal component* of A is the singular vector corresponding to the largest singular value of A.

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 13, which minimized the the vertical distance.

Inquiry 24.5 (✠3.16): Consider the data sets and lines below.



1. For each of the grids above, indicate which of the three lines you think corresponds to the linear least squares approximation and which corresponds to the first principal component.
2. Check your answers by computing the least squares linear approximation and the first principal component to the data sets. Use the interactive plot (link here).
3. Which approximation do you think is better? Why?
4. Try to come up with data for which the difference between the two lines is as big as posible.

The key idea for this inquiry is that least squares minimizes vertical distance and the first principal component minimizes perpendicular distance. "Distance" means the sum of the lengths from each point to the line.

24.2 PCA for higher dimensions

So far we saw data with two coordinates, but very often the data we see is many-dimensional, and has more than one important component. Now we analyze the principal components (that is, singular vectors) corresponding to the several largest singular values.

Example 24.6. The data from Example 24.1 can be augmented with extra data about the change in student and instructor numbers from the previous year. This gives 4-dimensional data, which can not be easily visualized on a page.

iestade	akad_pers_2019	akad_pers_2020	stud_2019	stud_2020
Latvijas Universitāte	1182	1531	15250	15260
Rīgas Tehniskā universitāte	930	904	14383	14006
Daugavpils Universitāte	194	182	2163	2068
⋮	⋮	⋮	⋮	⋮
Latvijas Nacionālā aizsardzības akadēmija	10	10	269	262

If we want to consider the change (percent), then we need to normalize the data, to make sure that a change in every coordinate is taken into account similarly.

Definition 24.7: Let $\mathbf{x} \in \mathbf{R}^n$. The *normalization* of $\mathbf{x}$ is a vector $\hat{\mathbf{x}} \in \mathbf{R}^n$ that is either:

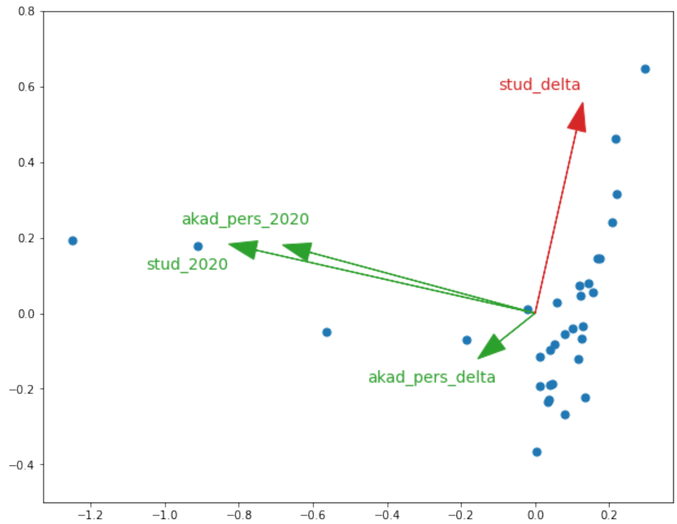
- a multiple of $\mathbf{x}$ so that it has unit length: $\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$
- a shift and scale the vector so that it lies in $[0,1]^n$: $\hat{\mathbf{x}} = \frac{\mathbf{x}-\mathbf{m}}{M-m}$, where $m = \min_i x_i$, $M = \max_i x_i$, and $\mathbf{m} = [m\ m\ \cdots\ m]^T$.

The second case is also called *min-max normalization*, and is the normalization used here.

We normalize each row, then center it at zero, then compute the sample covariance matrix, and finally get its SVD. The matrices U and Σ from the SVD are below.

$$U = \begin{bmatrix} -0.61 & 0.161 & -0.011 & -0.776 \\ -0.754 & 0.167 & -0.09 & 0.629 \\ -0.096 & -0.075 & 0.992 & 0.046 \\ -0.224 & -0.97 & -0.094 & -0.025 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 0.802 & 0 & 0 & 0 \\ 0 & 0.363 & 0 & 0 \\ 0 & 0 & 0.164 & 0 \\ 0 & 0 & 0 & 0.0145 \end{bmatrix}$$



Looking at the first two columns of U (the first two singular vectors), we see that the second coordinate (student number) has the largest magnitude for the first singular vector $\mathbf{u}_1$, and the last coordinate (change in student number) has the largest magnitude for the second singular vector $\mathbf{u}_2$:

$$\mathbf{x}_{\text{new}} = \text{proj}_{\text{span}(\mathbf{u}_1,\mathbf{u}_2)}(\mathbf{x}_{\text{old}}) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad x_1 = \frac{\mathbf{u}_1 \bullet \mathbf{x}_{\text{old}}}{\mathbf{u}_1 \bullet \mathbf{u}_1}, \quad x_2 = \frac{\mathbf{u}_2 \bullet \mathbf{x}_{\text{old}}}{\mathbf{u}_2 \bullet \mathbf{u}_2}.$$

The first two singular vectors are the "defining directions" of the data, and they most similar to the number of students and the change in students, respectively.

Definition 24.8: Let $A \in \mathcal{M}_{m\times n}$, considered as n samples in m coordinates.

- 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 $\text{trace}(S)$.

The variance of the data from Example 24.1 is either $\text{trace}(S) = S_{11} + S_{22}$ or $\text{trace}(\Sigma) = \Sigma_{11} + \Sigma_{22}$, since the sum of the eigenvalues of a matrix is the trace of the matrix. The singular value of the first principal component accounts for $\sigma_1/\text{trace}(S) \approx 0.99$, 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.

24.3 Exercises

Exercise 24.1. (✠3.17) This question is about the 4 point interactive found on the course website (link here).

1. Create an arrangement of the points with the largest angle possible between the two approximations that you can find. Do you think any angle is possible? Justify your reasoning.

2. Create an arrangement of the points with the largest difference between the sums of the distances that you can find. Besides all points being on a line, what situations give the same sums of distances?

Exercise 24.2. (✠3.16) 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 24.3. (✠3.16) 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 24.4. (✠3.16)

1. Create a matrix of 3-dimensional data for which first two principal components are the vectors $\left[\begin{smallmatrix} 1 & 0 & 0 \end{smallmatrix}\right]^T$ and $\left[\begin{smallmatrix} 0 & 1 & 0 \end{smallmatrix}\right]^T$. 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."

Part IV

Extensions

Lecture 25: Complex numbers

Chapters 9.1 and 9.2 in Strang

- Fact 1: All the math we have done so far can be considered over $\mathbf{C}$ instead of $\mathbf{R}$
- Fact 2: Complex number adition and multiplication have geometric meaning

✠ Standard 4.01: Express a complex number in one of four different ways

✠ Standard 4.02: Translate known properties of vectors and matrices to Hermitian vectors and matrices

In this lecture we will take some time to introduce fully the topic of coplex numbers. Fortunately, almost all the results we have seen so far with matrices over $\mathbf{R}$ apply to matrices over $\mathbf{C}$ as well.

25.1 The space of complex numbers

Definition 25.1: The *complex numbers* are elements of the set $\mathbf{C} = \{x + iy \ : \ x, y \in \mathbf{R}\}$. The symbol i is the *imaginary number*, having the property that $i^2 = -1$. For every $z = x + iy \in \mathbf{C}$:

- the *standard form* of z is $x + iy$.
- in *Cartesian*, or *rectangular* coordinates, the number z is written (x, y).

The *real part* of z is x and its *imaginary part* is y. If $x = 0$, then z is a *purely imaginary number*.

Let $z = x + iy$ and $w = a + ib$ be complex numbers and $c \in \mathbf{R}$. Complex number addition and multiplication, and real number multiplication are defined in the following way:

$$\begin{aligned} z + w &= (a + x) + i(y + b) \\ zw &= xa + ixb + iya + i^2tb = (xa - yb) + i(xb + ya) \\ cz &= cx + icy \end{aligned}$$

Inquiry 25.2 (✠4.02): The set $\mathbf{C}$ along with complex number addition and scalar multiplication as above form a vector space.

1. Show that the function $f\colon \mathbf{C} \to \mathbf{R}^2$, given by $f(x + iy) = (x, y)$ is a bijection.
2. With the bijection from above, the complex number $z = 1 + i$ could be considered as the vector $\mathbf{v} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right] \in \mathbf{R}^2$. Compute the square $z \cdot z$ and the dot product $\mathbf{v} \bullet \mathbf{v}$. Why do you get two dfferent results?
3. For any $z \in \mathbf{C}$, will $z \cdot z$ always be a real number? Give an example when it is and another example when it isn't.
4. Describe a surjective function $\mathbf{C} \to \mathbf{R}$ that takes in a complex number, and outputs a real number.

Example 25.3. 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 25.4: Let $z = x + yi \in \mathbf{C}$. The (*complex*) *conjugate* of z is $\overline{z} = z^* = x - iy$. The *absolute value*, or *modulus* of z is

$$|z| = \sqrt{z\overline{z}} = \sqrt{(x+iy)(x-iy)} = \sqrt{x^2+y^2}.$$

Taking the conjugate twice returns back the original number: $(z^*)^* = z$.

Proposition 25.5. Let $z = x + iy, w = a + ib \in \mathbf{C}$. 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:

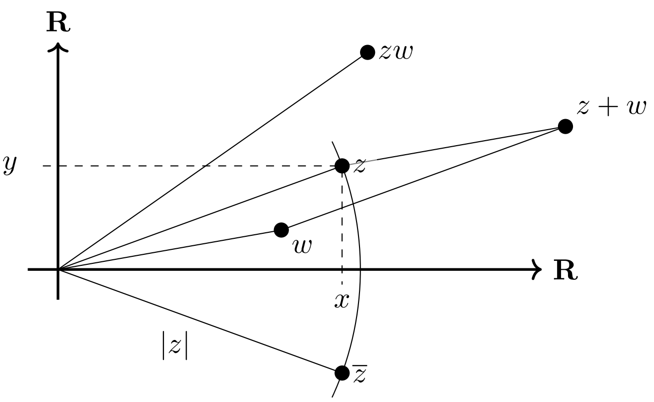
1. $|z| = 0$ iff $z = 0$
2. $|\overline{z}| = |z|$
3. $|zw| = |z||w|$
4. $|z+w| \leqslant |z| + |w|$

Definition 25.6: The third way to express $z = x+iy \in \mathbf{C}$ 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

$$x + iy = r\cos(\theta) + ir\sin(\theta) = re^{i\theta},$$

where the second equality is known as *Euler's formula*. This last expression is in *exponential form*.

Remark 25.7. 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.



$$\overline{z} = x - iy = r_z\cos(\theta_z) - ir_z\sin(\theta_z)$$
$$z = x + iy = r_z\cos(\theta_z) + ir_z\sin(\theta_z)$$
$$z + w = r_w\cos(\theta_w) + ir_w\sin(\theta_w)$$
$$zw = r_z r_w e^{i(\theta_z+\theta_w)}$$

Remark 25.8. Putting complex numbers into polar coordinates makes computations in standard form much easier. For $z = re^{i\theta}$ and $n \in \mathbf{N}$, we have:

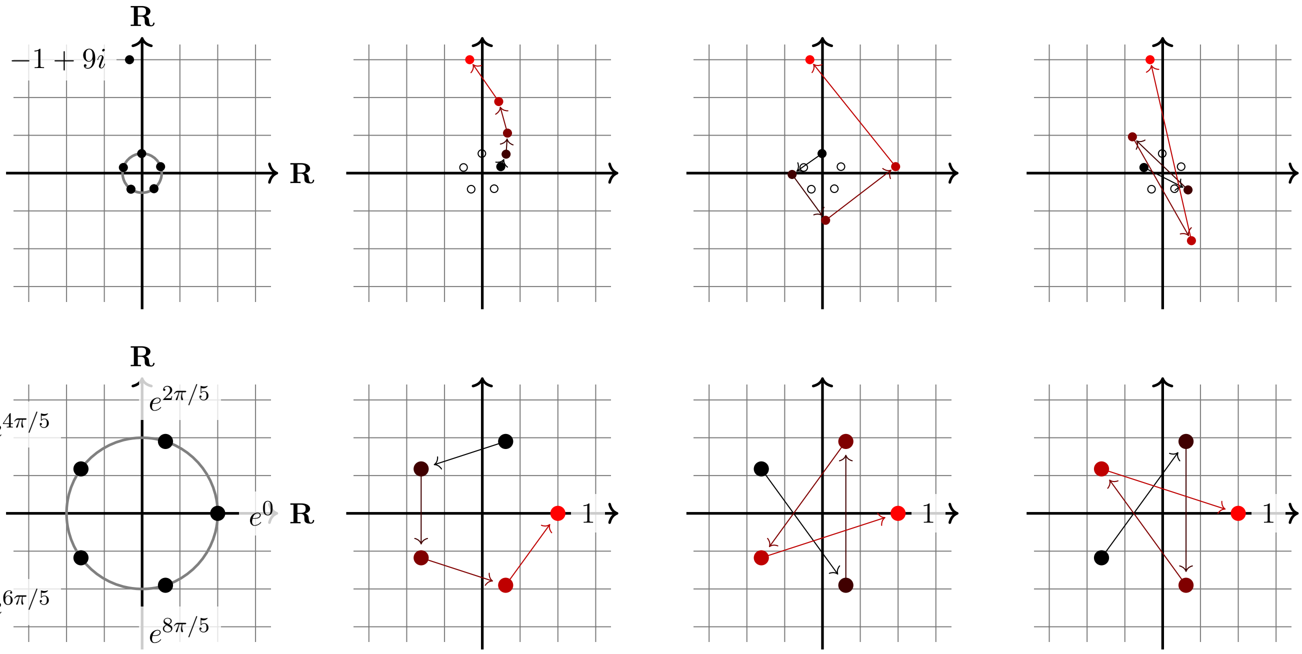
- (De Moivre's theorem) $z^n = (re^{i\theta})^n = r^n e^{in\theta}$
- (complex roots) the nth roots of z are $r^{1/n}e^{i(\theta+2k\pi)/n}$, for every $k = 0, 1, \ldots, n-1$.

For the second point, when $z = 1 + 0i$, then the kth root of z is called the *kth root of unity*.

Inquiry 25.9 (✠4.01): This inquiry is about the different forms of complex numbers.

1. Express $z = 5\cos(\pi/4) + 5i\sin(\pi/4)$ in standard form.
2. Express $w = -\sqrt{3} - i$ in polar form.
3. Find the 4th roots of $p = 1 + i$ in Cartesian coordinates.
4. Explain why finding nth roots of unity is much easier in polar coordinates than in rectangular coordinates.

Example 25.10. 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 25.11. The space of complex numbers is a 2-dimensional vector space over $\mathbf{R}$ via the identification of Cartesian coordinates. However, it is a 1-dimensional vector space over $\mathbf{C}$.

25.2 Complex vectors and complex matrices

Just like we generalized numbers to vectors, we generalize complex numbers to complex matrices. We now talk about the vector space $\mathbf{C}^n$, of vectors having n components, and the matrix space $\mathcal{M}_{m\times n}(\mathbf{C})$, of $m \times n$ matrices with complex number entries.

Remark 25.12. Multiplication of complex numbers may be viewed as matrix multiplication. Making a correspondence between $z = x + iy \in \mathbf{C}$ and $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \in \mathbf{R}^2$, as in Inquiry 25.2, reveals a correspondence for multiplication:

$$(a + ib)(x + iy) \quad \leftrightarrow \quad \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Definition 25.13: 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 $\mathbf{z} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = x + iy = z$, 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, with $A^* \in \mathcal{M}_{n\times m}(\mathbf{C})$ whenever $A \in \mathcal{M}_{m\times n}(\mathbf{C})$ defined by $(A^*)_{ij} = \overline{A_{ji}}$.

Definition 25.14: The square matrix $A \in \mathcal{M}_{n\times n}(\mathbf{C})$ is is *Hermitian* if $A = A^*$.

We will see nice properties of Hermitian matrices later in the course. For now we consider some of their properties.

Proposition 25.15. Let $A, B \in \mathcal{M}_{n\times n}(\mathbf{C})$ be Hermitian. Then:

- the entries on the diagonal of A are real numbers
- the identity $(AB)^* = B^* A^*$ holds

Inquiry 25.16 (✠4.02): Let $A \in \mathcal{M}_{n\times n}(\mathbf{C})$ be Hermitian, and let $\mathbf{v} \in \mathbf{C}^n$.

1. Expand the product $(\mathbf{v}^* A \mathbf{v})^*$ to show that it is Hermitian. How many rows and columns does the product have, and in what space must it be?
2. Find the complete solution to $\left[\begin{smallmatrix} 0 & 3+i \\ 3-i & 0 \end{smallmatrix}\right] \mathbf{z} = \left[\begin{smallmatrix} i \\ 2-i \end{smallmatrix}\right]$, for $\mathbf{z} \in \mathbf{C}^2$.

25.3 Exercises

Exercise 25.1. (✠4.02) Show that every complex number $z = x + iy$ 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 25.2. (✠4.02) Prove all the claims of Proposition 9.5, 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 25.3. (✠4.01) This question is about proving Euler's formula $\cos(\theta) + i\sin(\theta) = e^{i\theta}$.

1. Take the derivative of $f(\theta) = (\cos(\theta) + i\sin(\theta))e^{-i\theta}$ with respect to θ.
2. Explain why the result of the previous step means that $f(\theta)$ is constant.
3. Evaluate f at $\theta = 0$ to find this constant from the previous step.
4. Rearrange to get Euler's formula.

Lecture 26: Graphs

Chapter 10.1 in Strang's "Linear Algebra" and IV.6 in Strang's "Learning from Data"

- Fact 1: Graphs may be directed or undirected, and may or may not have weight associated to vertices or edges.
- Fact 2: Row reducing the incidence matrix gives a spanning tree.
- Fact 3: Something about connectedness

✠ Standard 4.03: Construct the four matrices associated to graph (adjacency, incidence, Laplacian, transition probability), and reconstruct the graph from them

✠ Standard 4.04: Find a spanning tree of a graph using row reduction on the incidence matrix

---

In this lecture we take a brief break to set up a new interpretation of matrices. We will apply the tools of matrix algebra already seen so far to graphs. A *network* is just a fancy name for a *graph*, with perhaps more structure. In general, both "graph" and "network" refer to the same thing.

26.1 The structure of graphs

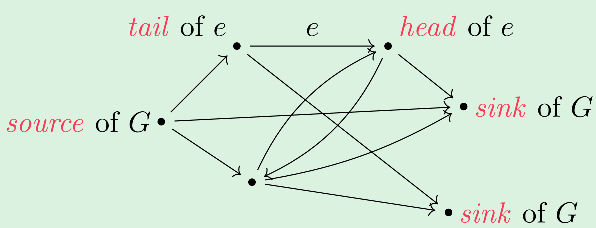
Definition 26.1: A *graph* G is a pair of sets (V, E), where $V = \{v_1, \ldots, v_n\}$ is a finite set and every element of E is a set $\{v_i, v_j\}$, for $1 \leqslant i < j \leqslant n$. The elements of V are called *vertices* (singular *vertex*) and the elements of E are called *edges*.

The above definition only conains the most basic information, but with a little work we can get much more.

Remark 26.2. Additional structure may be placed upon graphs in the following ways:

- Definition 26.1 is for an *undirected* graph, as demonstrated by the fact that an edge is a set, so $\{v_i, v_j\} = \{v_j, v_i\}$. 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 $i \neq j$.
- The set V of vertices in Definition 26.1 is an unordered set, but the naming often gives vertices an order. When the elements of V have an order, the edges have a natural order as well, with $\{v_i, v_j\} \leqslant \{v_k, v_\ell\}$ whenever $i < k$, or $i = k$ and $j < \ell$. Here we assumed without loss of generality that $i = \min\{i, j\}$ and $k = \min\{k, \ell\}$.

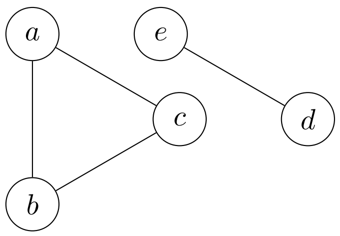
Definition 26.3: Let $G = (V, E)$ be a directed graph. In the edge $e = (t, h) \in E$, the vertex t is called the *tail* and the vertex h is called the *head* of e. If $v \in V$ only appears as a tail in edges, then v is called a *source* of G. If v only appears as a head, then v is called a *sink* of G.

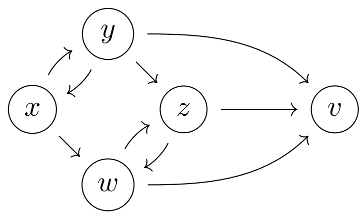


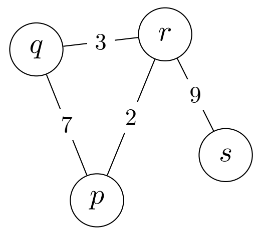
The egde $e = (t, h)$ is an *outgoing* edge of t and an *incoming* edge of h.

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 (such as an edge $\{v_i, v_i\}$). Graphs without repeated edges and without self loops are called *simple* graphs.

Example 26.4. Here are some examples of graphs and their associated adjacency matrices.







$V = \{a, b, c, d, e\}$

$E = \{\{a, b\}, \{b, c\}, \{d, e\}, \{c, a\}\}$

$V = \{x, y, z, v\}$

$E = \{(x, y), (y, x), (x, w), (y, z), (y, v), (w, z), (z, w), (w, v), (z, v)\}$

$V = \{p, q, r, s\}$

$E = \{(p, q), (p, r), (q, r), (r, s)\}$

$$\begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ x \\ y \\ z \\ w \\ v \end{array} \begin{array}{c} \begin{array}{ccccc} x & y & z & w & v \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

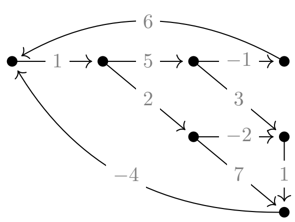
$$\begin{array}{c} \\ p \\ q \\ r \\ s \end{array} \begin{array}{c} \begin{array}{cccc} p & q & r & s \end{array} \\ \begin{bmatrix} 0 & 7 & 2 & 0 \\ 7 & 0 & 3 & 0 \\ 2 & 3 & 0 & 9 \\ 0 & 0 & 9 & 0 \end{bmatrix} \end{array}$$

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. The graph on the right is a *weighted graph*, meaning every edge has a (potentially negative) number associated to it.

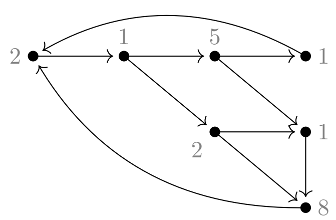
Definition 26.5: A graph $G = (V, E)$ is *weighted* when accompanied by a function $w\colon E \to \mathbf{R}$. This is sometimes called an *edge-weighted* graph to distinguish it from a *vertex-weighted* graph, which needs a function $w\colon V \to \mathbf{R}$.

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.

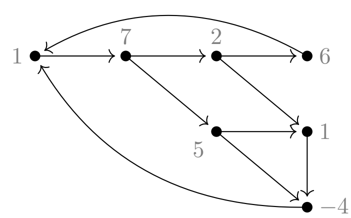
Example 26.6. Here is an example of an edge-weighted directed graph G and two vertex-weighted graphs G_1, G_2 that are built following the comment above.



G



G_1



G_2

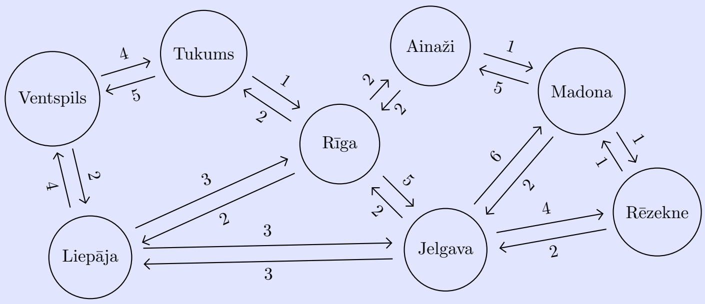
> **Definition 26.7:** Let $G = (V, E)$ be a graph and $A \in \mathcal{M}_{n\times n}$ its adjacency matrix. For $v_k \in V$, the *degree* of v_k is the number of edges in E in which v_k appears. Or, it is the sum
>
> $$\deg(v_k) = \underbrace{\sum_{i=1}^{n} A_{ik}}_{G \text{ undirected}} \qquad \text{or} \qquad \deg(v_k) = \underbrace{\overbrace{\sum_{i=1}^{n} A_{ik}}^{\textit{out-degree}} + \overbrace{\sum_{j=1}^{n} A_{kj}}^{\textit{in-degree}}}_{G \text{ directed}}.$$
>
> The out-degree of v_k is denoted $\operatorname{outdeg}(v_k)$, and the in-degree is denoted $\operatorname{indeg}(v_k)$. If every vertex $v \in V$ has $\deg(v) = k$, then G is called a *k-regular* graph.

Remark 26.8. Let $G = (V, E)$ be a graph and $A \in \mathcal{M}_{|V|\times|V|}$ be its adjacency matrix. There are three other matrices associated to G:

- 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 *Laplacian matrix* $L \in \mathcal{M}_{|V|\times|V|}$, defined as $L = N^T N$. If G is undirected, then $L = D - A$, where D is a diagonal matrix with $D_{ii} = \deg(v_i)$ and A is the adjacency matrix
- the *transition probability matrix* $T \in \mathcal{M}_{|V|\times|V|}$, where the (i, j)-entry of T is defined as the "probability" of going from vertex i to vertex j. For different types of graphs, there are different definitions:

$$\underbrace{T_{ij} = \begin{cases} 0 & \text{if } A_{ij} = 0 \\ \frac{1}{\deg(v_i)} & \text{else} \end{cases}}_{\text{undirected, unweighted graph}} \qquad \underbrace{T_{ij} = \begin{cases} 0 & \text{if } A_{ij} = 0 \\ \frac{1}{\operatorname{outdeg}(v_i)} & \text{else} \end{cases}}_{\text{directed, weighted graph}} \qquad \underbrace{T_{ij} = \begin{cases} 0 & \text{if } A_{ij} = 0 \\ \frac{w(v_i, v_j)}{\sum_k w(v_i, v_k)} & \text{else} \end{cases}}_{\text{weighted graph}}$$

Most often the adjacency matrix is used, since it is square and the graph can be easily reconstructed from it.

> **Inquiry 26.9 (✠4.03):** Consider the following (imagined) traffic observations between among cities.
>
> 
>
>
> 1. Construct three matrices associated to this graph: adjacency, incidence, Laplacian.
> 2. Construct the transition probability matrix using the weights as shown.
>
> The transition probability matrix is a *stochastic* matrix, which can be multiplied with itself to see how a system evolves.

Definition 26.10: 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 $e \in E'$.

26.2 Patterns in graphs

Definition 26.11: Let $G = (V, E)$ be a graph.

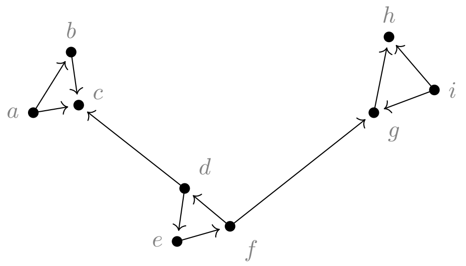
- A *path* in G is an ordered sequence of distinct vertices $v_1, \dots, v_n$ for which v_i and v_{i+1} form an edge, for every i.
- A *walk* in G is the same as a path, but the vertices do not need to be distinct.
- A *cycle*, or *loop* in G is a path for which v_n and v_1 form an edge.

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 26.11, 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 26.12. Consider the following directed graph and associated sequences of vertices.

$G =$



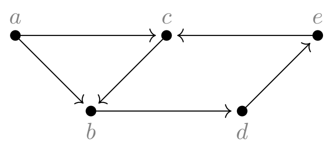
$$\begin{aligned} P_1 &= (d, e, f, g, h) \\ P_2 &= (c, d, e, f, g, h) \\ W_1 &= (f, d, e, f, g, h) \\ W_2 &= (a, b, c, a, c, a, b, c) \\ C_1 &= (a, b, c) \\ C_2 &= (d, e, f) \end{aligned}$$

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.

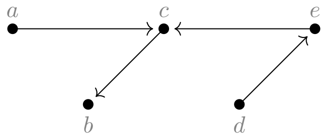
Remark 26.13. Let $G = (V, E)$ be the graph given below, with incidence matrix N.

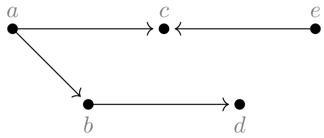
$G =$



$$N = \begin{array}{c} \\ ab \\ ac \\ bd \\ cb \\ de \\ ec \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \end{array}$$

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.

$T_1 =$ 

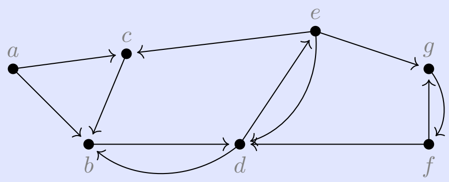
$T_2 =$ 

$$ab = ac + cb \qquad\qquad cb = ab - ac$$
$$bd = -cb - de - ec \qquad\qquad de = ac - ab - bd - ec$$

Proposition 26.14. 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.

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.

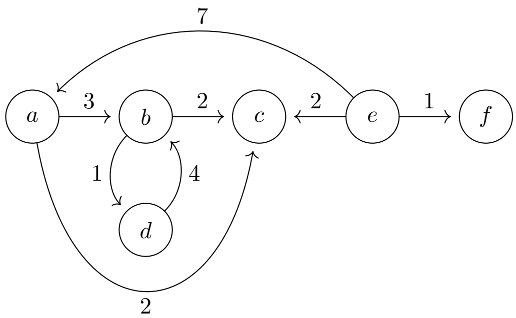
Inquiry 26.15 (✠4.04): Consider the following directed graph G.



1. Construct the incidence matrix for G.
2. Row reduce the matrix to find a spanning tree T of G.
3. Is this spanning tree unique for G? Would row reducing by hand (instead of a computer) yield a different tree?

26.3 Exercises

Exercise 26.1. (✠4.03, 4.04) Consider the following directed graph:

$G =$ 

1. Compute the adjacency A, incidence N, Laplacian L, and transition probability T matrices for G. Use the weighted definition for T.
2. Row reduce N and give the resulting spanning tree of G.
3. Using a computer make an educated guess as to what $\lim_{n\to\infty} T^n$ could be.
4. Let $\hat{T}$ be the same as T, but with the (c, c) and (f, f) entries 1 (instead of 0) on the diagonal. Using a computer make an educated guess as to what $\lim_{n\to\infty} \hat{T}^n$ could be.

The last two questions address the *network flow*, where matrix multiplication represents movement along the edges, and the weights represent the probability of moving along a given edge (relative to all the weights outgoing from the tail node).

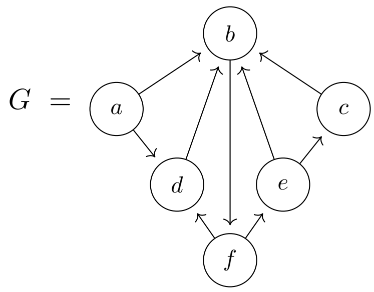
Exercise 26.2. (✠4.03) Let $G = (V, E)$ be a directed simple graph, and let A be its adjacency matrix.

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 26.3. (✠4.03) Use induction to prove Proposition 26.14.

Exercise 26.4. (✠4.03, 4.03) 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.

The preferred (but not required) order of the edges is the lexicographic order: $ab, ad, bf, cb, db, eb,$ ec, fd, fe.

Index of notation

Index