Math 318
Advanced Linear Algebra – Tools and Applications

Lecture Notes by Sam Roven & Rekha Thomas

July 2, 2021

Contents

Preliminaries

Introduction

- **Math 318 vs Math 308**: Math 318 is a second course in linear algebra that follows Math 308. The course begins with a quick review of eigenvalues and diagonalization which is the last topic in Math 308, and the only topic from Math 308 that we will revisit explicitly. You will be expected to remember and use the concepts you learned in Math 308, so it will be helpful to have your Math 308 notes and textbook handy. You can also find a full set of class notes for Math 308 at `samroven.com/linear`. Unless explicitly stated, you are allowed to use any fact/theorem from Math 308 in solving problems in Math 318. To test your readiness for Math 318, please take the self diagnostic test at `https://sites.math.washington.edu/∼m318diagnostictest/` before the course.

- **What is Math 318 about?** The aim of this course is to further develop the properties of matrices, linear maps and other concepts from linear algebra to a point where you start seeing some of the powerful applications of the subject. Linear algebra is one of the most useful math languages in modern day applications such as those in machine learning, big data, statistics, optimization, natural sciences, social sciences and engineering. Many of you may have already encountered applications of linear algebra. One of our aims will be to understand the mathematics behind these applications and their algorithms.

  In Math 308 you encountered matrices. Here is a rough classification of the different lives of a matrix $A \in \mathbb{R}^{m \times n}$:

  1. A matrix is a way to organize computation. For example if you want to solve a linear system of equations, you express it as $A\mathbf{x} = \mathbf{b}$ and then solve the system via Gaussian elimination on the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$.
  2. A matrix can be thought of as a way to organize data. For example, consider the case where the n columns of a matrix are indexed by different foods you commonly eat like granola bar, chocolate milk, tofu scramble etc, the m rows by food groups such as protein, carbohydrate, fat etc, and the entries in a column record the amount of protein, carbs, fat etc in one unit of the food indexing that column. When you compute $A\mathbf{x}$ where $\mathbf{x}$ is a vector whose entries are the amount of each food you eat in a day, you see how much of each food group you are consuming that day.
  3. A matrix $A \in \mathbb{R}^{m \times n}$ also represents a linear map from $\mathbb{R}^n \to \mathbb{R}^m$ that sends $\mathbf{x} \in \mathbb{R}^n$ to $A\mathbf{x} \in \mathbb{R}^m$. Many physical operations like projections, rotations, reflections etc can be encoded as linear transformations, each represented by a matrix. This is a powerful use of matrices which can be used to model many real world situations.

  In Math 308 you encountered all these uses of matrices and there was a lot of emphasis on (1). In this course, we will focus mostly on (3) and also (2).

Math Notation

- $\mathbb{R}^{m\times n} :=$ the vector space of all $m \times n$ matrices with real entries.
- $\mathbb{C}^{m\times n} :=$ the vector space of all $m \times n$ matrices with complex entries.
- $\neq$ means "not equal to", e.g., $2 \neq 3$.
- $\in$ denotes "element of ", e.g. $1 \in \{1, 2, 3\}$ means 1 is an element of the set $\{1, 2, 3\}$.
  $\notin$ denotes "not element of ", e.g. $5 \notin \{1, 2, 3\}$ means 5 is not an element of the set $\{1, 2, 3\}$.
- $\subseteq$ denotes "subset of" or "is contained in", e.g. $\{1, 2\} \subseteq \{1, 2, 3\}$ means the set $\{1, 2\}$ is a subset of the set $\{1, 2, 3\}$.
- $\forall$ denotes "for all", e.g. $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \in \text{span}\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ translates to "For all vectors $\mathbf{x}$ in $\mathbb{R}^n$, $\mathbf{x}$ is in the span of $\mathbf{e}_1$ through $\mathbf{e}_n$".
- $\exists$ denotes "there exists", e.g. If $\mathbf{a}$ is a multiple of $\mathbf{b}$ then $\exists$ some scalar r such that $\mathbf{a} = r\mathbf{b}$.
- $:$ or $|$ denotes the word "such that", e.g. $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ denotes the set of all vectors of the form $a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n$ such that a_i are any real numbers. This is written as

  $$\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\} = \{a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n \,:\, a_i \in \mathbb{R}\}$$

  or

  $$\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\} = \{a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n \,|\, a_i \in \mathbb{R}\}$$

- $\Rightarrow$ means "implies", e.g., The statement "if p is a prime number bigger than 2 then p is odd" can be written as "p prime, $p > 2 \Rightarrow p$ is odd".
- $\Leftrightarrow$ means "if and only if", e.g., p even $\Leftrightarrow p$ is a multiple of 2.
- Compact notation for a matrix $A \in \mathbb{R}^{m\times n}$ (a_{ij} is the entry in row i and column j):

  $$(a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{n2} & \cdots & a_{mn} \end{bmatrix}$$

Acknowledgments

These are informal lecture notes for Math 318, provided for your convenience. In particular, they are brief and follow the class lectures closely. It is always advisable to consult linear algebra books if you need more or multiple explanations of concepts. These notes closely follow Chapters 4,6,7,9, and 10 in the book *Introduction to Linear Algebra, Fifth Edition* by Gilbert Strang which is recommended for this course. Many of the homework problems are inspired by (but not usually exactly) Strang's exercises. When this is the case, the problem number in Strang's book is written next to the exercise. These notes also draw from the book *Thirty-three Miniatures* by Jiri Matousek.

Caution!

These notes are a work in progress and may have errors. If you find mistakes or typos please let your instructor know.

Chapter 1

Eigenvalues and Diagonalization

1.1 Eigenvalues and Eigenvectors

Consider the *vector space* $\mathbb{R}^{m\times n}$ of all $m \times n$ matrices with entries in $\mathbb{R}$. Being a vector space means that the elements of $\mathbb{R}^{m\times n}$ (which are matrices) satisfy the following under addition and scalar multiplication:

1. for all $A, B \in \mathbb{R}^{m\times n}$, their sum $A + B \in \mathbb{R}^{m\times n}$,

2. for any $A \in \mathbb{R}^{m\times n}$ and scalar $r \in \mathbb{R}$, the product $rA \in \mathbb{R}^{m\times n}$, and

3. the zero matrix $\mathbf{0} \in \mathbb{R}^{m\times n}$.

The set of all square matrices of size $n \times n$ is denoted as $\mathbb{R}^{n\times n}$. The set of all diagonal matrices of size $n \times n$ is a *subspace* of $\mathbb{R}^{n\times n}$. Every subspace is a vector space – it just happens to live in a bigger vector space.

A matrix $A \in \mathbb{R}^{m\times n}$ represents the linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, that sends $\mathbf{x} \in \mathbb{R}^n$ to $A\mathbf{x} \in \mathbb{R}^m$, written as $\mathbf{x} \mapsto A\mathbf{x}$. It is usual to denote the linear transformation also by A and write $A : \mathbb{R}^n \to \mathbb{R}^m$ such that $\mathbf{x} \mapsto A\mathbf{x}$. Note that A is used both to denote the linear transformation and its matrix representation.

Throughout this chapter we consider a square matrix $A \in \mathbb{R}^{n\times n}$. Then both $\mathbf{x}$ and its image $A\mathbf{x}$ live in $\mathbb{R}^n$. Normally the vector $\mathbf{x}$ and its image $A\mathbf{x}$ are not related in any simple way, but for some vectors $\mathbf{x}$, it may happen that $A\mathbf{x}$ is simply a scaling of $\mathbf{x}$. Such vectors tell us about parts (subspaces) of the domain $\mathbb{R}^n$ that only scale (and hence remain fixed) under the action of A. This is important information about the linear transformation A and brings us to the notion of eigenvalues and eigenvectors.

Definition 1.1.1. Let $A \in \mathbb{R}^{n\times n}$. If there exists a non-zero vector $\mathbf{x} \in \mathbb{R}^n$ and some scalar $\lambda \in \mathbb{R}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ we say that **$\mathbf{x}$ is an eigenvector of A with eigenvalue λ**.

Question: Should $\mathbf{0} \in \mathbb{R}^n$ be allowed as an eigenvector of A? If yes, what would be its eigenvalue? Do you see why the above definition requires eigenvectors to be non-zero?

Definition 1.1.2. The **eigenspace of eigenvalue λ**, denoted as E_λ, is the set of all vectors $\mathbf{x} \in \mathbb{R}^n$ that are eigenvectors of A with eigenvalue λ. Mathematically, $E_\lambda = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \lambda\mathbf{x}\}$.

It follows from the definition that $E_\lambda = \{\mathbf{x} \in \mathbb{R}^n : (A - \lambda I)\mathbf{x} = \mathbf{0}\} = \mathrm{Null}(A - \lambda I)$ where the notation $\mathrm{Null}(B)$ denotes the nullspace of the matrix B. Since E_λ is the nullspace of a matrix, it is a subspace of $\mathbb{R}^n$. Any vector in E_λ only scales by λ under the linear transformation A.

The nullspace of a square matrix is non-zero if and only if the matrix is singular (not invertible) which is if and only if the determinant of the matrix is zero. Therefore, $E_\lambda = \mathrm{Null}(A - \lambda I) \neq \{\mathbf{0}\}$ if and only if $\det(A - \lambda I) = 0$.

Proposition 1.1.3. *A scalar $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if* $\det(A - \lambda I) = 0$

This proposition enables the computation of all eigenvalues of A: Let the variable t denote an unknown eigenvalue and set up the equation $\det(A - tI) = 0$. On the left is a polynomial in t of degree n. Its roots are the eigenvalues of A. The polynomial $\det(A - tI)$ is called the *characteristic polynomial of A*.

Question: Do you see why $p(t) = \det(A - tI)$ is a polynomial in t of degree n? **Hint**: *Think of the cofactor formula for computing a determinant. We'll see more of this soon.*

Example 1.1.4. Let $A \in \mathbb{R}^{2\times 2}$ be the matrix that represents reflection in $\mathbb{R}^2$ about the line $y = x$. This is a linear transformation (check if you wish). Recall how one finds the matrix of a linear transformation.

Matrix of a linear transformation

Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, the matrix for T can be written as

$$\begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$$

where $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis of $\mathbb{R}^n$. **Remember: A linear transformation is completely determined by where it takes a basis of the domain.**

Using this fact, we find that the matrix representing the above reflection is $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Before we compute the eigenvalues and eigenvectors of this matrix, let us see if we can just find them using geometry. Which vectors are only scaled by the above reflection? There are two natural guesses based on geometry:

1. If $\mathbf{x}$ lies on the line $y = x$, then it remains unchanged under reflection about the line $y = x$ and hence all vectors of the form $r \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ where $r \in \mathbb{R}$ is an eigenvector of A with eigenvalue 1.
2. If $\mathbf{x}$ is perpendicular to the line $y = x$ then it goes to $-\mathbf{x}$ under this reflection. Therefore, all vectors of the form $r \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ where $r \in \mathbb{R}$ is an eigenvector of A with eigenvalue -1.

We can now double check our predictions by solving the characteristic polynomial equation:

- Step 1: Compute that $\det(A - \lambda I) = \lambda^2 - 1$.
- Step 2: Solving $\det(A - \lambda I) = \lambda^2 - 1 = 0$ we find that the eigenvalues of A are $\lambda = \pm 1$.
- Step 3: For each λ, we find a basis of $E_\lambda = \text{Null}(A - \lambda I)$

$$\lambda = 1 \implies A - I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \text{ which has echelon form } \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$$

This gives the linear system $-x_1 + x_2 = 0 \implies x_1 = x_2$ and $E_1 = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$.

$$\lambda = -1 \implies A + I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{ which has echelon form } \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

This gives the linear system $x_1 + x_2 = 0 \implies x_1 = x_2$ and $E_{-1} = \text{Span}\left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$.

Indeed, we had found all the eigenvalues and eigenvectors of A using geometry.

1.1.1 The characteristic polynomial

We now look at the characteristic polynomial more closely. Recall that $p(\lambda) = \det(A - \lambda I)$ is a degree n polynomial in the variable λ and its roots are the eigenvalues of A. (It is usual to use λ for the variable t when we talk about characteristic polynomials, but you should remember that this λ is a variable and not any particular eigenvalue.) By *roots* of a polynomial $p(t)$ we mean the values of t for which $p(t) = 0$. What can we say about the roots of the characteristic polynomial $p(\lambda)$ or more generally, a univariate polynomial $p(t)$ of degree n? To fully understand this we need to recall *complex numbers.*

The *imaginary* i is defined to be the positive square root of -1, i.e., $i^2 = -1$. An expression of the form $a + ib$ where $a, b \in \mathbb{R}$ is called a complex number, e.g., $2 + 3i, \sqrt{5} - 2i$. The set of all complex numbers is denoted $\mathbb{C}$. Real numbers are special cases of complex numbers, i.e., $\mathbb{R} = \{a + 0i \,:\, a \in \mathbb{R}\}$. Therefore, $\mathbb{R} \subseteq \mathbb{C}$. The conjugate of a complex number $a + ib$ is $a - ib$, and the conjugate of $a - ib$ is $a + ib$. Check that the conjugate of a real number a is a again.

We could do this whole course over $\mathbb{C}$ but for simplicity, and for the sake of geometry, we will stick to $\mathbb{R}$. Almost all the results we will see also hold for complex matrices with slight modifications. We will revisit complex matrices and vector spaces at the end of this course.

Fundamental Theorem of Algebra

Let $p(t)$ be a degree n polynomial in the variable t with complex numbers as coefficients. Then $p(t)$ has (counting multiplicities), precisely n complex roots $\mu_1, \ldots, \mu_n$. Equivalently, there is some constant $c \in \mathbb{C}$ such that $p(t)$ can be factored as

$$p(t) = c(t - \mu_1)(t - \mu_2) \cdots (t - \mu_n).$$

We make a few comments:

1. The *multiplicity* of a root μ_i is the number of times it appears as a root of $p(t)$. If we denote the multiplicity of μ_i by α_i, then we can factor $p(t)$ as

   $$p(t) = c(t - \mu_1)^{\alpha_1}(t - \mu_2)^{\alpha_2} \cdots (t - \mu_k)^{\alpha_k}$$

   where $\alpha_1 + \alpha_2 + \cdots + \alpha_k = n$.

2. If all the coefficients of the polynomial $p(t)$ are real, and $a + ib$ is a complex root of $p(t)$, then so is $a - ib$ which is called the *complex conjugate* of $a + ib$. *Do you see why? Try a small example.*

   Example 1.1.5. (a) $p(t) = t^2 - 1 = 0 \implies t = \pm 1$ and $t^2 - 1 = (t - 1)(t + 1)$.
   (b) $p(t) = 2t^2 + 1 = 0 \implies t^2 = -\frac{1}{2} \implies t = \pm \frac{i}{\sqrt{2}}$ and $2t^2 + 1 = 2(t - \frac{i}{\sqrt{2}})(t + \frac{i}{\sqrt{2}})$.

   Notice that the coefficients of both polynomials above are real numbers. In the first case, $p(t)$ has only real roots and in the second case it has two complex roots that are conjugates.

3. If $p(\lambda) = \det(A - \lambda I)$ is the characteristic polynomial of the square matrix $A \in \mathbb{R}^{n \times n}$ then all coefficients of $p(\lambda)$ are real. However, the some roots of $p(\lambda)$ may be complex and if so, they come in conjugate pairs. The factorization in this case looks like

   $$p(\lambda) = \det(A - \lambda I) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n).$$

   Question: Do you see why $(-1)^n$ appears as the constant in front of the factorization?

This brings us to the first general fact about eigenvectors and eigenvalues of a matrix $A \in \mathbb{R}^{n \times n}$.

Theorem 1.1.6. *A matrix $A \in \mathbb{C}^{n \times n}$ (in particular, in $\mathbb{R}^{n \times n}$) has n eigenvalues (counting multiplicities).*

Let's recap why. The characteristic polynomial $p(\lambda) = \det(A - \lambda I)$ of a matrix $A \in \mathbb{C}^{n \times n}$ is a degree n polynomial with complex coefficients. Therefore, by the Fundamental Theorem of Algebra it has n roots (counting multiplicities) and these roots are the eigenvalues of A.

1.1.2 Arithmetic and geometric multiplicities

Definition 1.1.7. Suppose the characteristic polynomial of A factors as

$$p(\lambda) = \det(A - \lambda I) = (-1)^n (\lambda - \lambda_1)^{\alpha_1} (\lambda - \lambda_2)^{\alpha_2} \cdots (\lambda - \lambda_k)^{\alpha_k}$$

where $\alpha_1 + \alpha_2 + \cdots + \alpha_k = n$ and $\lambda_1, \ldots, \lambda_k$ are the eigenvalues of A. The **algebraic multiplicity** of the eigenvalue λ_i is the exponent α_i, denoted as $\mathrm{AM}(\lambda_i)$.

Definition 1.1.8. The **geometric multiplicity** of the eigenvalue λ_i, denoted $\mathrm{GM}(\lambda_i)$, is the dimension of the eigenspace E_{λ_i}, i.e., $\mathrm{GM}(\lambda_i) = \dim(E_{\lambda_i})$.

These two multiplicities are sometimes different as the following example shows.

Example 1.1.9. Let $A = \begin{bmatrix} 8 & -9 \\ 4 & -4 \end{bmatrix}$. Check that $p(\lambda) = (\lambda - 2)^2$ and $\mathrm{AM}(2) = 2$. On the other hand,

$$A - \lambda I = \begin{bmatrix} 8-\lambda & -9 \\ 4 & -4-\lambda \end{bmatrix} \implies A - 2I = \begin{bmatrix} 6 & -9 \\ 4 & -6 \end{bmatrix}$$

which shows that the columns of $A - 2I$ are linearly dependent. Since it is not the zero matrix, we conclude that $\mathrm{rank}(A - 2I) = 1$ which implies that $\mathrm{nullity}(A - 2I) = 1$. Hence, $\dim(E_2) = \dim(\mathrm{Null}(A - 2I)) = 1$, and $\mathrm{GM}(2) < \mathrm{AM}(2)$.

In general, we have the following relationship between the two multiplicities.

Proposition 1.1.10. *For an eigenvalue λ of $A \in \mathbb{R}^{n \times n}$, $GM(\lambda) \leq AM(\lambda)$ always.*

1.1.3 Complex eigenvalues

From the Fundamental Theorem of Algebra we know that some eigenvalues of a real matrix can be complex. Rotation matrices are good examples of real matrices with complex eigenvalues. We have no easy way to see geometry in complex eigenvectors since we can only draw in $\mathbb{R}^n$, but we will develop some machinery later that will provide insight.

Example 1.1.11. Let A denote the linear transformation that rotates any vector in $\mathbb{R}^2$ by $\pi/2$. Clearly there is no real vector that is only scaled under rotation by $\pi/2$ and so we do not expect any real eigenvectors for the matrix of this linear transformation.

Check that the matrix representing this rotation is

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

In general, the matrix that represents rotation in $\mathbb{R}^2$ by θ is

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

We compute the characteristic polynomial of A and see that

$$\det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + 1 = 0 \implies \lambda = \pm i$$

Both eigenvalues are complex in this case. Further, $E_i = \mathrm{Span}\left\{ \begin{bmatrix} i \\ 1 \end{bmatrix} \right\}$ and $E_{-i} = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ i \end{bmatrix} \right\}$.

Example 1.1.12. Now suppose B denotes rotation in $\mathbb{R}^2$ by π rather than $\pi/2$. Then we do expect real eigenvectors since every vector $\mathbf{x} \in \mathbb{R}^2$ will simply flip sign and become $-\mathbf{x}$ under this rotation. Hence -1 is an eigenvalue and we might guess that $E_{-1} = \mathbb{R}^2$. Let's confirm this with calculations.

We have three ways to compute B: (i) compute B using the standard basis of $\mathbb{R}^2$, or (ii) compute B from the R_θ in the previous example by setting $\theta = \pi$, or ((iii) we could obtain B from A by recalling that the product of matrices corresponds to the composition of linear transformations. Since rotating by π is the same as rotating twice by $\pi/2$, we have that

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

and hence,

$$\det \begin{bmatrix} -1-\lambda & 0 \\ 0 & -1-\lambda \end{bmatrix} = (-1-\lambda)^2 = (\lambda+1)^2 = 0 \implies \lambda = -1 \text{ with } \mathrm{AM}(-1) = 2.$$

Therefore -1 is the only eigenvalue of B with arithmetic multiplicity 2, and B is the rare rotation matrix with only real eigenvalues. Check that

$$E_{-1} = \mathrm{Null}(A^2 + I) = \mathrm{Null}(-I + I) = \mathrm{Null}(O_{22}) = \mathbb{R}^2$$

and so our guess above was correct and $\mathrm{GM}(-1) = \mathrm{AM}(-1) = 2$.

1.1.4 Eigenvalues of powers and inverses

We now state a few general facts about eigenvalues and eigenvectors under various matrix operations. You will derive more facts in homework.

1. If λ is an eigenvalue of A then λ^k is an eigenvalue of A^k, and if $\mathbf{x}$ is an eigenvector of A with eigenvalue λ then it is also an eigenvector of A^k with eigenvalue λ^k.

   Why?

   $$A\mathbf{x} = \lambda\mathbf{x} \implies A^2\mathbf{x} = A(A\mathbf{x}) = A\lambda\mathbf{x} = \lambda A\mathbf{x} = \lambda^2\mathbf{x} \implies A^3\mathbf{x} = A(A^2\mathbf{x}) = A\lambda^2\mathbf{x} = \lambda^2 A\mathbf{x} = \lambda^3\mathbf{x}$$

   Continuing with this process we can see that

   $$A^k\mathbf{x} = A(A^{k-1})\mathbf{x} = A\lambda^{k-1}\mathbf{x} = \lambda^{k-1}A\mathbf{x} = \lambda^k\mathbf{x}$$

   Question: does this account for all eigenvalues of A^k?

2. If λ is an eigenvalue of A and A is invertible, then $\lambda^{-1} = \frac{1}{\lambda}$ is an eigenvalue of A^{-1}.

   Why? If $A\mathbf{x} = \lambda\mathbf{x}$, we can multiply both sides of this equation on the left by A^{-1} (which we know exists!), and get $\mathbf{x} = A^{-1}\lambda\mathbf{x}$ which implies that $\frac{1}{\lambda}\mathbf{x} = A^{-1}\mathbf{x}$. Note that again the eigenvalues change but the eigenvectors don't.

In compact notation, we write a matrix $A \in \mathbb{R}^{m\times n}$ as

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Here is a quick notational refresher, which we will use *a lot*.

Definition 1.1.13. The *transpose* of a matrix $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, denoted as $A^\top$, is the matrix obtained by swapping rows and columns of A. That is $A^\top = (a_{ji}) \in \mathbb{R}^{n\times m}$.

Example 1.1.14. If $A = \begin{bmatrix} 1 & -1 & 0 \\ 3 & 1 & -2 \end{bmatrix}$ then $A^\top = \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}$.

In homework you will have to think about the relationship between the eigenvalues of A and $A^\top$.

1.2 Diagonalization

Definition 1.2.1. A matrix $A \in \mathbb{R}^{n\times n}$ is **diagonalizable** if there exists an invertible matrix $U \in \mathbb{R}^{n\times n}$ and a diagonal matrix $\Lambda \in \mathbb{R}^{n\times n}$ (Λ is the capital lambda) such that $A = U\Lambda U^{-1}$.

Suppose $A = U\Lambda U^{-1}$ where $\Lambda = \operatorname{diag}((\lambda_1, \lambda_2, \dots, \lambda_n)$. Then multiplying both sides by U on the right, we get that $AU = U\Lambda$. if $\mathbf{u}_i$ denotes the ith column of U, then this implies that $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$ for each $i = 1, \dots, n$. Therefore, λ_i, the entry in position (i, i) of Λ, is an eigenvalue of A with eigenvector $\mathbf{u}_i$, which is the the ith column of U. For A to be diagonalizable, U needs to be invertible which means that A has to have n linearly independent eigenvectors, i.e., $\mathbb{R}^n$ has a basis consisting of eigenvectors of A.

Theorem 1.2.2. *A matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if it has n linearly independent eigenvectors.*

Recall that the eigenvectors of A lie in the eigenspaces of A. We saw earlier that the dimension of the eigenspace E_λ can be smaller than the multiplicity of λ. If this happens for some eigenvalue λ, then we will not have enough linearly independent eigenvectors of A to span $\mathbb{R}^n$ and A will not be diagonalizable. So we can write down a finer (equivalent) condition for diagonalizability as follows.

Theorem 1.2.3. *$A \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if $\mathrm{AM}(\lambda) = \mathrm{GM}(\lambda)$ for all eigenvalues λ.*

We now look at three examples:

Example 1.2.4. Recall $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ from 1.1.4. The eigenvalues of this matrix were $\lambda = \pm 1$ with eigenspaces $E_1 = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ and $E_{-1} = \mathrm{Span}\left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$. Using the basis vectors for our eigenspaces we obtain the eigenbasis of $\mathbb{R}^2$ given by $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$. A is therefore, diagonalizable:

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

In this example, $\mathrm{AM}(\lambda) = \mathrm{GM}(\lambda)$ for both eigenvalues $\lambda = \pm 1$.

Example 1.2.5. Now consider the matrix $B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ from Example 1.1.12. This matrix has only one eigenvalue $\lambda = -1$ but since $E_{-1} = \mathbb{R}^2$ it has two linearly independent eigenvectors. Hence B is diagonalizable. For instance, we could take the standard basis of $\mathbb{R}^2$ as two eigenvectors in E_{-1} and write

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

Example 1.2.6. Here is an example of a matrix that is not diagonalizable. Let $A = \begin{bmatrix} 6 & -1 \\ 1 & 4 \end{bmatrix}$. Computing eigenvalues and eigenvectors, we see the following:

$$\det \begin{bmatrix} 6-\lambda & -1 \\ 1 & 4-\lambda \end{bmatrix} = \lambda^2 - 10\lambda + 25 = (\lambda - 5)^2 = 0 \implies \lambda = 5 \text{ with } \mathrm{AM}(5) = 2$$

Since eigenvalues are not distinct, A may not be diagonalizable. Computing the eigenspace of $\lambda = 5$,

$$E_5 = \text{Null}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} = \text{Null}\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

so $\text{GM}(5) = 1$ and $\text{AM}(5) \neq \text{GM}(5)$. This means A is not diagonalizable.

In general, it takes some work to see if a matrix is diagonalizable. Here is a special case in which a matrix is always diagonalizable.

Theorem 1.2.7. *If $A \in \mathbb{R}^{n\times n}$ has n* ***distinct*** *eigenvalues, then A is diagonalizable.*

Proof. First note that if λ_i is an eigenvalue of A it **must** have some non-zero eigenvector $\mathbf{x}$ such that $A\mathbf{x} = \lambda_i\mathbf{x}$. Moreover, for any $\mathbf{x} \in E_{\lambda_i}$, we know that $c\mathbf{x} \in E_{\lambda_i}$ for any $c \in \mathbb{R}$, thus if $\mathbf{x} \in E_{\lambda_i}$ then $\text{Span}\{\mathbf{x}\} \subseteq E_{\lambda_i}$. This means that for each eigenvalue λ_i, we have $\text{GM}(\lambda_i) \geq 1$.

If A has n distinct eigenvalues, then we can list them out as $\lambda_1, \lambda_2, \ldots, \lambda_n$ and are guaranteed that $\lambda_i \neq \lambda_j$ for all $i \neq j$. This means that $\text{AM}(\lambda_i) = 1\ \forall i$. From Proposition 1.1.10 we can conclude that

$$1 \leq \text{GM}(\lambda_i) \leq \text{AM}(\lambda_i) = 1$$

hence $\text{AM}(\lambda_i) = \text{GM}(\lambda_i)\ \forall i$. The result now follows from Theorem 1.2.3. □

1.2.1 Applications of Diagonalization

Powers of matrices

Powers of diagonalizable matrices are very easy to compute. If $A = U\Lambda U^{-1}$ then

$$A^k = \underbrace{(U\Lambda U^{-1})(U\Lambda U^{-1})\cdots(U\Lambda U^{-1})}_{k \text{ times}} = U\Lambda^k U^{-1}$$

Recall that a power of a diagonal matrix is easy to compute: $\Lambda^k = \text{diag}(\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k)$.

For the A in Example 1.1.4, we see that

$$A^{53} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{53} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}^{53}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

A diagonalizable matrix behaves like a diagonal matrix

The action of a diagonal matrix Λ is very easy to understand, since it sends $\mathbf{x}$ to $\Lambda\mathbf{x} = (\lambda_1 x_1, \lambda_2 x_2, \ldots, \lambda_n x_n)^\top$. Check this on an example. Diagonalizable matrices behave like diagonal matrices if we are willing to change bases as we now explain. This is perhaps the greatest use of diagonalization!

Suppose A is diagonalizable and $A = U\Lambda U^{-1}$. This means that the columns of U which we denote by $\mathbf{u}_1, \ldots, \mathbf{u}_n$ form a basis for $\mathbb{R}^n$. Take a vector $\mathbf{x} \in \mathbb{R}^n$ represented in the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Recall that the coordinates of $\mathbf{x}$ in the basis $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is the vector $\mathbf{y}$ such that $U\mathbf{y} = \mathbf{x}$ or alternately, $\mathbf{y} = U^{-1}\mathbf{x}$. This formula is extremely important to remember!

Change of basis formula

If a point in $\mathbb{R}^n$ has coordinates $\mathbf{x}$ with respect to the standard basis, then its coordinates with respect to the basis $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of $\mathbb{R}^n$ is $\mathbf{y} = U^{-1}\mathbf{x}$, where U is the matrix whose ith column is $\mathbf{u}_i$.

Now suppose we use the basis $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ instead of the standard basis of $\mathbb{R}^n$. Then the coordinates of $\mathbf{x}$ with respect to the basis $\mathcal{U}$ is $\mathbf{y} = U^{-1}\mathbf{x}$ and the coordinates of $A\mathbf{x}$ with respect to the basis $\mathcal{U}$ is

$$U^{-1}A\mathbf{x} = U^{-1}U\Lambda U^{-1}\mathbf{x} = \Lambda U^{-1}\mathbf{x} = \Lambda\mathbf{y}.$$

Therefore in the $\mathcal{U}$ basis, a vector $\mathbf{y}$ is sent to $\Lambda\mathbf{y}$ which rather simple to understand. In other words, the action of the matrix A is the same as the action of the diagonal matrix Λ if we are willing to change the basis of $\mathbb{R}^n$ from the standard basis to the eigenbasis $\mathcal{U}$ of A.

1.3 Similar Matrices

We finish the chapter with a completely new notion.

Definition 1.3.1. Two matrices A and C are **similar**, denoted as $A \sim C$, if there exists an invertible matrix B such that

$$A = BCB^{-1}$$

There is one main example that probably comes to mind.

Example 1.3.2. If A is diagonalizable, then $A = U\Lambda U^{-1}$ for some U, and hence $A \sim \Lambda$.

Theorem 1.3.3. *If $A \sim C$ then A and C have the same eigenvalues.*

Proof. If $A \sim C$ then $\exists$ an invertible matrix B such that $A = BCB^{-1}$. Our goal is to show that A and C have the same eigenvalues. Suppose $C\mathbf{x} = \lambda\mathbf{x}$, then since $AB = BC$ we know that $AB\mathbf{x} = BC\mathbf{x}$, hence

$$A(B\mathbf{x}) = B(C\mathbf{x}) = B\lambda\mathbf{x} = \lambda B\mathbf{x}$$

This implies that λ is an eigenvalue of A (with eigenvector $B\mathbf{x}$), hence all eigenvalues of C are also eigenvalues of A. Similarly we could start with an eigenvalue of A and show that it is an eigenvalue of C. Therefore, we have that the set of eigenvalues of C is a subset of the set of eigenvalues of A and the set of eigenvalues of A is a subset of the set of eigenvalues of C. This means that the eigenvalues of A and C are the same. *In general, if F and G are two sets and $F \subseteq G$ and $G \subseteq F$, then $F = G$. This is because, the assumptions imply that $F \subseteq G \subseteq F$, but since the left and right ends are equal, it must be that there is equality throughout.* □

1.4 Determinants and Permutations

In this section we will see a fact about the determinant of a square matrix that you perhaps did not see in Math 308. This will help us understand the characteristic polynomial a bit more.

You might know that the number called *n factorial* (written as $n!$) is $1 \times 2 \times \cdots \times (n-1) \times n$. A *permutation* of the numbers $1, 2, \ldots, n$ is an ordering of the numbers. For example, there are $2 = 2!$ permutations of $1, 2$, namely 12 and 21 (we write the numbers in a string with no commas separating them). There are $6 = 1 \times 2 \times 3 = 3!$ permutations of $1, 2, 3$, namely $123, 132, 213, 231, 312, 321$. There are $24 = 4!$ permutations of $1, 2, 3, 4$. In general, there are $n! = 1 \times 2 \times \cdots \times (n-1) \times n$ permutations of $1, 2, \ldots, n$. You can also think of a permutation as a function $\pi : \{1, 2, 3, \ldots, n\} \to \{1, 2, 3, \ldots, n\}$. For example, the permutation 1432 is the function π that sends $1 \mapsto 1, 2 \mapsto 4, 3 \mapsto 3, 4 \mapsto 2$, alternately, $\pi(1) = 1, \pi(2) = 4, \pi(3) = 3, \pi(4) = 2$.

It turns out that there is a close connection between all permutations of n (which is a short way of saying all permutations of $1, 2, \ldots, n$) and the determinant of a square matrix. Let's illustrate on the following 3×3 symbolic matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Computing the determinant of A by cofactor expansion (say along the first row) we have

$$\begin{aligned}\det(A) &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}\end{aligned}$$

The above expression is a polynomial of degree 3 in the variables $a_{11}, \ldots, a_{33}$. It consists of 6 *terms* and each term is the product of a *coefficient* which is $+1$ or -1 and a *monomial* which is a product of variables. For example, the second term in the above polynomial is $-a_{11}a_{23}a_{32}$ whose coefficient is -1 and monomial is $a_{11}a_{23}a_{32}$. Notice the following facts:

- There are 3! terms in the determinant of the above 3×3 matrix.
- Each monomial consists of three variables, one from each row and column of A (i.e., two variables in a monomial do not come from the same row or same column and all rows and columns appear).

This makes us suspect that there is a term in the determinant for each permutation of 3. Let's check. First, note that we have already written each term so that the rows the variables come from are in the order $1, 2, 3$. We can always do this since every row contributes exactly one variable to a monomial and multiplication of real and complex numbers is commutative. Once we have fixed the order of the rows to be 1,2,3, notice that the order in which the column indices appear in each monomial is a permutation of 3, and all $6 = 3!$ permutations appear in the terms of the determinant. Let's check on our example:

$$\underbrace{a_{11}a_{22}a_{33}}_{123} - \underbrace{a_{11}a_{23}a_{32}}_{132} - \underbrace{a_{12}a_{21}a_{33}}_{213} + \underbrace{a_{12}a_{23}a_{31}}_{231} + \underbrace{a_{13}a_{21}a_{32}}_{312} - \underbrace{a_{13}a_{22}a_{31}}_{321}$$

The coefficient that appears in front of a monomial is also determined by the permutation of the column indices. A pair i, j is said to be *inverted* in a permutation π if $i < j$ but $\pi(i) > \pi(j)$. For example in the permutation 1432, 2 and 4 are inverted since $\pi(2) = 4 > \pi(4) = 2$ even though $2 < 4$. We say that a permutation is *even* if it has an even number of inverted pairs and *odd* if it has an odd number of inverted pairs. The *sign of a permutation* is 1 if the permutation is even and -1 if the permutation is odd.

Notice that the coefficient in front of the term in the determinant indexed by a permutation π is exactly the sign of π. For example take the third term $-a_{12}a_{21}a_{33}$ which is indexed by the permutation 213. There is exactly one inversion in this permutation which is given by the pair $(1, 2)$ since $\pi(1) = 2 > \pi(2) = 1$. The pairs $(1, 3)$ and $(2, 3)$ are not inverted in this permutation. Therefore, this is an odd permutation and its sign is -1 which is the coefficient of the term $-a_{12}a_{21}a_{33}$. Check every term in the above determinant and make sure you agree that the coefficient in front of the term in the determinant indexed by a permutation π is exactly the sign of π. Let's check that the same behavior can be seen in 2×2 matrices:

Example 1.4.1.

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

The first term corresponds to the permutation 12 with sign 1, and the second term corresponds to the permutation 21 with sign -1.

This brings us to the following general theorem (which we will not prove):

Theorem 1.4.2. *If $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, then*

$$\det(A) = \sum_{\pi \text{ permutation of } n} \operatorname{sign}(\pi) a_{1\pi(1)}a_{2\pi(2)} \cdots a_{n\pi(n)}.$$

This formula helps us understand the characteristic polynomial of a matrix even more. Let's take the 3×3 matrix A from above again. Then:

$$p(\lambda) = \det(A - \lambda I) = \det \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{bmatrix}.$$

The term in the determinant indexed by the permutation 123 is $(a_{11}-\lambda)(a_{22}-\lambda)(a_{33}-\lambda) = -\lambda^3 + \cdots$ No other term in the determinant contributes a λ^3 or even λ^2, only lower powers of λ, so the λ^3 we got from the first term will not cancel out (write out the determinant and check!). Therefore, $p(\lambda)$ is a polynomial in λ of degree 3 and the coefficient of λ^3 is $(-1)^3 = -1$.

Do you now see why if $A \in \mathbb{R}^{n\times n}$, then $p(\lambda) = \det(A - \lambda I)$ is a degree n polynomial in λ and its leading coefficient is $(-1)^n$? **Hint**: As above, look at the term in the determinant indexed by the permutation $123\cdots n$. What is the coefficient of this term and what is the highest power of λ in it and why does this highest power not cancel out when you take all terms together?

1.5 A Quick Primer on Logic

In this last section, we give a quick and simple introduction to making logical arguments in mathematics. You will be asked to show that a statement "P" implies the statement "Q", written as $P \implies Q$ and read as "P implies Q" or "if P then Q". The statement P is the *hypothesis* (which is always true) and the statement Q is the *conclusion.* While no formal proof writing skills are expected in this course, you will often be asked to *argue* that something is true or false. Here are three common ways to make such arguments.

Direct: The direct method is the most straightforward. Given an *if-then* statement of the form $P \implies Q$, one assumes the hypothesis P and works to conclude Q.

Example 1.5.1. Fix a matrix $A \in \mathbb{R}^{m\times n}$ and let $S = \Big\{\mathbf{x} \in \mathbb{R}^n \colon A^2\mathbf{x} = A\mathbf{x}\Big\}$. Argue that if $\mathbf{x}, \mathbf{y} \in S$, then $\mathbf{x}+\mathbf{y} \in S$.

Here P is "$\mathbf{x}, \mathbf{y} \in S$" and Q is "$\mathbf{x}+\mathbf{y} \in S$". Doing this directly, we *assume* that $\mathbf{x}, \mathbf{y} \in S$ and want to conclude that $\mathbf{x}+\mathbf{y} \in S$. Since $\mathbf{x}, \mathbf{y} \in S$ we know that $A^2\mathbf{x} = A\mathbf{x}$ and $A^2\mathbf{y} = A\mathbf{y}$. Adding them up, we check that they are still in S:

$$A^2(\mathbf{x}+\mathbf{y}) = A^2\mathbf{x} + A^2\mathbf{y} = A\mathbf{x} + A\mathbf{y} = A(\mathbf{x}+\mathbf{y})$$

Note that the second equality is where we used our assumption P. The rest came from matrix algebra.

The next method is often a good idea when the direct method looks hard.

Contrapositive: Given the statement,"if P then Q", its contrapositive is the statement " if Q is false then P is false" written as $\sim Q \Rightarrow \sim P$.

Indeed, $P \Rightarrow Q$ says that Q is true whenever P is true. This is the same as saying that if Q is false, then P must have been false too.

It is important to note that $P \Rightarrow Q$ is NOT the same as either $Q \Rightarrow P$ or $\sim P \Rightarrow \sim Q$. For example, consider the following true statement " if p is a multiple of 4 then p is even". Here P is "p is a multiple of 4" and Q is "p is even". The contrapositive $\sim Q \implies \sim P$ is "if p is not even then p is not a multiple of 4" which is true. Check that the following are NOT true:

$$Q \implies P\colon \quad p \text{ even} \implies p \text{ is a multiple of } 4$$
$$\sim P \implies \sim Q\colon \quad p \text{ not a multiple of } 4 \implies p \text{ is not even}$$

Example 1.5.2. Consider the statement, "if $x^2 - 6x + 5$ is even, then x is odd". In doing this directly, we would assume that there is some number a such that $x^2 - 6x + 5 = 2a$ but then we would need to show that $x = 2b + 1$ for some number b, and this feels hard. It turns out the contrapositive makes this much more tractable. The contrapositive statement is that "if x is not odd, then $x^2 - 6x + 5$ is not even", in other words, "if x is even, then $x^2 - 6x + 5$ is odd". Assuming that $x = 2a$ for some number a, we plug it into the equation and get that

$$(2a)^2 - 6(2a) + 5 = 4a^2 - 12a + 4 + 1 = 2(2a^2 - 6a + 2) + 1$$

If we set $b = 2a^2 - 6a + 2$ then the last expression above is of the form $2b + 1$ which is odd and so $x^2 - 6x + 5$ is odd.

The last technique has a slightly different flavor but can also be very effective if you're stuck.

Contradiction: Suppose we need to prove that $P \Rightarrow Q$. If by assuming P is true **and** Q is false we are able to get a contradiction (an obviously false statement), then it must be that something in our assumption was wrong. Since the hypothesis P is never wrong, it must be that assuming "Q is false" was wrong. We then conclude that if P is true then Q is also true.

Example 1.5.3. Consider the statement, "If $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ is a basis for $\mathbb{R}^n$, then $m = n$". Here P is "$\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ is a basis of $\mathbb{R}^n$" and Q is "$m = n$". To prove what we want by contradiction we assume "$\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ is a basis of $\mathbb{R}^n$" and $m \neq n$. The latter means that either $n < m$ or $n > m$. We consider both cases:

- If $n < m$ then we have more than n vectors and no set of more than n vectors can be linearly independent in $\mathbb{R}^n$. Therefore, $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ is not a basis of $\mathbb{R}^n$ which contradicts our assumption.
- If $n > m$, then we have a basis for $\mathbb{R}^n$ consisting of fewer than n vectors, but no set of fewer than n vectors can ever span $\mathbb{R}^n$, giving us another contradiction.

Therefore, our assumption that $m \neq n$ must have been false and we conclude that if $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ is a basis for $\mathbb{R}^n$, then $m = n$.

Here are two further logic facts:

When are two sets equal: By definition, two sets, A and B, are equal if every element of A is also an element of B, and similarly, every element of B is an element of A. If only one of these conditions hold, say every element of A is an element of B, but not every element of B is an element of A, then we say A is a *subset* of B and write $A \subseteq B$.

The key idea to prove that $A = B$ is to take an arbitrary element of one set, and show it belongs to the other, and then repeat the process in the other direction, i.e., we need to show the following:

1. Pick an arbitrary element $a \in A$, and show that $a \in B$. Since a was arbitrary, we conclude that every element of A lies in B and so $A \subseteq B$.
2. Similarly, pick an arbitrary element $b \in B$ and show that $b \in A$. This shows that $B \subseteq A$.

Picking an arbitrary element means picking an element that has the property specified in the hypothesis P but nothing more. For example, if P says that "x is even", then set $x = 2a$ which is the general representation of an even number. Don't pick $x = 4a$ or $x = 10$ since these represent only *some* even numbers.

To summarize, we have that $A = B$ if and only if $A \subseteq B$ *and* $B \subseteq A$. We saw this in action in the proof of Theorem 1.3.3. This will prove to be useful a number of times throughout the course. Remember, at its core, **many** of the things we look at are sets! For example, $\mathrm{Col}(A), \mathrm{Row}(A), \mathrm{Null}(A), \mathrm{Range}(T)$, and $\ker(T)$ are all sets, so if we want to argue that any two of them are equal, we use the above method.

Chapter 2

Difference Equations

In this chapter we see an application of diagonalization to solving difference equations. We will develop the theory we need by working out an example. The material in this chapter comes from Strang's Chapter 6.2.

2.1 Fibonacci sequence

You might have heard of the *Fibonacci sequence* $0, 1, 1, 2, 3, 5, 8, \ldots$ defined recursively as follows: if F_k denotes the k^{th} Fibonacci number, where $k = 0, 1, 2, 3, \ldots$, then

$$F_0 = 0,\ F_1 = 1,\ F_{k+2} = F_{k+1} + F_k.$$

This recursive definition let's us enumerate any Fibonacci number we want

$$F_0 = 0,\ F_1 = 1,\ F_2 = 1,\ F_3 = 2,\ F_4 = 3,\ F_5 = 5,\ F_6 = 8,\ F_7 = 13,\ F_8 = 21,\ F_9 = 33,\ F_{10} = 54, \ldots$$

The numbers get large relatively quickly and it becomes cumbersome to use the recursive definition to find a large Fibonacci number like F_{100}. However, linear algebra makes this easy as we now see.

Define $\mathbf{w}_k := \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$ which means that $\mathbf{w}_{k+1} = \begin{bmatrix} F_{k+2} \\ F_{k+1} \end{bmatrix}$. Check that $\mathbf{w}_k$ and $\mathbf{w}_{k+1}$ can be related via multiplication by a matrix A. Indeed, if $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, then $A\mathbf{w}_k = \mathbf{w}_{k+1}$. Let's check:

$$A\mathbf{w}_k = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{w}_k = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \begin{bmatrix} F_{k+1} + F_k \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} F_{k+2} \\ F_{k+1} \end{bmatrix} = \mathbf{w}_{k+1}.$$

Now suppose we start this process with $\mathbf{w}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. We know that $A\mathbf{w}_0 = \mathbf{w}_1$ and multiplying both sides of this equation with A we see that

$$\mathbf{w}_2 = A\mathbf{w}_1 = A(A\mathbf{w}_0) = A^2\mathbf{w}_0$$

Continuing to multiply by powers of A on both sides, we see that $\mathbf{w}_k = A^k\mathbf{w}_0$ and hence $\mathbf{w}_{100} = A^{100}\mathbf{w}_0$. Recall that we can "easily" compute A^{100} if A is diagonalizable.

We find the eigenvalues of A by computing the roots of its characteristic polynomial:

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

which yields the eigenvalues $\lambda_1 = \frac{1+\sqrt{5}}{2} \sim 1.618$ and $\lambda_2 = \frac{1-\sqrt{5}}{2} \sim -0.618$. Now computing eigenvectors we get that $\mathbf{u}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$ is an eigenvector of λ_1 and $\mathbf{u}_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$ is an eigenvector of λ_2. Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is a basis of $\mathbb{R}^2$, we can diagonalize A and write $A = U\Lambda U^{-1}$. This means that

$$\mathbf{w}_k = A^k \mathbf{w}_0 = U\Lambda^k U^{-1}\mathbf{w}_0$$

Putting everything together we have that

$$\begin{bmatrix} F_{101} \\ F_{100} \end{bmatrix} = \mathbf{w}_{100} = A^{100}\mathbf{w}_0 = U\Lambda^{100}U^{-1}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = U \begin{bmatrix} (\frac{1+\sqrt{5}}{2})^{100} & 0 \\ 0 & (\frac{1-\sqrt{5}}{2})^{100} \end{bmatrix} U^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Multiplying everything out we could find F_{100}, but this still requires some matrix computations and computing Λ^{100} is not too easy in this example since it requires finding $(\frac{1+\sqrt{5}}{2})^{100}$ and $(\frac{1-\sqrt{5}}{2})^{100}$.

We will now see that there is an easier way to find F_{100}. In fact, there is a closed form expression for the kth Fibonacci number F_k purely in terms of the eigenvalues and eigenvectors of A.

2.2 The general setting

Before we start, we point out an important interpretation of matrix-vector multiplication. These are very useful facts that we will need over and over again in this class. Please test these facts on some examples and make sure you understand these interpretations very very well.

Matrix-vector multiplication as a linear combination of columns/rows

Let $A \in \mathbb{R}^{m\times n}$ with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ and rows $\mathbf{b}_1^\top, \mathbf{b}_2^\top, \dots, \mathbf{b}_m^\top$ where $\mathbf{a}_i \in \mathbb{R}^m$ and $\mathbf{b}_j \in \mathbb{R}^n$.

- The vector $A\mathbf{x}$ is the linear combination of the columns of A by $x_1, x_2, \dots, x_n$. It is a column vector whose ith entry is the dot product of the ith row of A with $\mathbf{x}$. Mathematically, if $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{R}^{m\times n}$ and $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$, then $A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$ and $(A\mathbf{x})_i = \mathbf{b}_i^\top \mathbf{x}$.

- The vector $\mathbf{y}^\top A$ is the linear combination of the rows of A by $y_1, \dots, y_m$. It is a row vector whose jth entry is the dot product of $\mathbf{y}$ and the jth column of A. Mathematically, if $\mathbf{y} = (y_1, \dots, y_m)^\top$ and $A = \begin{bmatrix} \mathbf{b}_1^\top \\ \mathbf{b}_2^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix} \in \mathbb{R}^{m\times n}$, then $\mathbf{y}^\top A = y_1\mathbf{b}_1^\top + y_2\mathbf{b}_2^\top + \ldots + y_m\mathbf{b}_m^\top$ and $(\mathbf{y}^\top A)_j = \mathbf{y}^\top \mathbf{a}_j$.

Using the above (check!) we get the following two very useful facts about the Col(A), the column space of A which is the span of the columns of A, and Row(A), the rowspace of A which is the span of the rows of A:

Column and row space of a matrix

$$\mathrm{Col}(A) = \{A\mathbf{x} \,:\, \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m \qquad \text{and} \qquad \mathrm{Row}(A) = \{\mathbf{y}^\top A \,:\, \mathbf{y} \in \mathbb{R}^m\} \subseteq \mathbb{R}^n$$

Now let's come back to difference equations. Suppose a matrix $A \in \mathbb{R}^{n\times n}$ sends an *initial state vector* $\mathbf{w}_0$ to the k^{th} *state vector*, $\mathbf{w}_k$ via the relation $A^k\mathbf{w}_0 = \mathbf{w}_k$. Recall that we got this from the *difference equation* $A\mathbf{w}_k = \mathbf{w}_{k+1}$ that related two consecutive states of some system whose kth state vector is $\mathbf{w}_k$. If

A is diagonalizable, then $A = U\Lambda U^{-1}$ and $\mathbf{w}_k = A^k\mathbf{w}_0 = U\Lambda^k U^{-1}\mathbf{w}_0$.

Now set $\mathbf{c} := U^{-1}\mathbf{w}_0$. Recall that in the eigenbasis of $\mathbb{R}^n$ consisting of the columns of U, the coordinates of $\mathbf{w}_0$ is $\mathbf{c} = U^{-1}\mathbf{w}_0$. You can find the vector $\mathbf{c}$ by solving the system $U\mathbf{c} = \mathbf{w}_0$.

Now consider $\mathbf{w}_k = A^k\mathbf{w}_0$ again where we substitute $\mathbf{c}$ for $U^{-1}\mathbf{w}_0$:

$$\mathbf{w}_k = U\Lambda^k U^{-1}\mathbf{w}_0 = U\Lambda^k\mathbf{c} = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix}\begin{bmatrix}\lambda_1^k & 0 & \dots & 0\\ 0 & \lambda_2^k & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n^k\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ \vdots\\ c_n\end{bmatrix}$$

$$= \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix}\begin{bmatrix}\lambda_1^k c_1\\ \lambda_2^k c_2\\ \vdots\\ \lambda_n^k c_n\end{bmatrix} = c_1\lambda_1^k\mathbf{u}_1 + c_1\lambda_2^k\mathbf{u}_2 + \cdots + c_n\lambda_n^k\mathbf{u}_n$$

which expresses the kth state vector $\mathbf{w}_k$ as a linear combination of the eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ of A. The scalars involved in the combination come from the eigenvalues of A as well as the vector $\mathbf{c}$. We summarize the above calculations in the following proposition.

Proposition 2.2.1. *Suppose we have a difference equation $A\boldsymbol{w}_k = \boldsymbol{w}_{k+1}$ with initial state vector $\boldsymbol{w}_0$ and $A \in \mathbb{R}^{n\times n}$. Then $\boldsymbol{w}_k = A^k\boldsymbol{w}_0$. If A is diagonalizable then we can explicitly solve for $\boldsymbol{w}_k$ as:*

$$\boldsymbol{w}_k = c_1\lambda_1^k\boldsymbol{u}_1 + c_2\lambda_2^k\boldsymbol{u}_2 + \cdots + c_n\lambda_n^k\boldsymbol{u}_n = \Sigma_{i=1}^n c_i\lambda_i^k\boldsymbol{u}_i.$$

In the above expression, the vector $\mathbf{c} = (c_1, c_2, \dots, c_n)^\top$ is the solution of the linear system $U\mathbf{c} = \boldsymbol{w}_0$, and the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are independent eigenvectors of A with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$.

2.3 Back to the Fibonacci sequence

Now let's come back to computing F_{100}. Solving $U\mathbf{c} = \mathbf{w_0}$, we see that

$$\mathbf{c} = \begin{bmatrix}\frac{1}{\lambda_1-\lambda_2}\\ \frac{-1}{\lambda_1-\lambda_2}\end{bmatrix}.$$

Therefore,

$$\mathbf{w}_k = \underbrace{\frac{1}{\lambda_1-\lambda_2}}_{c_1}\lambda_1^k\underbrace{\begin{bmatrix}\lambda_1\\ 1\end{bmatrix}}_{\mathbf{u}_1} + \underbrace{\frac{-1}{\lambda_1-\lambda_2}}_{c_2}\lambda_2^k\underbrace{\begin{bmatrix}\lambda_2\\ 1\end{bmatrix}}_{\mathbf{u}_2} = \begin{bmatrix}\frac{\lambda_1^{k+1}-\lambda_2^{k+1}}{\lambda_1-\lambda_2}\\ \\ \frac{\lambda_1^k-\lambda_2^k}{\lambda_1-\lambda_2}\end{bmatrix}.$$

In particular, you can check that

$$\mathbf{w}_0 = \begin{bmatrix}1\\ 0\end{bmatrix} = \underbrace{\frac{1}{\lambda_1-\lambda_2}}_{c_1}\underbrace{\begin{bmatrix}\lambda_1\\ 1\end{bmatrix}}_{\mathbf{u}_1} - \underbrace{\frac{1}{\lambda_1-\lambda_2}}_{c_2}\underbrace{\begin{bmatrix}\lambda_2\\ 1\end{bmatrix}}_{\mathbf{u}_2}.$$

Since F_k is the second coordinate of $\mathbf{w}_k = \begin{bmatrix}F_{k+1}\\ F_k\end{bmatrix}$, we conclude that

$$F_k = \frac{\lambda_1^k - \lambda_2^k}{\lambda_1 - \lambda_2}.$$

What happens as k goes to infinity, i.e., what is $\lim_{k\to\infty} \mathbf{w}_k = \lim_{k\to\infty} c_1\lambda_1^k\mathbf{u}_1 + c_1\lambda_2^k\mathbf{u}_2$?
Since $\lambda_2 = -0.618 \in [-1, 0]$, $\lim_{k\to\infty} \lambda_2^k = 0$. Therefore,

$$\lim_{k\to\infty} c_1\lambda_1^k\mathbf{u}_1 + c_1\lambda_2^k\mathbf{u}_2 = \lim_{k\to\infty} c_1\lambda_1^k\mathbf{u}_1 = \left(\lim_{k\to\infty} c_1\lambda_1^k\right)\mathbf{u}_1.$$

This means that $\mathbf{w}_k$ becomes proportional to the first eigenvector $\mathbf{u}_1$ in the long run. Looking back at the formula for F_k we also see that in the long run, $F_k \sim \frac{\lambda_1^k}{\lambda_1-\lambda_2} = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^k$.

The key observation is that in this Fibonacci example, the second eigenvalue of A tended to 0 as k got large and $\mathbf{w}_k$ limits to a multiple of the first eigenvector $\mathbf{u}_1$ of A. In other words, the long term behavior of $\mathbf{w}_k$ is controlled by the eigenvector that had the largest eigenvalue (called the dominant eigenvector and eigenvalue). Keep this in mind as we go to the next chapter where we will see situations in which the dominant eigenvalue and eigenvector are guaranteed to control long term behavior of states.

2.4 Fibonacci numbers and the golden ratio – an aside

Definition 2.4.1. The **golden ratio**, denoted by the greek letter ϕ, is defined to be $\phi = \frac{1+\sqrt{5}}{2}$. It is also a root of the quadratic polynomial $x^2 - x - 1$.

The golden ratio came about from the idea of trying to draw the "perfect" rectangle, that is, a rectangle that was the most pleasing to the human eye. It appears everywhere in nature and is intimately related to the Fibonacci sequence. Recall that ϕ is the dominant eigenvalue of the matrix A in the Fibonacci example.

Suppose we were a plant, one that grows upwards and has leaves coming off of its main stem. How would we grow more and more leaves and ensure that the leaf spacing maximized sunlight on the surface of our leaves? We could start with leaf 1, and then rotate halfway around the stem to let leaf 2 grow there. That is, leaf 2 is a rotation of $\frac{1}{2}$ units around the stem from leaf 1. Next, we would want leaf 3 to be positioned so it does not block too much sunlight from leaves 1 and 2. If we drew a picture and though about it for a bit, we would end up rotating $\frac{3}{5}$ units away from leaf 2. How about leaf 4? We would rotate this one $\frac{5}{8}$ units from leaf 3. Assuming the plant was immortal, we would continue this process indefinitely, with a new spacing for each leaf.

What would happen as we proceed? If we write it out we see that the we obtain a sequence of fractions $\frac{1}{2}, \frac{3}{5}, \frac{5}{8}, \ldots, \frac{F_k}{F_{k+1}}$ and the limit of this sequence as k tends to infinity is

$$\lim_{k\to\infty} \frac{F_k}{F_{k+1}} = \frac{1}{\phi}$$

In other words, $\lim_{k\to\infty} \frac{F_{k+1}}{F_k} = \phi$, the golden ratio!

Chapter 3

Markov Matrices

In this chapter we will look at the eigenvalues and eigenvectors of positive and nonnegative matrices. Our focus will be on a subset of positive and non-negative matrices that are said to be *Markov.* We will see several applications where difference equations involving Markov matrices arise. The eigenvalues and eigenvectors of Markov matrices have special properties which play a crucial role in these applications. This material is based on Strang's Chapter 10.3.

Let's start with an example that allows us to practice what we learned about solving difference equations in the last chapter.

3.1 Longterm Behavior of Rental Cars

Suppose we want to understand the behavior of rental cars in Seattle. Cars are constantly being rented in Seattle and returned elsewhere, and cars from elsewhere are being returned in Seattle. Suppose:

- At the start 2% of all rental cars are in Seattle (i.e., fraction of cars in Seattle is 0.02 and fraction outside Seattle is 0.98).
- Every month 20% of Seattle cars leave and 5% of outside cars come in.

Question 3.1.1. What fraction of rental cars are in Seattle in the long run?

Step 1: Define the kth state vector of rental cars to be $\mathbf{w}_k$ whose first coordinate is the fraction of cars in Seattle and second coordinate is the fraction of cars outside Seattle at the end of month k. Then $\mathbf{w}_0 = \begin{bmatrix} 0.02 \\ 0.98 \end{bmatrix}$.

Step 2: Determine $\mathbf{w}_1$. Since 20% of cars leave every month and 5% come in, we know that after one month the fraction of cars in Seattle is

$$(0.8)(0.02) + (0.05)(0.98)$$

and the fraction of cars outside Seattle is

$$(0.2)(.02) + (0.95)(0.98)$$

Therefore,

$$\mathbf{w}_1 = \begin{bmatrix} (0.8)(0.02) + (0.05)(0.98) \\ (0.2)(0.02) + (0.95)(0.98) \end{bmatrix} = \underbrace{\begin{bmatrix} 0.8 & 0.05 \\ 0.2 & 0.95 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 0.02 \\ 0.98 \end{bmatrix}}_{\mathbf{w}_0}$$

Step 3: Using the same reasoning as in the Fibonacci example, we can conclude that at the end of month k, our k^{th} state vector is $\mathbf{w}_k = A^k\mathbf{w}_0$. The eigenvalues and eigenvectors of A are

$$\lambda_1 = 1, \mathbf{u}_1 = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} \qquad \lambda_2 = 0.75, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Check that

$$\mathbf{w}_0 = 1 \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} + 0.18 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Now we apply the result in Proposition 2.2.1 to find $\mathbf{w}_k$. From the expression for $\mathbf{w}_0$, we get that $c_1 = 1$ and $c_2 = 0.18$. Therefore,

$$\mathbf{w}_k = (1)(1^k) \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} + \underbrace{(0.18)(0.75)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}}_{\to 0 \text{ as } k \to \infty}$$

Once again, since the second eigenvalue 0.75 is between 0 and 1, its powers $(0.75)^k$ will tend to 0 as k goes to infinity. Further the dominant eigenvalue is 1, and $\lim_{k\to\infty} \mathbf{w}_k = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$. Therefore, in the long run, 20% of cars end up in Seattle. Once again, in the limit, $\mathbf{w}_k$ is proportional to the dominant eigenvector of A.

3.2 Positive Markov Matrices

Definition 3.2.1. Let $A = (a_{ij}) \in \mathbb{R}^{m\times n}$.

- A is a **positive** matrix, denoted $A > 0$, if $a_{ij} > 0$ for all i, j.
- A is a **non-negative matrix**, denoted $A \geq 0$ if $a_{ij} \geq 0$ for all i, j.
- A is **Markov** (also known as *stochastic*) if A is non-negative and the entries in each column of A add up to 1, i.e., $a_{ij} \geq 0 \ \forall i, j$ and $\sum_{i=1}^m a_{ij} = 1 \ \ \forall j$.
- A is a **positive Markov** matrix if it is positive and the entries in each column of A add up to 1, i.e., $a_{ij} > 0 \ \ \forall i, j$ and $\sum_{i=1}^m a_{ij} = 1 \ \ \forall j$.

Example 3.2.2. The matrix $A = \begin{bmatrix} .8 & .05 \\ .2 & .95 \end{bmatrix}$ in the rental car example is positive Markov. The matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ in the Fibonacci example is non-negative but neither positive nor Markov.

Let us first develop some facts about Markov matrices. Suppose $\mathbf{1} = (1, 1, \cdots, 1)^\top$ is the vector of all ones. Recall from the facts about matrix-vector multiplication that the condition that all columns of a matrix A adds up to 1 is the same as saying $\mathbf{1}^\top A = \mathbf{1}^\top$. Therefore, if A is Markov then $\mathbf{1}^\top A = \mathbf{1}^\top$. We argue the following about powers of positive Markov matrices.

Lemma 3.2.3. *If A is positive Markov then A^k is also positive Markov.*

Proof. Suppose A is positive Markov and $\mathbf{v}$ is a non-negative vector that is not $\mathbf{0}$. Then

1. $A\mathbf{v} > 0$ since each entry of $A\mathbf{v}$ is the dot product of a row of A which is all positive with a non-negative vector whose entries are not all zero.
2. If further, $\mathbf{1}^\top\mathbf{v} = 1$ (i.e., the entries of $\mathbf{v}$ add to 1), then $\mathbf{1}^\top A\mathbf{v} = \mathbf{1}^\top\mathbf{v} = 1$ which means that the entries of $A\mathbf{v}$ also sum to 1.

Now consider $A^2 = A\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix}$. The jth column of A^2 is $A\mathbf{a}_j$. Since A is positive Markov and $\mathbf{a}_j$ is a column of A, $\mathbf{a}_j > 0$ and its entries sum to 1. So thinking of $\mathbf{a}_j$ as the $\mathbf{v}$ in the previous paragraph we get that $A\mathbf{a}_j > 0$ and its entries sum to 1. Thus we have shown that A^2 is positive Markov. Repeating this argument we get that A^k is positive Markov.

□

Theorem 3.2.4. (**Perron's theorem for positive matrices**) *If $A > 0$ then A has a* dominant eigenvalue *λ_A with the following properties:*

1. *$\lambda_A > 0$ and it has an eigenvector $\boldsymbol{u}_A > 0$ (has all positive entries) called a* dominant eigenvector *of A.*
2. *$\mathrm{AM}(\lambda_A) = 1$.*
3. *If μ is another eigenvalue of A, then $|\mu| < \lambda_A$.*
4. *If μ is a non-dominant eigenvalue of A, then it has no eigenvector that is non-negative.*

The proof of this theorem is beyond the scope of this class, so we will just accept it. If you are interested, a proof can be found in the book *Numerical Linear Algebra* by Lloyd Trefethen and David Bau. The above theorem says more if A is a positive Markov matrix.

Theorem 3.2.5. (**Perron's theorem for positive Markov matrices**) *If A is positive Markov then A has all the properties stated in Theorem 3.2.4, and in addition:*

1. *$\lambda_A = 1$ and thus if μ is another eigenvalue of A, then $|\mu| < 1$.*
2. *If $\boldsymbol{w}_0 \geq 0$ and $\boldsymbol{w}_k = A^k\boldsymbol{w}_0$, then $\lim_{k\to\infty} A^k\boldsymbol{w}_0 = c\boldsymbol{u}_A$ where $c \geq 0$.*
   In other words, the in the long run, $\boldsymbol{w}_k$ tends to a non-negative multiple of $\boldsymbol{u}_A$.

Example 3.2.6. Recall that in the rental car problem, $\mathbf{w}_k$ tends to $\mathbf{u}_A = \begin{bmatrix}.2\\.8\end{bmatrix}$ which is the dominant eigenvector of the positive Markov matrix $A = \begin{bmatrix}.8 & .05\\.2 & .95\end{bmatrix}$ in that problem. The dominant eigenvalue was indeed 1 and the other eigenvalue 0.75 is smaller than 1 in absolute value. The constant $c = 1$.

We prove Theorem 3.2.5 using Theorem 3.2.4.

Proof. 1. Let A be a positive Markov matrix and let $\mathbf{1}$ be the vector of all ones as before. Since the column entries of A sum to 1, the row entries of $A^\top$ sum to 1 which means $A^\top\mathbf{1} = \mathbf{1}$. Hence 1 is an eigenvalue of $A^\top$ with eigenvector $\mathbf{1}$. Since $A^\top$ is a positive matrix, it can have only one positive eigenvector by part 4 of Theorem 3.2.4 and this eigenvector is the dominant eigenvector of $A^\top$. Therefore, $\mathbf{1}$ is the dominant eigenvector of $A^\top$ and 1 is the dominant eigenvalue of $A^\top$. Since A and $A^\top$ have the same eigenvalues their biggest eigenvalues are also the same. We conclude that $\lambda_A = 1$.

The second part of the statement follows from Theorem 3.2.4.

2. To prove the second statement, we assume for simplicity that A is diagonalizable. We will only need this special case. If $A = U\Lambda U^{-1}$ then by Proposition 2.2.1 we know that

$$\mathbf{w}_k = c_1(1)^k\mathbf{u}_A + c_2\lambda_2^k\mathbf{u}_2 + \cdots + c_n\lambda_n^k\mathbf{u}_n.$$

Since we know that $|\lambda_i| < 1$ for all $i = 2, 3, \ldots, n$, we get that $\lim_{k\to\infty}\lambda_i^k = 0$ for all non-dominant eigenvalues. From this we can conclude that

$$\lim_{k\to\infty}\mathbf{w}_k = \lim_{k\to\infty} c_1(1)^k\mathbf{u}_A + \underbrace{c_2\lambda_2^k\mathbf{u}_2 + \cdots + c_n\lambda_n^k\mathbf{u}_n}_{\to 0 \text{ as } k\to\infty} = c_1\mathbf{u}_A.$$

How do we see that $c_1 > 0$? From Lemma 3.2.3 we know that A^k is positive Markov. Therefore, $(A^k)^\top \mathbf{1} = \mathbf{1}$. Since $\mathbf{w}_0 \geq 0$ and $\mathbf{w}_0 \neq \mathbf{0}$,

$$(A^k \mathbf{w}_0)^\top \mathbf{1} = \mathbf{w}_0^\top (A^k)^\top \mathbf{1} = \mathbf{w}_0^\top \mathbf{1} > 0.$$

Therefore, $(\lim_{k\to\infty} A^k \mathbf{w}_0)^\top \mathbf{1} = (c_1 \mathbf{u}_A)^\top \mathbf{1} > 0$ which implies that $c_1 > 0$ since we know that $\mathbf{u}_A^\top \mathbf{1} > 0$. □

Warning: If A is only non-negative Markov and not positive Markov, Perron's theorem fails. A counterexample is $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$. Note that the absolute value of the second largest eigenvalue is NOT strictly smaller than the dominant eigenvalue. Even so, not all hope is lost.

Proposition 3.2.7. *If $A \geq 0$ but A^k is positive Markov for some $k > 1$, then 1 is still the unique dominant eigenvalue of A and $|\mu| < 1$ for all other eigenvalues μ of A.*

Proof. If μ is an eigenvalue of A, then μ^k is an eigenvalue of A^k. Since A^k is a positive Markov matrix, by Theorem 3.2.5, 1 is its dominant eigenvalue. This means that $1 = \lambda^k$ for some eigenvalue λ of A, and $|\mu^k| < 1$ for all other eigenvalues μ. Thus $\lambda_A = 1$ (*why do we know $\lambda_A \neq -1$ or some other root of 1?*), and $|\mu| < 1$ for all other eigenvalues μ. □

Here is a situation where this Proposition becomes helpful.

Example 3.2.8. Suppose we have three groups with populations p_1, p_2, p_3 respectively and after each week, each group splits in half and joins the others. This behavior can be modeled by a non-negative Markov matrix. If $\mathbf{w}_0 = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$, then after one month we have

$$\mathbf{w}_1 = A\mathbf{w}_0 = \underbrace{\begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}}_{\text{non-negative Markov}} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$$

$$\mathbf{w}_2 = A^2\mathbf{w}_0 = \underbrace{\begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}}_{\text{positive Markov}} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}.$$

Since A^2 is positive Markov the previous proposition tells us that 1 is the dominant eigenvalue of A and $|\mu| < 1$ for any other eigenvalue μ of A. Computing eigenvalues and eigenvectors we see that $\lambda_A = 1$ with $\mathbf{u}_A = (1/3, 1/3, 1/3)^\top$ and $\lambda_2 = \lambda_3 = -1/2$.

Let's compute with Julia.

```
julia> using LinearAlgebra

julia> A = [1/2 1/4 1/4; 1/4 1/2 1/4; 1/4 1/4 1/2]
317Array{Float64,2}:
 0.5   0.25  0.25
 0.25  0.5   0.25
 0.25  0.25  0.5
```

```
julia> eigvals(A)
3-element Array{Float64,1}:
 0.24999999999999975
 0.25
 0.9999999999999998

julia> eigvecs(A)
317Array{Float64,2}:
 -0.0698557   0.813503  -0.57735
 -0.669586   -0.467248  -0.57735
  0.739442   -0.346255  -0.57735
```

Since Julia runs a numerical algorithm it does not often output integer answers, even if the true answer is an integer. The output 0.9999999999999998 should be interpreted as 1. The eigenvector of 1 that Julia gives us is the last column of the eigenvector output. Note that we can scale the eigenvector and have an eigenvector of 1. So if we flip the sign we get the positive eigenvector $(0.57735, 0.57735, 0.57735)^\top$. We can scale it further to get $(1/3, 1/3, 1/3)^\top$.

By Perron's theorem, $\mathbf{w}_k$ will approach a positive multiple of $\mathbf{u}_A$ in the long run. Since all coordinates of $\mathbf{u}_A$ are equal, we should expect that in the long run the three populations will equalize. Stop for a moment and think about whether this is what you would have expected from just the statement of the problem before you did any computation at all.

To see a computational example, suppose $\mathbf{w}_0 = (8, 16, 32)^\top$. Then you can compute that

$$\mathbf{w}_0 = \begin{bmatrix} 8 \\ 16 \\ 32 \end{bmatrix}, \mathbf{w}_1 = \begin{bmatrix} 24 \\ 20 \\ 12 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 16 \\ 18 \\ 22 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 20 \\ 19 \\ 17 \end{bmatrix} \cdots \longrightarrow 56 \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}.$$

The number 56 is the sum of the populations at the start which remains constant as you advance k.

3.3 Non-negative Markov Matrices

We already saw that Perron's theorem fails for non-negative matrices. However, the eigenvalues and eigenvectors of non-negative and non-negative Markov matrices still have some special properties as we see in the following theorem.

Theorem 3.3.1. *(***Frobenius Theorem***) If $A \geq 0$ then A has a dominant eigenvalue λ_A such that*

1. *$\lambda_A \geq 0$ with eigenvector $\boldsymbol{u}_A \geq 0$.*

2. *If μ is another eigenvalue of A then $|\mu| \leq \lambda_A$.*

Example 3.3.2. The matrix $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ has a single eigenvalue 0 with multiplicity 2. Therefore, $\lambda_A = 0$ and $\mathrm{AM}(\lambda_A) = 2$. Also, $E_0 = \{\mathbf{x} \in \mathbb{R}^2 : x_2 = 0\}$ which is the x_1 axis of $\mathbb{R}^2$. There are certainly nonnegative vectors in E_0 such as $(1, 0)$. This example shows you that if A is only non-negative, then the dominant eigenvalue is also only non-negative and it can be repeated. Further, we can only guarantee that the dominant eigenvector is non-negative, not positive.

As with Perron's theorem, we can say something more if A is non-negative and Markov. Besides the properties already guaranteed by Theorem 3.3.1, we get the following.

Theorem 3.3.3. *If A is non-negative and Markov, then $\lambda_A = 1$.*

Example 3.3.4. Consider the $n \times n$ identity matrix I_n which has only one eigenvalue $\lambda = 1$ with multiplicity n. The eigenspace $E_1 = \mathbb{R}^n$ so 1 has a non-negative eigenvector. In fact, it has a positive eigenvector.

We write the proof of part of this theorem in case you are interested to see it. This is optional and uses (slightly) advanced facts, but is within your reach. We need a theorem that we will not prove.

Theorem 3.3.5. *(**Gershgorin's Theorem**) For any $A \in \mathbb{R}^{n \times n}$ and any eigenvalue λ of A, there is some i such that $|a_{ii} - \lambda| \leq \sum_{j \neq i} |a_{ij}|$.*

What this theorem says is that every eigenvalue of A is close to at least one of the diagonal entries a_{ii} of A. More precisely, if you plot all the diagonal entries of A on the number line $\mathbb{R}$ and for each a_{ii} mark the interval centered at a_{ii} and distance $\sum_{j \neq i} |a_{ij}|$ on either side of a_{ii}, then λ will lie in at least one these intervals. Note that $\sum_{j \neq i} |a_{ij}|$ is the sum of the absolute values of all the entries in row i except a_{ii}.

We can now argue why Theorem 3.3.3 is true. Suppose we denote the (i, j) entry of $A^\top$ by a'_{ij}. Since A is non-negative Markov, the rows of $A^\top$ sum to 1 and that all entries of $A^\top$ are non-negative. Therefore,

$$\sum_{j \neq i} |a'_{ij}| = \sum_{j \neq i} a'_{ij} = 1 - a'_{ii} = 1 - a_{ii}.$$

Suppose μ is an eigenvalue of $A^\top$ then

$$|\mu| - a'_{ii} = |\mu| - |a'_{ii}| \leq |\mu - a'_{ii}| \leq \sum_{j \neq i} |a'_{ij}| = 1 - a_{ii}$$

We have used lots of different facts in the above chain: the first equality is because $a'_{ii} \geq 0$ and hence $a'_{ii} = |a'_{ii}|$. The inequality that comes next follows from the triangle inequality for distances in $\mathbb{R}$. The next inequality is from Geshgorin's theorem and the last equality is what we showed two lines before. Altogether we have shown that

$$|\mu| - a'_{ii} \leq 1 - a_{ii}$$

but since $a'_{ii} = a_{ii}$ we conclude that $|\mu| \leq 1$.

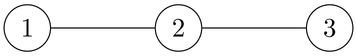
Now we use the fact that all eigenvalues of A and $A^\top$ are the same, so μ is also an eigenvalue of A. Also recall that if A is Markov, then 1 is an eigenvalue of A since 1 is an eigenvalue of $A^\top$ because $A^\top \mathbf{1} = \mathbf{1}$. So we conclude that $\lambda_A = 1$ and $|\mu| \leq 1$ for all other eigenvalues μ of A.

3.4 Adjacency Matrices of Graphs

We now give a family of non-negative matrices that will come up over and over again in this class.

A **graph** G is a collection of **nodes** or **vertices**, labeled $1, 2, \ldots, n$ and **edges** which connect pairs of nodes. We let $\{i, j\}$ denote the edge connecting nodes i and j.

Here is an example of a graph.



This graph has three nodes: $1, 2, 3$ and two edges $\{1, 2\}, \{2, 3\}$.

Graphs are denoted as $G = (V, E)$ where V is the set of nodes in G and E is the set of edges in G. Our example graph above would be written as $G = (\underbrace{\{1, 2, 3\}}_{V}, \underbrace{\{\{1, 2\}, \{2, 3\}\}}_{E})$.

To any graph G we can associate a matrix A_G, known as the *adjacency matrix* of G.

Definition 3.4.1. The **adjacency matrix** A_G of a graph $G = (V, E)$ with n nodes is a $n \times n$ matrix defined as follows. The rows and columns of A are indexed by the node labels $1, \dots, n$ and the (i, j)-entry of A is 1 if the pair $\{i, j\}$ is an edge in G and 0 otherwise. That is $A_G = (a_{ij})$ where

$$a_{ij} = \begin{cases} 1 & \text{if there exists an edge from } i \text{ to } j, \text{i.e., } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

The adjacency matrix for the graph G from above would be

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

Notice that this matrix is non-negative (but not Markov). Computing its eigenvalues and eigenvectors:

$$\lambda_1 = \sqrt{2} \quad \mathbf{u}_1 = \begin{bmatrix} \sqrt{2} \\ 2 \\ \sqrt{2} \end{bmatrix}, \qquad \lambda_2 = 0 \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \qquad \lambda_3 = -\sqrt{2} \quad \mathbf{u}_3 = \begin{bmatrix} \sqrt{2} \\ -2 \\ \sqrt{2} \end{bmatrix}$$

Note several things:

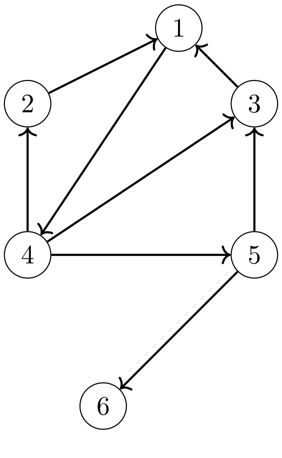
- The dominant eigenvalue $\lambda_{A_G} = \sqrt{2}$ is positive but not 1.
- A_G has a smaller eigenvalue $\mu = -\sqrt{2}$ such that $|\mu| = \lambda_{A_G}$.
- The dominant eigenvalue has a non-negative (in fact, positive) eigenvector.

Over the coming weeks we will be looking at lots of graphs and their adjacency matrices. We finish this section with a powerful application of what we have learned in this chapter.

3.5 Google Page Rank

The Google page rank algorithm is due to Larry Page and Sergey Brin in 1998, the founders of Google, and is used to rank webpages. The central idea is to decide the "rank" or "importance" of a webpage by the importance of pages that link to it. When we submit a query, we want to see the most relevant pages at the top of the list, so a useful measurement of importance is needed. We will see how positive Markov matrices and difference equations give us an algorithm. This material is taken from Chapter 12 of the book *Math Bytes* by Tim Chartier.

We can think of the world wide web as a *directed* graph, with nodes indicating webpages and edges indicating links from one page to another. Suppose our world wide web consists of 6 webpages and links as follows.



From this graph we construct a *page rank matrix*, which will be a 6×6 matrix of probabilities, that models movement between pages. We compute it according to the following rules:
When you are at page i, you have two choices

1. Teleportation: jump to a different page by typing the url of the webpage directly into the browser.
2. Follow a link.

If you decide to teleport, assume that all pages are equally like for you to teleport to. If you decide to follow a link, assume that all links are equally likely. You may want to assign different probabilities to these moves but let's keep it simple. Let's also decide in this example when we will teleport and when we will follow a link. Say we roll a dice and do the following:

- If we roll $1, 2, 3, 4$, or 5, then we follow a link.
- If we roll a 6, then we teleport.

The (i, j) entry of the page rank matrix A is the probability of going **from** page j **to** page i. Let's compute the first column of this matrix in our example. The entries are a_{i1} for $i = 1, 2, 3, 4, 5, 6$. Remember, to compute a_{i1} we only focus on going **from** page 1 **to** page i.

- a_{11}: We can only teleport from 1 to 1. We have a 1/6 chance of rolling a 6 and a 1/6 chance of teleporting to 1 out of the 6 possible webpages. The total probability $a_{11} = (1/6)(1/6) = 1/36$.
- a_{21}: Similarly to the previous case, we have a 1/6 chance of teleporting from 1 to 2 and since all webpages are equally likely, the probability that we teleport from 1 to 2 is $(1/6)(1/6) = 1/36$. We can also follow a link with probability 5/6 (i.e. rolling anything but a 6), there is no link from 1 to 2 in our graph. So in total, we move from 1 to 2 with probability $a_{21} = (1/6)(1/6) + (5/6)(0) = 1/36$.
- a_{31}: This is the same as the above, i.e., $a_{31} = 1/36$.
- a_{41}: This is the only computation that is different because there is a link from 1 to 4. As before, we teleport with probability 1/6 (chance of rolling a 6), so the teleportation probability from 1 to 4 is 1/36. We will follow a link with probability 5/6 (chance of rolling anything other than a 6), and a 1/1 chance of linking from 1 to 4 since 4 is the only link we can follow from 1. Note that if 1 had k outgoing edges, then the chance of linking to any of those pages would then be $1/k$. Therefore the total probability of going from 1 to 4 is $a_{41} = (1/6)(1/6) + (5/6)(1) = 31/36$.
- a_{51}: This is the same as a_{21}.
- a_{61}: This is the same as a_{21}.

Summarizing, the first column of our page rank matrix is $\begin{bmatrix} 1/36 \\ 1/36 \\ 1/36 \\ 31/36 \\ 1/36 \\ 1/36 \end{bmatrix}$. Computing all entries by the same procedure as above, we get the page rank matrix:

$$A = \begin{bmatrix} 1/36 & 31/36 & 31/36 & 1/36 & 1/36 & 6/36 \\ 1/36 & 1/36 & 1/36 & 11/36 & 1/36 & 6/36 \\ 1/36 & 1/36 & 1/36 & 11/36 & 16/36 & 6/36 \\ 31/36 & 1/36 & 1/36 & 1/36 & 1/36 & 6/36 \\ 1/36 & 1/36 & 1/36 & 11/36 & 1/36 & 6/36 \\ 1/36 & 1/36 & 1/36 & 1/36 & 16/36 & 6/36 \end{bmatrix}.$$

The computation of the last column is a bit special since page 6 has no links at all. Therefore, from page 6 you can only teleport and so with probability 1 you will teleport from page 6 and each page has an equal probability of being teleported to. Therefore, $a_{i6} = 1 \cdot \frac{1}{6} = \frac{1}{6}$ for all i. You should compute several entries by hand to make sure you understand the procedure. We make several key observations about this matrix and see how the theorems we saw in this chapter help.

Important Observations

- Teleportation **guarantees** that our matrix A will be positive since every entry will be at least 1/36.
- Since the entries in column j are the probabilities of going from page j to the others, the entries must add up to 1, so A will be Markov. Check that all columns of A add up to 1. This also provides shortcuts to the computation.

Since the page rank matrix A is positive Markov, Perron's theorem applies. Using Julia we can compute its eigenvalues and eigenvectors.

```
julia> using LinearAlgebra

julia> A = [1/36 31/36 31/36 1/36 1/36 6/36;
       1/36 1/36 1/36 11/36 1/36 6/36;
       1/36 1/36 1/36 11/36 16/36 6/36;
       31/36 1/36 1/36 1/36 1/36 6/36;
       1/36 1/36 1/36 11/36 1/36 6/36;
       1/36 1/36 1/36 1/36 16/36 6/36]
617Array{Float64,2}:
 0.0277778  0.861111   0.861111   0.0277778  0.0277778  0.166667
 0.0277778  0.0277778  0.0277778  0.305556   0.0277778  0.166667
 0.0277778  0.0277778  0.0277778  0.305556   0.444444   0.166667
 0.861111   0.0277778  0.0277778  0.0277778  0.0277778  0.166667
 0.0277778  0.0277778  0.0277778  0.305556   0.0277778  0.166667
 0.0277778  0.0277778  0.0277778  0.0277778  0.444444   0.166667

julia> eigvals(A)
6-element Array{Complex{Float64},1}:
    -0.2977044274773357 - 0.6329450281890806im
    -0.2977044274773357 + 0.6329450281890806im
   -0.22223919020536226 + 0.0im
 -5.782547471497545e-18 + 0.0im
    0.12320360071558936 + 0.0im
     1.0000000000000013 + 0.0im

julia> eigvecs(A)
617Array{Complex{Float64},2}:
  0.233465+0.505988im   0.233465-0.505988im   0.122729+0.0im  ?    0.20872+0.0im  -0.599066+0.0im
 0.0893727-0.243577im  0.0893727+0.243577im    0.48738+0.0im     0.0323265+0.0im  -0.253028+0.0im
  0.198012-0.133645im   0.198012+0.133645im  -0.426388+0.0im      0.141653+0.0im  -0.358456+0.0im
 -0.691375-0.0im       -0.691375+0.0im       -0.108766+0.0im      0.443701+0.0im  -0.588717+0.0im
 0.0893727-0.243577im  0.0893727+0.243577im    0.48738+0.0im     0.0323265+0.0im  -0.253028+0.0im
 0.0811519+0.114811im  0.0811519-0.114811im  -0.562336+0.0im  ?  -0.858726+0.0im  -0.194924+0.0im
```

The dominant eigenvalue of A is listed at the end in the list of eigenvalues and its eigenvector is listed last in the list of eigenvectors. Observe that the dominant eigenvalue is 1 with eigenvector

$$\begin{bmatrix}-0.599066\\-0.253028\\-0.358456\\-0.588717\\-0.253028\\-0.194924\end{bmatrix}.$$

This is not positive, but since all scalings of an eigenvector is again an eigenvector we can scale by -1 and take the dominant eigenvector to be

$$\begin{bmatrix}0.599066\\0.253028\\0.358456\\0.588717\\0.253028\\0.194924\end{bmatrix}$$

which is positive. Let's now scale once more by dividing by the sum of the coordinates to get a vector of probabilities, and set

$$\mathbf{u}_A = \begin{bmatrix}0.26658\\0.11259\\0.15951\\0.26197\\0.11259\\0.08674\end{bmatrix}.$$

Definition 3.5.1. We call $\mathbf{u}_A$ the **page rank vector** of our system of webpages. It is the dominant eigenvector of the page rank matrix and is always scaled to be a vector of probabilities that add up to 1.

Now here is how the page rank algorithm works. We assume all pages are equally important at the beginning, so our initial state vector, whose entries rank the importance of each webpage at time 0, is $\mathbf{w}_0 = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)^\top$. Let's think about what $A\mathbf{w}_0$ means. It's first entry is the dot product of the first row of A with $\mathbf{w}_0$. The first row of A records the probabilities with which all the different pages transfer to page 1. Therefore, the first entry of $A\mathbf{w}_0$ is the importance page 1 has after one iteration of teleporting and linking from all vertices. Applying the same logic to all entries of $A\mathbf{w}_0$ we see that $\mathbf{w}_1 = A\mathbf{w}_0$ is the vector of rankings of pages after one iteration. Continuing we see that $\mathbf{w}_k = A^k\mathbf{w}_0$ is the vector of rankings after k iterations. Therefore, by Theorem 3.2.5, we have $\lim_{k\to\infty}\mathbf{w}_k = c_1(1^k)\mathbf{u}_A$. That is, the importance of webpages, in the long run, is a positive multiple of the dominant eigenvector $\mathbf{u}_A$ of A, and in particular, proportional to $\mathbf{u}_A$. Looking at $\mathbf{u}_A$ in our example, we see that the most important webpage in our system is webpage 1, the second most important is webpage 4 and so on.

Chapter 4

Orthogonality and Projections

In this chapter we will see the notion of *orthogonality* of vectors and subspaces. This material is based on Chapter 4 in Strang's book. Geometrically, orthogonality is the same as *perpendicularity*, at least in $\mathbb{R}^n$. This important concept will play a fundamental role in much of the theory and applications we will see in the rest of this class.

4.1 Orthogonal Vectors and Subspaces

Let $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{R}^n$. Recall that the **dot product** of $\mathbf{v}$ and $\mathbf{w}$ is

$$\mathbf{v}^\top \mathbf{w} = v_1 w_1 + \cdots + v_n w_n = \sum_{i=1}^{n} v_i w_i.$$

The **norm** of $\mathbf{v}$ is

$$||\mathbf{v}|| = \sqrt{\mathbf{v}^\top \mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2} = \sqrt{\sum_{i=1}^{n} v_i^2}$$

and the **squared norm** of $\mathbf{v}$ is

$$||\mathbf{v}||^2 = \mathbf{v}^\top \mathbf{v}.$$

Check that if $\mathbf{v} \in \mathbb{R}^n$, then its norm $\|\mathbf{v}\|$ is the length of the vector $\mathbf{v}$. You may also remember that if the angle between $\mathbf{v}$ and $\mathbf{w}$ is θ, then

$$\mathbf{v}^\top \mathbf{w} = \mathbf{w}^\top \mathbf{v} = \|\mathbf{v}\| \|w\| \cos\theta.$$

Therefore, if $\mathbf{v}$ and $\mathbf{w}$ are perpendicular, then $\theta = \pi/2$ and $\mathbf{v}^\top \mathbf{w} = 0$ since $\cos \pi/2 = 0$.

Definition 4.1.1. For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we say that $\mathbf{v}$ and $\mathbf{w}$ are **orthogonal** if $\mathbf{v}^\top \mathbf{w} = 0$ or equivalently, if $\mathbf{w}^\top \mathbf{v} = 0$. We write $\mathbf{v} \perp \mathbf{w}$.

If $\mathbf{v} \perp \mathbf{w}$, then draw a picture and check that the Pythagorus theorem implies the following.

Theorem 4.1.2. *If $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ and $\boldsymbol{v}^\top \boldsymbol{w} = 0$, then*

$$||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2 = ||\boldsymbol{v} - \boldsymbol{w}||^2 = ||\boldsymbol{v} + \boldsymbol{w}||^2$$

We can extend the notion of orthogonality of vectors in $\mathbb{R}^n$, to subspaces of $\mathbb{R}^n$.

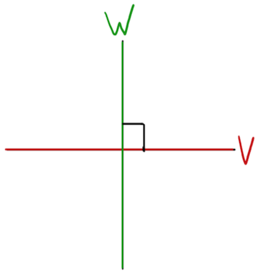
Definition 4.1.3. Let V and W be subspaces of $\mathbb{R}^n$. We say that V and W are **orthogonal**, denoted $V \perp W$, if every $\mathbf{v} \in V$ is orthogonal to every $\mathbf{w} \in W$.

To prove that two subspaces V and W are orthogonal, you need to pick an *arbitrary* vector $\mathbf{v} \in V$ and an *arbitrary* vector $\mathbf{w} \in W$ and argue that $\mathbf{v}^\top \mathbf{w} = 0$. Here *arbitrary* means that you assume for $\mathbf{v}$ only properties that all vectors in $\mathbf{v}$ would have, nothing special. Same for $\mathbf{w}$. Then since you assumed nothing special for $\mathbf{v}$ and $\mathbf{w}$ if $\mathbf{v}^\top \mathbf{w} = 0$ then you can conclude that this is true for any $\mathbf{v} \in V$ and $\mathbf{w} \in W$. We will see this technique in the examples below.

We cannot visualize subspaces in high dimension, but drawing some examples in $\mathbb{R}^2$ and $\mathbb{R}^3$ will be helpful to understand what is happening, and gives you something to hold on to.

Examples you can see

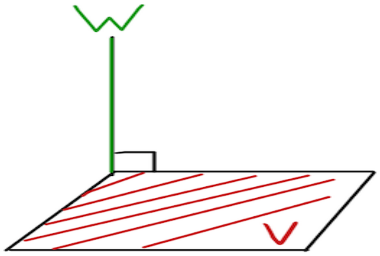
Example 4.1.4. Let $V = \text{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}\right\}$ and $W = \text{Span}\left\{\begin{bmatrix}0\\1\end{bmatrix}\right\}$. We see that these are orthogonal subspaces



because any (*arbitrary*) vector in V looks like $\begin{bmatrix}c\\0\end{bmatrix}$ for some $c \in \mathbb{R}$ and any (*arbitrary*) vector in W looks like $\begin{bmatrix}0\\d\end{bmatrix}$ for some $d \in \mathbb{R}$, and their dot product is $c \cdot 0 + 0 \cdot d = 0$. Since we made the argument using *arbitrary* vectors in V and W, the result holds for any two vectors in V and W, and we can conclude that V and W are orthogonal. This is also clear from the picture – check that any $\mathbf{v} \in V$ is perpendicular to any $\mathbf{w} \in W$.

In the previous example we had orthogonal subspaces of equal dimension but this is not always the case.

Example 4.1.5. Consider the subspaces $V = \text{Span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$ and $W = \text{Span}\left\{\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$ in $\mathbb{R}^3$.



Again $V \perp W$ because any vector in V looks like $\mathbf{v} = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$ for some $a, b \in \mathbb{R}$ and any vector in W looks like $\mathbf{w} = \begin{bmatrix} 0 \\ 0 \\ c \end{bmatrix}$ for some $c \in \mathbb{R}$ and $\mathbf{v}^\top \mathbf{w} = 0$.

An example you can't see

Example 4.1.6. Let $V = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ and $W = \operatorname{Span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$. Use similar arguments to the previous examples to check that $V \perp W$.

We will now see that some subspaces that we know well are in fact orthogonal. A matrix $A \in \mathbb{R}^{m \times n}$ has three natural subspaces associated to it:

- **Column space of** A: $\operatorname{Col}(A) = \{A\mathbf{x} \,:\, \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$, the span of the columns of A.
- **Row space of** A: $\operatorname{Row}(A) = \{\mathbf{y}^\top A \,:\, \mathbf{y} \in \mathbb{R}^m\} \subseteq \mathbb{R}^n$, the span of the rows of A.
- **Null space of** A: $\operatorname{Null}(A) = \{\mathbf{x} \in \mathbb{R}^n \,:\, A\mathbf{x} = \mathbf{0}\} \subseteq \mathbb{R}^n$.

Here is some convention to remember. All vectors are assumed to be column vectors. So if we write $\mathbf{u} \in \mathbb{R}^n$ we mean a column vector. If we write $\mathbf{u}^\top$ then we mean a row vector. To save space in these notes we write $(1, 1, 1)^\top$ to mean the *column vector* whose entries are all ones. This makes sense because $(1, 1, 1)$ is a row vector and $(1, 1, 1)^\top$ is its transpose which is a column vector. Technically column vectors and row vectors live in *dual vector spaces* but we won't really go there in this class because the dual vector space of $\mathbb{R}^n$ is isomorphic to $\mathbb{R}^n$ again and we allow row vectors and column vectors to live together in $\mathbb{R}^n$. Therefore, in the above, $\operatorname{Row}(A)$ is considered to be a subspace of $\mathbb{R}^n$ although it is filled with row vectors. Whereas $\operatorname{Null}(A)$ and $\operatorname{Col}(A)$ are filled with column vectors. Note that $\operatorname{Null}(A)$ and $\operatorname{Col}(A)$ live in different spaces.

Proposition 4.1.7. *For any matrix $A \in \mathbb{R}^{m \times n}$,* $\operatorname{Row}(A) \perp \operatorname{Null}(A)$.

Proof. We begin by writing A in terms of its rows.

$$A = \begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix}$$

Pick an $\mathbf{x} \in \operatorname{Null}(A)$ and a row $\mathbf{b}_i^\top$ of A. We first show that $\mathbf{b}_i^\top \mathbf{x} = 0$ for all i. (Note that since $\mathbf{b}_i^\top$ is already a row vector the dot product is just $\mathbf{b}_i^\top \mathbf{x}$.) Recall that the entries of $A\mathbf{x}$ are dot products of the rows of A with $\mathbf{x}$, so we have that

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = A\mathbf{x} = \begin{bmatrix} \mathbf{b}_1^\top \mathbf{x} \\ \vdots \\ \mathbf{b}_m^\top \mathbf{x} \end{bmatrix}.$$

Since two vectors are equal if and only if their entries are equal, we get that $\mathbf{b}_i^\top \mathbf{x} = 0$ for all i.

Now recall that $\operatorname{Row}(A) = \operatorname{Span}\{\mathbf{b}_1^\top, \cdots, \mathbf{b}_m^\top\} \subseteq \mathbb{R}^n$. Therefore, any vector in $\operatorname{Row}(A)$ is of the form $c_1 \mathbf{b}_1^\top + \cdots + c_m \mathbf{b}_m^\top$ for some $c_1, \ldots, c_m \in \mathbb{R}$, and we have that

$$(c_1 \mathbf{b}_1^\top + \cdots + c_m \mathbf{b}_m^\top)\mathbf{x} = c_1 \mathbf{b}_1^\top \mathbf{x} + \cdots + c_m \mathbf{b}_m^\top \mathbf{x} = 0.$$

Since we have shown that the dot product of an arbitrary vector in $\operatorname{Row}(A)$ and an arbitrary vector in $\operatorname{Null}(A)$ is 0, we can conclude that $\operatorname{Row}(A)$ is orthogonal to $\operatorname{Null}(A)$. □

We could have made this argument short by using matrix notation. Since $\mathrm{Row}(A) = \{\mathbf{y}^\top A : \mathbf{y} \in \mathbb{R}^m\}$ and $\mathrm{Null}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$, if we took and element $\mathbf{y}^\top A \in \mathrm{Row}(A)$ and an element $\mathbf{x} \in \mathrm{Null}(A)$ then $(\mathbf{y}^\top A)\mathbf{x} = \mathbf{y}^\top (A\mathbf{x}) = \mathbf{y}^\top \mathbf{0} = 0$. Hence, $\mathrm{Row}(A) \perp \mathrm{Null}(A)$.

As we saw in the proof of Proposition 4.1.7, to show the orthogonality of V and W it is enough to show that any basis vector of V is orthogonal to any basis vector of W. Here is the reason, more formally.

Proposition 4.1.8. *Let V and W be subspaces in $\mathbb{R}^n$ with bases $\mathcal{B}_V = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ and $\mathcal{B}_W = \{\boldsymbol{w}_1, \ldots, w_l\}$. Then $V \perp W$ if and only if $\boldsymbol{v}_i^\top \boldsymbol{w}_j = 0$ for all $i = 1, \ldots, k$ and $j = 1, \ldots, l$.*

Proof. If $V \perp W$, then any vector $\mathbf{v} \in V$ is orthogonal to any vector $\mathbf{w} \in W$. In particular, $\mathbf{v}_i^\top \mathbf{w}_j = 0$ for all $i = 1, \ldots, k$ and $j = 1, \ldots, l$.

We now argue the converse. By the definition of bases, we know that any $\mathbf{v} \in V$ and $\mathbf{w} \in W$ look like

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_k, \quad \mathbf{w} = \beta_1 \mathbf{w}_1 + \cdots + \beta_l \mathbf{w}_l$$

for scalars $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_l \in \mathbb{R}$. If $\mathbf{v}_i^\top \mathbf{w}_j = 0$ for all i and j, then

$$\mathbf{v}^\top \mathbf{w} = (\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_k)^\top (\beta_1 \mathbf{w}_1 + \cdots + \beta_l \mathbf{w}_l) = \alpha_1 \beta_1 \mathbf{v}_1^\top \mathbf{w}_1 + \alpha_1 \beta_2 \mathbf{v}_1^\top \mathbf{w}_2 + \cdots + \alpha_k \beta_l \mathbf{v}_k^\top \mathbf{w}_l = 0$$

This means that any vector in V is orthogonal to any vector in W and $V \perp W$. □

Example 4.1.9. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \end{bmatrix}$. We have $\mathrm{Row}(A) = \mathrm{Span}\{(1,2,3), (-1,0,2)\}$. Before computing the null space of A, we know by rank-nullity that it must be one-dimensional. Carrying out the usual computation we get that $\mathrm{Null}(A) = \mathrm{Span}\left\{\begin{bmatrix} 2 \\ -5/2 \\ 1 \end{bmatrix}\right\}$. Computing the dot products we have that

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ -5/2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -5/2 \\ 1 \end{bmatrix} = 0.$$

Since it's enough to check orthogonality on bases, we conclude that $\mathrm{Row}(A) \perp \mathrm{Null}(A)$.

Now you may be wondering if there is a nice analog of this Proposition 4.1.7 for $\mathrm{Col}(A)$, and sure enough there is! The key to seeing this fact is to notice that $\mathrm{Col}(A) = \mathrm{Row}(A^\top)$ for any matrix A. The following is then an immediate corollary of Proposition 4.1.7.

Proposition 4.1.10. $\mathrm{Col}(A) \perp \mathrm{Null}(A^\top)$ *for any matrix $A \in \mathbb{R}^{m \times n}$.*

In Strang's book he calls the four subspaces $\mathrm{Row}(A), \mathrm{Col}(A), \mathrm{Null}(A), \mathrm{Null}(A^\top)$ associated to a matrix A, the *four fundamental subspaces* of A. The space $\mathrm{Null}(A^\top)$ is often called the **left nullspace of** A since

$$\mathrm{Null}(A^\top) = \{\mathbf{y}^\top : \mathbf{y}^\top A = 0\}.$$

We think of $\mathrm{Null}(A^\top) \subseteq \mathbb{R}^m$ as filled with row vectors similar to $\mathrm{Row}(A)$. We have seen that these subspaces are paired up by orthogonality:

$$\mathrm{Row}(A) \perp \mathrm{Null}(A), \quad \mathrm{Null}(A^\top) \perp \mathrm{Col}(A).$$

Before we end this section, let's note a very useful fact about matrix-matrix multiplication.

Rows and Columns of a Product of Matrices

Let $A \in \mathbb{R}^{m\times k}$ with rows $\mathbf{a}_1^\top, \dots, \mathbf{a}_m^\top$ and $B \in \mathbb{R}^{k\times n}$ with columns $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ where $\mathbf{a}_i \in \mathbb{R}^k$ and $\mathbf{b}_j \in \mathbb{R}^k$.

- The columns of AB are $A\mathbf{b}_1, \dots, A\mathbf{b}_n$:

$$AB = A\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n\end{bmatrix} = \begin{bmatrix}A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_n\end{bmatrix}$$

  Therefore, each column of AB is a linear combination of the columns of A.

- The rows of AB are $\mathbf{a}_1^\top B, \mathbf{a}_2^\top B, \dots, \mathbf{a}_m^\top B$:

$$AB = \begin{bmatrix}\mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_m^\top\end{bmatrix} B = \begin{bmatrix}\mathbf{a}_1^\top B \\ \mathbf{a}_2^\top B \\ \vdots \\ \mathbf{a}_m^\top B\end{bmatrix}.$$

  Therefore, each row of AB is a linear combination of the rows of B.

4.2 Orthogonal complements

In Example 4.1.4 we saw two orthogonal subspaces of the same dimension, while in Example 4.1.5 we had orthogonal subspaces of different dimensions. However, in both cases, the sum of their dimensions was equal to the dimension of the ambient space, $\mathbb{R}^2$ and $\mathbb{R}^3$ respectively. If this happens we will call the subspaces *complementary*. These examples were a nice coincidence but we could have also used the following example.

Example 4.2.1. Let $V = \text{Span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}\right\}$ and $W = \text{Span}\left\{\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$. Check that $V \perp W$, and $\dim(V) + \dim(W) = 2 < 3$.

In fact, any pair of axes in $\mathbb{R}^n$ are orthogonal subspaces. These provide examples of orthogonal subspaces that are not *complementary*. We will see that complementary orthogonal subspaces allow us to decompose our entire space $\mathbb{R}^n$ in a unique way.

Definition 4.2.2. Let V be a subspace of $\mathbb{R}^n$. The **orthogonal complement** of V, denoted $V^\perp$ is the subspace of all vectors orthogonal to V. That is

$$V^\perp = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{x}^\top\mathbf{y} = 0 \;\; \forall \mathbf{x} \in V\}$$

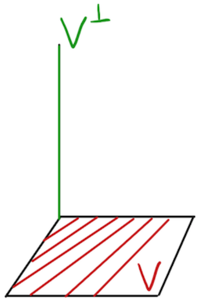
You should pause and check that the set of all vectors orthogonal to a subspace $V \subseteq \mathbb{R}^n$ is again a subspace of $\mathbb{R}^n$. We will now derive a series of facts about orthogonal complements.

Proposition 4.2.3. *Let $V \subseteq \mathbb{R}^n$ be a subspace. Then $V \cap V^\perp = \{\mathbf{0}\}$.*

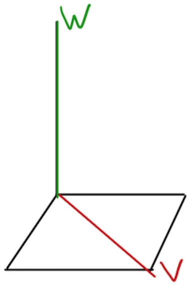
Proof. If $\mathbf{x} \in V \cap V^\perp$ and $\mathbf{x} \neq \mathbf{0}$ then $\mathbf{x}^\top\mathbf{x} = 0$ which implies that $||\mathbf{x}|| = 0$ and hence $\mathbf{x} = \mathbf{0}$. □

The four fundamental subspaces of a matrix provide examples of orthogonal complements.

Theorem 4.2.4. *For any matrix $A \in \mathbb{R}^{m\times n}$,* $\text{Row}(A)^\perp = \text{Null}(A)$ *and* $\text{Col}(A)^\perp = \text{Null}(A^\top)$.



Orthogonal Complements



$V \perp W$ but $W \neq V^\perp$

Proof. We'll show the first statement. The second one is similar. We saw that $\text{Row}(A) \perp \text{Null}(A)$ which means that all vectors in $\text{Null}(A)$ are orthogonal to the vectors in $\text{Row}(A)$. Therefore $\text{Null}(A) \subseteq \text{Row}(A)^\perp$ since $\text{Row}(A)^\perp$ consists of all vectors that are orthogonal to the vectors in $\text{Row}(A)$. We need to argue that there are no more vectors in $\text{Row}(A)^\perp$ beyond the ones in $\text{Null}(A)$. Let's argue by contradiction. Suppose there was a vector $\mathbf{z} \notin \text{Null}(A)$ (which means that $A\mathbf{z} \neq \mathbf{0}$) but $\mathbf{z} \in \text{Row}(A)^\perp$. Then $\mathbf{b}_i^\top \mathbf{z} = 0$ for all rows $\mathbf{b}_i^\top$ of A which means that

$$A\mathbf{z} = \begin{bmatrix} \mathbf{b}_1^\top \mathbf{z} \\ \vdots \\ \mathbf{b}_m^\top \mathbf{z} \end{bmatrix} = \mathbf{0}$$

which contradicts that $A\mathbf{z} \neq \mathbf{0}$. □

Check that in Example 4.1.9, $\text{Row}(A)$ and $\text{Null}(A)$ are complementary orthogonal subspaces. We will now argue that ALL pairs of complementary subspaces $(V, V^\perp)$ are of the form $V = \text{Row}(A)$ and $V^\perp = \text{Null}(A)$ for some matrix A.

Theorem 4.2.5. *Any subspace $V \subseteq \mathbb{R}^n$ is the row space of a matrix $A \in \mathbb{R}^{m\times n}$ and hence $V^\perp = \text{Null}(A)$. In particular, any pair of subspaces $(V, V^\perp)$ is of the form $(\text{Row}(A), \text{Null}(A))$ for some matrix A.*

Proof. Suppose $V = \text{Span}\{\mathbf{b}_1^\top, \ldots, \mathbf{b}_k^\top\}$. Then setting $A = \begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_k^\top \end{bmatrix}$ we get that $V = \text{Row}(A)$. From Theorem 4.2.4 we get that $\text{Null}(A) = \text{Row}(A)^\perp = V^\perp$. □

We'll now see various reasons that justify the word *complementary* for a subspace $V \subseteq \mathbb{R}^n$ and its orthogonal complement $V^\perp$.

Proposition 4.2.6. *If $V \subseteq \mathbb{R}^n$ is a subspace, then* $\dim(V) + \dim(V^\perp) = n$.

Proof. We just argued that for any subspace $V \subseteq \mathbb{R}^n$, there is a matrix $A \in \mathbb{R}^{m\times n}$ such that $V = \text{Row}(A)$ and $V^\perp = \text{Null}(A)$. Therefore, by the rank-nullity theorem,

$$\dim(V) = \dim(\text{Row}(A)) = \text{rank}(A) = n - \text{nullity}(A) = n - \dim(V^\perp),$$

and so $\dim(V) + \dim(V^\perp) = n$. □

We can in fact build a basis of $\mathbb{R}^n$ by taking the union of a basis of V and one of $V^\perp$.

Proposition 4.2.7. *If $\mathcal{B}_V = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r\}$ is a basis for a subspace $V \subseteq \mathbb{R}^n$ and $\mathcal{B}_{V^\perp} = \{\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_{n-r}\}$ is a basis for $V^\perp$ then*

$$\mathcal{B} = \mathcal{B}_V \cup \mathcal{B}_{V^\perp} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r, \boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_{n-r}\}$$

is a basis for $\mathbb{R}^n$.

Proof. By construction $\mathcal{B}$ has n elements and $\text{Span}(\mathcal{B}) \subseteq \mathbb{R}^n$, so to argue that $\mathcal{B}$ is a basis of $\mathbb{R}^n$ we only need to show that the n elements are linearly independent. Each set separately is linearly independent because we assumed that $\mathcal{B}_V$ and $\mathcal{B}_{V^\perp}$ were bases. We will now show that none of the $\mathbf{a}'_j$ are in $V = \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_r\}$ by assuming that they are and arriving at a contradiction.

Suppose $\mathbf{a}'_j \in \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_r\}$ for some j. Then there exist scalars $c_1, \ldots, c_r \in \mathbb{R}$ such that

$$\mathbf{a}'_j = c_1\mathbf{a}_1 + \cdots + c_r\mathbf{a}_r.$$

Note that ${\mathbf{a}'_j}^\top \mathbf{a}_i = 0$ for all $i = 1, \ldots, r$ since $\mathbf{a}_i \in V$ and $\mathbf{a}'_j \in V^\perp$. It follows that

$$||\mathbf{a}'_j||^2 = {\mathbf{a}'_j}^\top \mathbf{a}'_j = c_1 \underbrace{{\mathbf{a}'_j}^\top \mathbf{a}_1}_{=0} + c_2 \underbrace{{\mathbf{a}'_j}^\top \mathbf{a}_2}_{=0} + \cdots + c_r \underbrace{{\mathbf{a}'_j}^\top \mathbf{a}_r}_{=0} = 0$$

and thus $\mathbf{a}'_j = \mathbf{0}$. However this cannot be since $\mathbf{a}'_j$ was in the basis $\mathcal{B}_{V^\perp}$ and hence it cannot be $\mathbf{0}$.

By a similar argument we can show that no $\mathbf{a}_i \in \text{Span}\{\mathbf{a}'_1, \ldots, \mathbf{a}'_{n-r}\}$. Therefore, $\mathcal{B}$ consists of linearly independent vectors and we are done. □

Now we can derive the important fact that any vector in $\mathbb{R}^n$ is the sum of a vector in V and a vector in $V^\perp$. This decomposition is in fact unique (please think about why this is the case).

Proposition 4.2.8. *Every $\boldsymbol{b} \in \mathbb{R}^n$ can be written uniquely as*

$$\boldsymbol{b} = \underbrace{c_1\boldsymbol{a}_1 + \cdots + c_r\boldsymbol{a}_r}_{\boldsymbol{b}_V} + \underbrace{c'_1\boldsymbol{a}'_1 + \cdots + c'_{n-r}\boldsymbol{a}'_{n-r}}_{\boldsymbol{b}_{V^\perp}}$$

for some scalars $c_1, \ldots, c_r, c'_1, \ldots, c'_{n-r} \in \mathbb{R}$. That is, every vector $\boldsymbol{b} \in \mathbb{R}^n$ has the form $\boldsymbol{b} = \boldsymbol{b}_V + \boldsymbol{b}_{V^\perp}$ where $\boldsymbol{b}_V \in V$ and $\boldsymbol{b}_{V^\perp} \in V^\perp$.

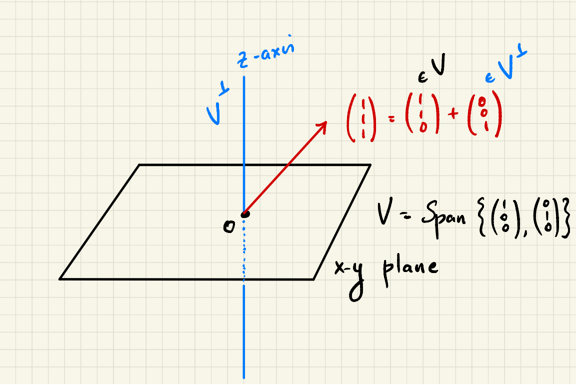
Proof. The decomposition follows immediately from the previous proposition. Why is it unique? In other words, what happens if $\mathbf{b} = \mathbf{b}_V + \mathbf{b}_{V^\perp}$ and $\mathbf{b} = \mathbf{b}'_V + \mathbf{b}'_{V^\perp}$ where $\mathbf{b}_V, \mathbf{b}'_V \in V$, $\mathbf{b}_{V^\perp}, \mathbf{b}'_{V^\perp} \in V^\perp$ and $\mathbf{b}_V \neq \mathbf{b}'_V$ (which implies that $\mathbf{b}_{V^\perp} \neq \mathbf{b}'_{V^\perp}$)? □

Warning: We may be inclined to think that Proposition 4.2.7 implies that any $\mathbf{b} \in \mathbb{R}^n$ lives in either V or $V^\perp$ but this is not the case. There is a fundamental difference between saying that $\mathbf{b} = \mathbf{b}_V + \mathbf{b}_{V^\perp}$ and saying that $\mathbf{b}$ in V or $V^\perp$. Here is a picture to drive this home.

Example 4.2.9. Let $\mathbf{b} = \begin{bmatrix}1\\1\\1\end{bmatrix}$, $V = \text{Span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$ and $V^\perp = \text{Span}\left\{\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$. We can write

$$\mathbf{b} = \underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\1\\0\end{bmatrix}}_{\mathbf{b}_V} + \underbrace{\begin{bmatrix}0\\0\\1\end{bmatrix}}_{\mathbf{b}_{V^\perp}}$$

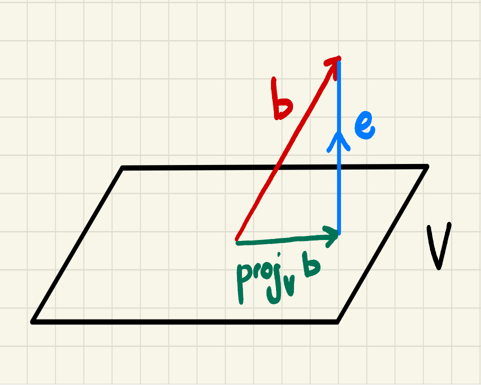
but $\mathbf{b} \notin V$ and $\mathbf{b} \notin V^\perp$.


z-axis
$V^\perp$
$\in V$
$\in V^\perp$
$\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\1\\0\end{pmatrix} + \begin{pmatrix}0\\0\\1\end{pmatrix}$
0
$V = \text{Span}\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$
x-y plane

4.3 Projections

In many mathematical applications one needs to project a vector $\mathbf{b}$ into a subspace V. This might be because $\mathbf{b}$ is a vector of observations and if it were a true data point in your model it would have been in V, but because of noise $\mathbf{b}$ ends up outside V, and now we want to find the closest element to $\mathbf{b}$ in V which we consider to be the "best solution" based on the observation $\mathbf{b}$. In this case the "best solution" is often the projection of $\mathbf{b}$ into V, denoted as $\text{proj}_V\mathbf{b}$. How does one project into V and how do we compute $\text{proj}_V\mathbf{b}$?

If $V \subseteq \mathbb{R}^n$ is a subspace and $\mathbf{b} \in \mathbb{R}^n$ is a vector, then either $\mathbf{b} \in V$ or $\mathbf{b} \notin V$. If $\mathbf{b}$ in V then we say that $\mathbf{b}$ is already the projection of $\mathbf{b}$ into V. If $\mathbf{b} \notin V$ then we do the following. Drop a perpendicular from the tip of the vector $\mathbf{b}$ into V. The vector in V whose head is at the end of the perpendicular is $\text{proj}_V\mathbf{b}$. Let's make the perpendicular into a vector too, by directing it from $\text{proj}_V\mathbf{b}$ to $\mathbf{b}$ and call this vector $\mathbf{e}$. See the figure below for $\mathbf{b}, \mathbf{e}$ and $\text{proj}_V\mathbf{b}$.



Mathematically, we have the following relationships:

- $\mathbf{b} = \text{proj}_V\mathbf{b} + \mathbf{e}$,
- $\text{proj}_V\mathbf{b} \in V$, and $\mathbf{e} \in V^\perp$ since $\mathbf{e}$ was perpendicular to all of V.

By Proposition 4.2.8 there is only one way to decompose $\mathbf{b}$ as the sum of a vector in V and a vector in $V^\perp$, so it must be that $\text{proj}_V\mathbf{b} = \mathbf{b}_V$ and $\mathbf{e} = \mathbf{b}_{V^\perp}$. We record this as follows.

Theorem 4.3.1. *If $V \subseteq \mathbb{R}^n$ is a subspace and $\boldsymbol{b} \in \mathbb{R}^n$ is a vector, then the projection of $\boldsymbol{b}$ onto V, which we denote as $\text{proj}_V\boldsymbol{b}$ is the vector $\boldsymbol{b}_V$. Similarly, the projection of $\boldsymbol{b}$ onto $V^\perp$ is $\boldsymbol{b}_{V^\perp}$.*

Check for yourself that if $\mathbf{b} \in V$, then $\mathbf{e} = \mathbf{0}$ and $\mathbf{b} = \mathbf{b}_V = \text{proj}_V\mathbf{b}$. Proposition 4.2.7 provides an algorithm to find $\text{proj}_V\mathbf{b}$:

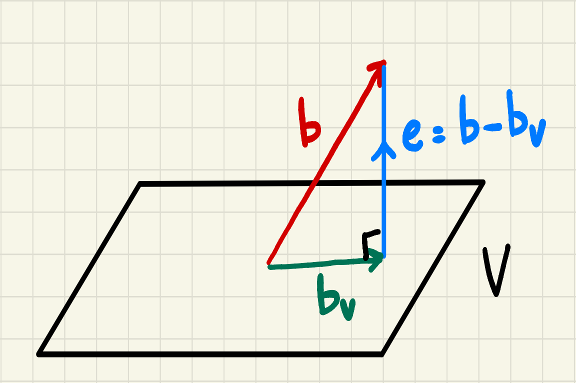
1. Compute a basis $\mathcal{B}_V = \{\mathbf{a}_1, \dots, \mathbf{a}_r\}$ of V.
2. Compute a basis $\mathcal{B}_{V^\perp} = \{\mathbf{a}'_1, \dots, \mathbf{a}'_{n-r}\}$ of $V^\perp$.
3. Write $\mathbf{b} = \mathbf{b}_V + \mathbf{b}_{V^\perp}$ by solving the system of linear equations

$$\mathbf{b} = c_1\mathbf{a}_1 + \cdots + c_r\mathbf{a}_r + c'_1\mathbf{a}'_1 + \cdots + c'_{n-r}\mathbf{a}'_{n-r}$$

4. The projection of $\mathbf{b}$ into V is the vector $\text{proj}_V\mathbf{b} = \mathbf{b}_V$.

This method works, but finding the bases $\mathcal{B}_V$ and $\mathcal{B}_{V^\perp}$ is time consuming. Moreover, this method does not give us an explicit matrix for the linear transformation that projects **any** vector into V. So we devise something better.

Suppose $V = \text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_k\} \subset \mathbb{R}^n$. Then if we set $A = \begin{bmatrix}\mathbf{a}_1 & \cdots \mathbf{a}_k\end{bmatrix} \in \mathbb{R}^{n\times k}$ we know that $\text{Col}(A) = V$ and every vector in V is of the form $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^k$. In particular, there exists some $\hat{\mathbf{x}}$ such that $\mathbf{b}_V = A\hat{\mathbf{x}}$ and $\mathbf{e} = \mathbf{b} - \mathbf{b}_V \perp V$.



From the fact that $\mathbf{b} - \mathbf{b}_V \perp V = \text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$ every $\mathbf{a}_i \perp \mathbf{b} - \mathbf{b}_V$ and so,

$$\mathbf{a}_1^\top(\mathbf{b} - \mathbf{b}_V) = \mathbf{a}_2^\top(\mathbf{b} - \mathbf{b}_V) = \cdots = \mathbf{a}_k^\top(\mathbf{b} - \mathbf{b}_V) = 0 \implies \begin{bmatrix}\mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_k^\top\end{bmatrix}(\mathbf{b} - \mathbf{b}_V) = A^\top(\mathbf{b} - \mathbf{b}_V) = \mathbf{0}.$$

Now we substitute for $\mathbf{b}_V = A\hat{\mathbf{x}}$ ($\hat{\mathbf{x}}$ is unknown), and get

$$A^\top(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \implies A^\top\mathbf{b} = A^\top A\hat{\mathbf{x}}.$$

Definition 4.3.2. The equations $A^\top\mathbf{b} = A^\top A\hat{\mathbf{x}}$ are called **normal equations**.

If we could invert $A^\top A$ we could solve for $\hat{\mathbf{x}}$ and then using that, find $\mathbf{b}_V = A\hat{\mathbf{x}}$. We could also solve the linear system $A^\top\mathbf{b} = A^\top A\hat{\mathbf{x}}$ to get $\hat{\mathbf{x}}$ but remember that if $A^\top A$ is not invertible then this system will give you infinitely many solutions for $\hat{\mathbf{x}}$ which does not make sense since we know the projection into V is unique. So either way it is important to understand when $A^\top A$ is invertible. As a sanity check, notice that $A^\top A \in \mathbb{R}^{k\times k}$ so it is a square matrix, and has a chance of being invertible.

Proposition 4.3.3. *If $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k$ are linearly independent, then $A^\top A$ is invertible.*

Proof. Set $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_k \end{bmatrix} \in \mathbb{R}^{n \times k}$. Recall that $A^\top A$ is invertible if and only if $\mathrm{Null}(A^\top A) = \{\mathbf{0}\}$. Suppose $\mathbf{y}$ is a non-zero vector in $\mathrm{Null}(A^\top A)$, i.e., $A^\top A\mathbf{y} = \mathbf{0}$. Then multiplying both sides (on the left) by $\mathbf{y}^\top$ and we get $\mathbf{y}^\top A^\top A\mathbf{y} = 0$ which gives the following chain of implications:

$$0 = \mathbf{y}^\top A^\top A\mathbf{y} = (A\mathbf{y})^\top A\mathbf{y} = \|A\mathbf{y}\|^2 \implies A\mathbf{y} = \mathbf{0}.$$

Therefore $\mathbf{y}$ is a non-zero vector in $\mathrm{Null}(A)$ but this is a contradiction since if the columns of A are linearly independent, then $\mathrm{Null}(A) = \{\mathbf{0}\}$. Therefore, we conclude that if $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly independent, then $A^\top A$ is invertible. □

This proposition allows us to find a linear transformation that projects into a subspace V. Let us assume that $\mathbf{a}_1, \ldots, \mathbf{a}_k$ is a basis for V and not just a spanning set, which guarantees they are linearly independent. Using Proposition 4.3.3 we get that if $A\hat{\mathbf{x}} = \mathbf{b}_V$ then

$$\hat{\mathbf{x}} = (A^\top A)^{-1}A^\top\mathbf{b} \quad \text{and hence} \quad \mathbf{b}_V = A(A^\top A)^{-1}A^\top\mathbf{b}.$$

Therefore the linear transformation that projects $\mathbf{b} \in \mathbb{R}^n$ to $\mathrm{proj}_V\mathbf{b} \in V$ sends $\mathbf{b} \mapsto A(A^\top A)^{-1}A^\top\mathbf{b}$.

Proposition 4.3.4. *Let V be a subspace of $\mathbb{R}^n$ with basis $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ and let $A = \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_k \end{bmatrix}$. Then projection into V is given by the* **projection matrix**

$$P = A(A^\top A)^{-1}A^\top.$$

Many of you may have seen the above formula for projection before in the context of some application. A takeaway from this chapter should be that the formula should not be applied blindly. It works only when the columns of A are linearly independent.

Check that $P^2 = P$ (recall the homework problem about matrices like this). We will see later that that any (symmetric) matrix $P \in \mathbb{R}^{n \times n}$ that satisfies $P^2 = P$ is the matrix of projection from $\mathbb{R}^n$ into some subspace, and that this subspace is not all of $\mathbb{R}^n$ if 0 is an eigenvalue of P (alternately, P has a nullspace). The homework problem further showed you that **all projection matrices are diagonalizable**. There is a lot more to learn about projection matrices.

We now give an important example.

Example 4.3.5. Let's project $\mathbf{b} = \begin{bmatrix} 3 \\ 4 \\ 4 \end{bmatrix}$ onto the line spanned by $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$.

Here $V = \mathrm{Span}\left\{\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}\right\}$. Going along with the procedure just described, we set $A = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ and then compute $A^\top A = \begin{bmatrix} 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} = 9$. This is clearly an invertible 1×1 matrix. The last step is to plug into the formula for a projection matrix and multiply by $\mathbf{b}$ to get $\mathrm{proj}_V\,\mathbf{b}$.

$$\mathbf{b}_V = A(A^\top A)^{-1}A^\top\mathbf{b} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} (\frac{1}{9}) \begin{bmatrix} 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} (\frac{1}{9})(18) = \begin{bmatrix} 4 \\ 4 \\ 2 \end{bmatrix}.$$

Check that $\begin{bmatrix} 4 \\ 4 \\ 2 \end{bmatrix}$ is indeed on the line spanned by $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$.

This example shows you that if $V = \mathrm{Span}\{\mathbf{u}\}$ is a line, then the projection of $\mathbf{b}$ onto V is $\mathbf{b}_V = \frac{\mathbf{u}(\mathbf{u}^\top\mathbf{b})}{(\mathbf{u}^\top\mathbf{u})}$.

The above example of projecting onto a line is so important that we are going to record it as a proposition. We will revisit this formula later in the quarter.

Proposition 4.3.6. *Let $V = \text{Span}\{\mathbf{u}\}$ be the line spanned by the vector $\mathbf{u} \in \mathbb{R}^n$. Then projection onto V is achieved by the linear transformation:*

$$P = \frac{\boldsymbol{u}\boldsymbol{u}^\top}{\boldsymbol{u}^\top\boldsymbol{u}} \quad \text{that sends} \quad \boldsymbol{x} \mapsto \frac{\boldsymbol{u}\boldsymbol{u}^\top}{\boldsymbol{u}^\top\boldsymbol{u}}\boldsymbol{x}$$

In particular, if $\|\boldsymbol{u}\| = 1$, then $\mathbf{u}^\top\boldsymbol{u} = 1$ and $P = \boldsymbol{u}\boldsymbol{u}^\top$.

Let's illustrate on a very familiar example. Suppose $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Then its projections onto the x-axis, y-axis and z-axis are respectively:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{e}_1\mathbf{e}_1^\top\mathbf{b} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{e}_1} \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{\mathbf{e}_1^\top} \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{\mathbf{b}},$$

$$\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \mathbf{e}_2\mathbf{e}_2^\top\mathbf{b} = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{e}_2} \underbrace{\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}}_{\mathbf{e}_2^\top} \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{\mathbf{b}},$$

$$\begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} = \mathbf{e}_3\mathbf{e}_3^\top\mathbf{b} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{e}_3} \underbrace{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}}_{\mathbf{e}_3^\top} \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{\mathbf{b}}.$$

Before we finish this section, let's note an important property of the vector $\mathbf{b}_V = \text{proj}_V\mathbf{b}$.

Proposition 4.3.7. *For any vector $\boldsymbol{b} \in \mathbb{R}^n$ and subspace V, $\text{proj}_V\boldsymbol{b}$ is the closest point (in Euclidean distance) in V to $\boldsymbol{b}$.*

Proof. If $\mathbf{b} \in V$ then $\text{proj}_V\mathbf{b} = \mathbf{b}$ and $\mathbf{b}$ is indeed the closet point to itself in V. So suppose $\mathbf{b} \notin V$ and $V = \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$. Then $V = \text{Col}(A)$ and any point in V looks like $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^k$. We are looking for $\mathbf{x}$ such that $\|A\mathbf{x} - \mathbf{b}\|$ is minimized. This is same as looking for $\mathbf{x}$ such that $\|A\mathbf{x} - \mathbf{b}\|^2$ is minimized.

Now recall that $\mathbf{b} = \mathbf{b}_V + \mathbf{e}$. Substituting for $\mathbf{b}$ we want to find $\mathbf{x} \in \mathbb{R}^k$ that minimizes $\|A\mathbf{x} - \mathbf{b}_V - \mathbf{e}\|^2$. However, since $A\mathbf{x} - \mathbf{b}_V \in V$ and $\mathbf{e} \in V^\perp$, by Pythagorus' theorem,

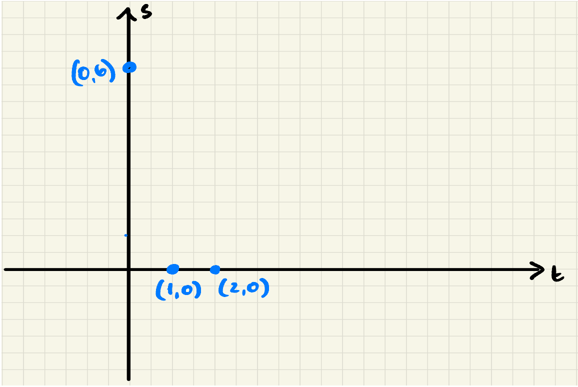
$$\|A\mathbf{x} - \mathbf{b}\|^2 = \|(A\mathbf{x} - \mathbf{b}_V) - \mathbf{e}\|^2 = \|(A\mathbf{x} - \mathbf{b}_V)\|^2 + \|\mathbf{e}\|^2.$$

The vector $\mathbf{e}$ is fixed since it is the perpendicular from $\mathbf{b}$ to V both of which are fixed. In particular, $\|\mathbf{e}\|^2$ is fixed. Therefore, to minimize $\|A\mathbf{x} - \mathbf{b}\|^2$ we can only minimize $\|(A\mathbf{x} - \mathbf{b}_V)\|^2$. This quantity cannot be negative and it will be zero if we set $\mathbf{x} = \hat{\mathbf{x}}$. Therefore, we get that $\hat{\mathbf{x}}$ is the vector that minimizes $\|A\mathbf{x} - \mathbf{b}\|^2$ and so $\mathbf{b}_V = A\hat{\mathbf{x}} = \text{proj}_V\mathbf{b}$ is the closest point in V to $\mathbf{b}$.

□

4.4 Linear Regression

A very important problem in statistics is to find the best line that fits a collection of observed data points in $\mathbb{R}^2$. We now use the theory of projections to find such a best fit line and make sense of what we mean by "best fit". Afterwards we will mention several applications of this method which is commonly called *least squares regression* or *linear regression.*



Problem: Find the line that best fits the points $(0,6),(1,0)$, and $(2,0)$.

From the picture we can see that no line passes through all 3 points. So we are going to find the best line we can and we'll see what that means.

Suppose we label the horizontal axis of $\mathbb{R}^2$ by the variable t and vertical axis by the variable s, so that any point in $\mathbb{R}^2$ has coordinates (t,s). A line L in this plane has the form $s = C + Dt$ where C and D are constants. If L passed through the 3 points then the following equations would have a solution (C,D):

$$\begin{aligned} 6 &= C + D \cdot 0 \\ 0 &= C + D \cdot 1 \\ 0 &= C + D \cdot 2 \end{aligned}$$

In other words, the linear system

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}}_{A} \begin{bmatrix} C \\ D \end{bmatrix} = \underbrace{\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{b}}$$

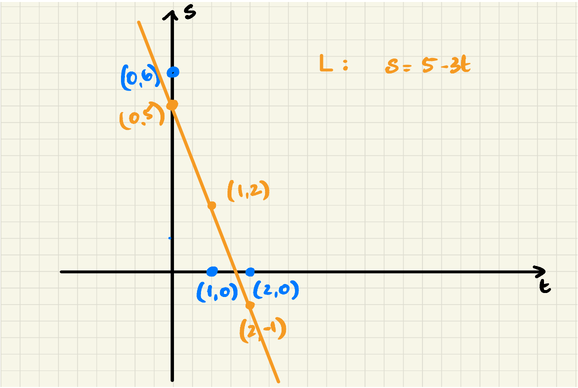
would have a solution. Note that the variables in this system are C and D.

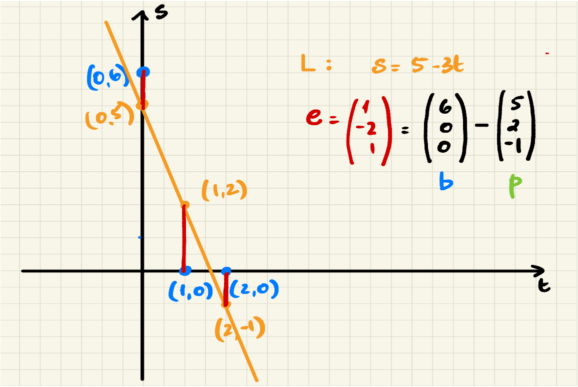
Since there is no such line, we know the system has no solution, i.e. $\mathbf{b} = (6,0,0)^\top \notin \mathrm{Col}(A)$. However, we learned that the closest point to $\mathbf{b}$ in $\mathrm{Col}(A)$ is $\mathrm{proj}_{\mathrm{Col}(A)}\mathbf{b}$ which we will call $\mathbf{p}$ for convenience – $\mathbf{p}$ for projection. Recall that $\mathbf{p} = A\hat{\mathbf{x}}$. Computing $\hat{\mathbf{x}}$ (using the normal equations) we get

$$\hat{\mathbf{x}} = \begin{bmatrix} 5 \\ -3 \end{bmatrix} \quad \text{and} \quad \mathbf{p} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}, \quad \text{or equivalently,} \quad \underbrace{\begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}}_{\mathbf{p}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 5 \\ -3 \end{bmatrix}}_{\hat{\mathbf{x}}}.$$

Definition 4.4.1. Define the best fit line L for the given data points (in blue) to be the line L (in orange) with $\begin{bmatrix} C \\ D \end{bmatrix} = \hat{\mathbf{x}} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$, i.e., L has equation $s = 5 - 3t$.

The points on L are of the form $(t, 5-3t)$ for each $t \in \mathbb{R}$. When $t = 0,1,2$ which were the first coordinates of the (blue) points we started with, we get the points $(0,5),(1,2),(2,-1)$ on L marked in orange. The vector of second coordinates of the orange points is exactly $\mathbf{p}$ by construction. Now recall that $\mathbf{b} - \mathbf{p} = \mathbf{e}$. Therefore, for each $t = 0,1,2$, we get that the difference in vertical height (shown in red) between the given point (t_i, b_i) and the point (t_i, p_i) is exactly e_i. Since $\hat{\mathbf{x}}$ minimizes $\|A\mathbf{x} - \mathbf{b}\|^2$ over all $\mathbf{x}$ and this minimum value is $\|\mathbf{e}\|^2$, we get that the line L we have constructed minimizes the sum of squares of the vertical distances between the ith data point (t_i, b_i) and the point (t_i, p_i) on L, i.e., $e_1^2 + e_2^2 + e_3^2$. Further note that since $(1,1,1)^\top$ is a column of A, it is in $\mathrm{Col}(A)$, and we know that $\mathbf{e} \perp \mathrm{Col}(A)$. Therefore, $0 = (1,1,1)\mathbf{e} = e_1 + e_2 + e_3$.





General algorithm for finding the best fit line by least squares regression

Input: Data points $(t_1, b_1), (t_2, b_2), \dots (t_m, b_m)$. Output: The line L that best fits the given data in the sense of minimizing the sum of squares of vertical distances between (t_i, b_i) and $(t_i, s_i) \in L$.

1. Set $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$

2. If there was a line $s = C + Dt$ through all points, then $\mathbf{b} \in \text{Col}(A)$ and

$$\underbrace{\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}}_{A} \begin{bmatrix} C \\ D \end{bmatrix} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{\mathbf{b}}$$

   would be feasible. Solve for C and D and find the best fit line L with equation $s = C + Dt$.

3. Otherwise, project $\mathbf{b}$ onto $\text{Col}(A)$ to get $\mathbf{p} = A\hat{\mathbf{x}}$. Then set $\begin{bmatrix} C \\ D \end{bmatrix} = \hat{\mathbf{x}}$ to get the best fit line L with equation $s = C + Dt$.

The values $e_i = b_i - (C + Dt_i)$ are the differences in height between the i^{th} data point and the point on L with the same t coordinate. The line L minimizes $||\mathbf{e}||^2 = e_1^2 + \cdots + e_m^2$. Further, $e_1 + \cdots + e_m = 0$ because $\mathbf{1} \in \text{Col}(A)$ and $\mathbf{e} \perp \text{Col}(A)$.

Linear regression has many many applications and is a standard method to make predictions based on past behavior. For example:

- Predict ice cream sales s on a day with temperature t based on past observations of the amount of ice cream sold on days with recorded temperatures.
- Predict house prices based on data collected for price vs square footage.
- Predict the number of goals a player would score based on past performances.
- Set the salary of a new employee based on data collected of salary vs experience.

In 2018 Washington state abolished the death penalty based on regression analysis done by two UW faculty – Katherine Beckett and Heather Evans – in the Department of Sociology. They found that black people were four times as likely to be sentenced to death as prisoners of other races. You can read more about this remarkable work at: https://magazine.washington.edu/feature/death-penalty-washington-state/

Chapter 5

Symmetric Matrices

In this chapter we will define *symmetric matrices* which are arguably the most important matrices you will see. We will prove the powerful *Spectral Theorem* for real symmetric matrices which will show that all symmetric matrices are diagonalizable. You can read about symmetric matrices in Chapter 6.4 in Strang's book. In the next chapter we will go onto another very important class of matrices called *positive semi-definite matrices* which form a subset of symmetric matrices. These classes of matrices appear in many applications as we will see.

5.1 Eigenvalues and Eigenvectors of Real Symmetric Matrices

Definition 5.1.1. A matrix $A \in \mathbb{R}^{n\times n}$ is **symmetric** if $A = A^\top$.

Example 5.1.2. The matrix $A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$ is symmetric. The matrix $B = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$ is not.

We now prove a sequence of facts about symmetric matrices that will get us to the Spectral Theorem. Recall that the conjugate of a complex number $z = a + bi$, is $\overline{z} = a - bi$. A complex number z is real if and only if $\overline{z} = z$. The complex conjugate satisfies a number of properties, such as

$$\overline{(a+bi)(c+di)} = (\overline{a+bi})(\overline{c+di}).$$

Define the conjugate of a matrix A, denoted $\overline{A}$, to be the matrix obtained by conjugating all entries of A. Note that A is a real matrix if and only if $\overline{A} = A$. Using these facts we get the following.

Proposition 5.1.3. *If $A \in \mathbb{R}^{n\times n}$ is symmetric then all eigenvalues of A are real.*

Proof. Suppose that $A\mathbf{x} = \lambda\mathbf{x}$ where λ may be some complex number and $\mathbf{x}$ may have complex entries. Taking conjugates of both sides of the eigenvalue equation and since $A = \overline{A}$, we get

$$\overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} \implies \overline{A}\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}} \implies A\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$$

We now transpose both sides of this equation and using $A = A^\top$ we have:

$$(A\overline{\mathbf{x}})^\top = (\overline{\lambda}\overline{\mathbf{x}})^\top \implies \overline{\mathbf{x}}^\top A^\top = \overline{\mathbf{x}}^\top\overline{\lambda} \implies \overline{\mathbf{x}}^\top A = \overline{\lambda}\overline{\mathbf{x}}^\top$$

Multiplying $A\mathbf{x} = \lambda\mathbf{x}$ by $\overline{\mathbf{x}}^\top$ on the left we get

$$\overline{\mathbf{x}}^\top A\mathbf{x} = \lambda\overline{\mathbf{x}}^\top\mathbf{x}.$$

Multiplying $\overline{\mathbf{x}}^\top A = \overline{\lambda}\overline{\mathbf{x}}^\top$ by $\mathbf{x}$ on the right we get

$$\overline{\mathbf{x}}^\top A\mathbf{x} = \overline{\lambda}\overline{\mathbf{x}}^\top\mathbf{x}.$$

Therefore, $\lambda\overline{\mathbf{x}}^\top\mathbf{x} = \overline{\lambda}\overline{\mathbf{x}}^\top\mathbf{x}$ which implies that $\lambda = \overline{\lambda}$ (since $\overline{\mathbf{x}}^\top\mathbf{x} \neq 0$) and so, $\lambda \in \mathbb{R}$. □

We can also see that if $A = A^\top$ then each eigenspace of A has a basis made up of real eigenvectors.

Proposition 5.1.4. *If λ is an eigenvalue of a symmetric matrix A, then E_λ has a basis of real eigenvectors.*

Proof. Recall that the eigenspace E_λ is the nullspace of $A - \lambda I$. Since $\lambda \in \mathbb{R}$, $A - \lambda I$ is a real matrix and hence its nullspace has a basis consisting of only real vectors. □

Example 5.1.5. Consider the symmetric matrix $A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$. It's eigenvalues are $\lambda = 0, 1, 3$ with eigenvectors $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, and $\mathbf{u}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ respectively. Notice that the three eigenvectors are mutually orthogonal:

$$\mathbf{u}_1^\top \mathbf{u}_2 = \mathbf{u}_1^\top \mathbf{u}_3 = \mathbf{u}_2^\top \mathbf{u}_3 = 0$$

Proposition 5.1.6. *Two real eigenvectors from different eigenspaces of a symmetric matrix are mutually orthogonal. That is, if $\boldsymbol{x}_1 \in E_{\lambda_1}$ and $\boldsymbol{x}_2 \in E_{\lambda_2}$ with $\lambda_1 \neq \lambda_2$, then $\boldsymbol{x}_1^\top \boldsymbol{x}_2 = 0$.*

Proof. Suppose $A\mathbf{x} = \lambda_1 \mathbf{x}$ and $A\mathbf{y} = \lambda_2 \mathbf{y}$ with $\lambda_1 \neq \lambda_2$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$A\mathbf{x} = \lambda_1 \mathbf{x} \implies (\lambda_1 \mathbf{x})^\top = (A\mathbf{x})^\top \implies \lambda_1 \mathbf{x}^\top = \mathbf{x}^\top A^\top = \mathbf{x}^\top A$$

Multiplying both sides of this equation on the right by $\mathbf{y}$ we get

$$\lambda_1 \mathbf{x}^\top \mathbf{y} = \mathbf{x}^\top A \mathbf{y} = \mathbf{x}^\top \lambda_2 \mathbf{y} \implies \lambda_1 \mathbf{x}^\top \mathbf{y} = \lambda_2 \mathbf{x}^\top \mathbf{y} \implies (\lambda_1 - \lambda_2)\mathbf{x}^\top \mathbf{y} = 0$$

Since $\lambda_1 \neq \lambda_2$ we must have $\mathbf{x}^\top \mathbf{y} = 0$. □

Proposition 5.1.7. *For each eigenvalue λ of a symmetric matrix $A \in \mathbb{R}^{n \times n}$, we have* $\mathrm{AM}(\lambda) = \mathrm{GM}(\lambda)$*.*

We might prove this fact later, but will assume it for now. Notice that this means that A is diagonalizable.

5.2 The Spectral Theorem for Real Symmetric Matrices

We now show that we can pick a very special eigenbasis for a symmetric matrix.

Definition 5.2.1.

1. The vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \mathbb{R}^n$ are **orthogonal** if $\mathbf{u}_i^\top \mathbf{u}_j = 0 \; \forall \; i \neq j$.
2. The vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \mathbb{R}^n$ are **orthonormal** if $\mathbf{u}_i^\top \mathbf{u}_j = 0 \; \forall \; i \neq j$ and $||\mathbf{u}_i|| = 1 \; \forall \; i$.
3. A matrix $Q \in \mathbb{R}^{n \times n}$ is **orthogonal** if its columns are orthogonal.
4. A matrix $Q \in \mathbb{R}^{n \times n}$ is **orthonormal** if its columns are orthonormal.

Example 5.2.2. The set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

consists of orthonormal vectors because all vectors have length 1 are are mutually orthogonal. The identity matrix I_3 is therefore, orthonormal. In fact, all identity matrices are orthonormal. Even if you scramble the columns of an identity matrix, you get an orthonormal matrix. In particular, all permutation matrices are orthonormal.

Example 5.2.3. Consider the three eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ from Example 5.1.5. The vectors are already pairwise orthogonal, so to get a orthonormal set of eigenvectors we just need to replace the given eigenvectors by scaled versions that have unit norm:

$$\mathbf{u}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix} \longrightarrow \mathbf{u}_1' = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix}1\\0\\-1\end{bmatrix} \longrightarrow \mathbf{u}_2' = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix}1\\-2\\1\end{bmatrix} \longrightarrow \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\1\end{bmatrix}.$$

It is a general fact that one can always find an orthonormal basis for any real vector space. This is done using a procedure called the *Gram-Schmidt algorithm.* For now let's assume that we can always find a orthogonal/orthonormal basis of any vector space.

Proposition 5.2.4. *If $A \in \mathbb{R}^{n\times n}$ is symmetric, then A has n orthonormal eigenvectors.*

Proof. By Proposition 5.1.7, we know that each eigenspace E_λ has maximum possible dimension, equal to $AM(\lambda)$. By applying the Gram-Schmidt procedure, we can obtain an orthonormal basis for each E_λ:

$$\mathcal{B}_{E_{\lambda_i}} = \{\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, \ldots, \mathbf{u}_{i_{AM(\lambda_i)}}\}$$

From Proposition 5.1.6, we have that the basis vectors in $\mathcal{B}_{E_{\lambda_i}}$ are all orthogonal to the basis vectors in $\mathcal{B}_{E_{\lambda_j}}$ for eigenvalues $\lambda_i \neq \lambda_j$. Taking the union of all these bases, we obtain n linearly independent eigenvectors of A, each of unit length and mutually orthogonal. □

We need one final fact before we can state the spectral theorem.

Proposition 5.2.5. *If $Q \in \mathbb{R}^{n\times n}$ is orthonormal, then $QQ^\top = Q^\top Q = I_n$. In particular, $Q^{-1} = Q^\top$.*

Proof. Suppose $Q = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix}$. Then $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are orthonormal. Therefore,

$$Q^\top Q = \begin{bmatrix}\mathbf{u}_1^\top\\ \vdots \\ \mathbf{u}_n^\top\end{bmatrix}\begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix} = \begin{bmatrix}||\mathbf{u}_1||^2 & & \mathbf{u}_j^\top\mathbf{u}_i\\ & \ddots & \\ \mathbf{u}_i^\top\mathbf{u}_j & & ||\mathbf{u}_n||^2\end{bmatrix} = \begin{bmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{bmatrix} = I_n$$

In words, the (i,j) entry of $Q^\top Q$ is the dot product of column i of Q (row i of $Q^\top$) and column j of Q. If $i \neq j$, this dot product is 0 since all columns of Q are orthogonal. If $i = j$ then this dot product is $\mathbf{u}_i^\top\mathbf{u}_i = \|\mathbf{u}_i\|^2 = 1$. Hence $Q^\top Q = I_n$.

Since Q is invertible, multiply both sides of $Q^\top Q = I_n$ by Q^{-1} to get $Q^\top = Q^{-1}$. Therefore we also have that $QQ^\top = QQ^{-1} = I_n$. □

Theorem 5.2.6. [Spectral Theorem] *If $A \in \mathbb{R}^{n\times n}$ is symmetric, then A has a diagonalization*

$$A = Q\Lambda Q^{-1} = Q\Lambda Q^\top$$

where Λ is a real diagonal matrix and Q is orthonormal.

Proof. The proof of this is a combination of the above propositions. Proposition 5.1.3 implies that all n eigenvalues of A are real, and so Λ is a real diagonal matrix. Proposition 5.1.7 and Proposition 5.2.4 imply that A is diagonalizable by an orthonormal matrix. Finally Proposition 5.2.5 tells us that $Q^{-1} = Q^\top$. □

Example 5.2.7. Consider the matrix $A = \begin{bmatrix}1 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 1\end{bmatrix}$ again. It has the diagonalization

$$A = Q\Lambda Q^{-1} = \begin{bmatrix}1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6}\\ 1/\sqrt{3} & 0 & -2/\sqrt{6}\\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6}\end{bmatrix}\begin{bmatrix}0 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 3\end{bmatrix}\begin{bmatrix}1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\\ 1/\sqrt{2} & 0 & -1/\sqrt{2}\\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6}\end{bmatrix}$$

where Q is orthonormal and Λ is real. The columns of Q are orthonormal eigenvectors of A.

5.3 Diagonalization and change of basis

We now recall the connection between diagonalization and change of bases from Chapter 1. Since all symmetric matrices are diagonalizable, their effect as linear transformations is the same as the effect of the eigenvalue matrix Λ, as long as we work in the basis given by the columns of the eigenvector matrix Q. Let $A = Q\Lambda Q^\top$ be an orthonormal diagonalization of the symmetric matrix A and suppose the set of columns of Q are $\mathcal{Q} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$. From the diagonalization we have

$$Q^\top A = Q^\top Q\Lambda Q^\top = \Lambda Q^\top$$

since $Q^\top Q = I$. Recall that the coordinates $\mathbf{y}$ with respect to $\mathcal{Q}$ of a vector $\mathbf{x} \in \mathbb{R}^n$ is $\mathbf{y} = Q^{-1}\mathbf{x} = Q^\top\mathbf{x}$. Similarly, the coordinates of $A\mathbf{x}$ with respect to $\mathcal{Q}$ is equal to

$$Q^\top Ax = \Lambda Q^\top x = \Lambda\mathbf{y}.$$

Therefore, the effect of sending $\mathbf{x} \mapsto A\mathbf{x}$ in standard coordinates is the same as $\mathbf{y} \mapsto \Lambda\mathbf{y}$ in the basis $\mathcal{Q}$. By the Spectral Theorem, the eigenbasis $\mathcal{Q}$ is orthonormal. Therefore, we have passed from the standard (orthonormal) basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ to a new orthonormal basis $\mathbf{u}_1, \dots, \mathbf{u}_n$ which is a rotation of the standard basis. We illustrate on a simple example.

Example 5.3.1. Consider the following (symmetric) matrix, along with its *orthogonal* diagonalization

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \underbrace{\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 5 \end{bmatrix}}_{\Lambda} \underbrace{(\frac{1}{5}) \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}}_{Q^{-1}}$$

Here, our eigenbasis $\mathcal{Q} = \left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ is only orthogonal and not orthonormal. We have not normalized the vectors to keep them nice because of which, $Q^{-1} \neq Q^\top$.

If $\mathbf{x} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$, then $A\mathbf{x} = \begin{bmatrix} 15 \\ 30 \end{bmatrix}$. After changing the coordinates of $\mathbf{x}$ and $A\mathbf{x}$ with respect to the eigenbasis $\mathcal{Q}$, we get

$$[\mathbf{x}]_{\mathcal{Q}} = \mathbf{y} = Q^{-1}\mathbf{x} = \frac{1}{5}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

and

$$[A\mathbf{x}]_{\mathcal{Q}} = Q^{-1}A\mathbf{x} = \frac{1}{5}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 15 \\ 20 \end{bmatrix} = \begin{bmatrix} 0 \\ 15 \end{bmatrix} = \Lambda\mathbf{y} = \begin{bmatrix} 0 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

In the eigenbasis $\mathcal{Q}$, the effect of A is stretching along the new orthogonal coordinate axes – crushing the first coordinate of any vector to the origin and scaling the second by a factor of 5.

Chapter 6

Positive Semidefinite Matrices

We now study a subclass of symmetric matrices called positive semidefinite matrices. Their eigenvalues have even more structure than that of symmetric matrices and they play a central role in a number of applications. You can read more about positive semidefinite matrices in Chapter 6.5 of Strang's book.

6.1 Symmetric Matrices and Quadratic Forms

We first observe that to every square matrix $A \in \mathbb{R}^{n\times n}$, we can associate a quadratic polynomial in n variables by multiplying A on the left by $\mathbf{x}^\top$ and on the right by $\mathbf{x}$ where $\mathbf{x} = (x_1, \dots, x_n)^\top$. To illustrate, for $n = 2$ we can construct the following quadratic polynomial:

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax+cy & bx+dy \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + bxy + cxy + dy^2 = ax^2 + (b+c)xy + dy^2.$$

Now notice that we could have gotten the same quadratic polynomial from a *symmetric matrix* as follows:

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & \frac{b+c}{2} \\ \frac{b+c}{2} & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + (b+c)xy + dy^2.$$

In general, a symmetric matrix A leads to the *quadratic polynomial*:

$$q_A(\mathbf{x}) := \mathbf{x}^\top A\mathbf{x} = \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i\neq j} 2a_{ij}x_ix_j$$

Since the degree of every monomial in this polynomial is the same, namely 2, this polynomial is said to be *homogeneous* and homogeneous polynomials are called *forms*. Therefore, we see that any symmetric matrix A gives rise to a *quadratic form* $q_A(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$. Conversely, every quadratic form in $x_1, \dots, x_n$ is of the form $\sum_{i=1}^{n} b_{ii}x_i^2 + \sum_{i\neq j} b_{ij}x_ix_j$ and hence comes from the unique symmetric matrix

$$B = \begin{bmatrix} b_{11} & \frac{b_{12}}{2} & \frac{b_{13}}{2} & \cdots & \frac{b_{1n}}{2} \\ \frac{b_{12}}{2} & b_{22} & \frac{b_{23}}{2} & \cdots & \frac{b_{2n}}{2} \\ \frac{b_{13}}{2} & \frac{b_{23}}{2} & b_{33} & \cdots & \frac{b_{3n}}{2} \\ \frac{b_{14}}{2} & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{b_{1n}}{2} & \cdots & \cdots & \cdots & b_{nn} \end{bmatrix}$$

by the construction $q_B(\mathbf{x}) = \mathbf{x}^\top B\mathbf{x}$. Therefore, we conclude that we can identify the set of all quadratic forms in n variables with the set of symmetric matrices.

Example 6.1.1. The quadratic form $2x^2 - 15xy + 3y^2 = q_A(x, y)$ for $A = \begin{bmatrix} 2 & -15/2 \\ -15/2 & 3 \end{bmatrix}$.

We now look at the shape of the graphs of the quadratic forms $q_A(\mathbf{x})$ as we slowly dial up the eigenvalues of A from some negative ones to all positive. Since $\det(A)$ is the product of eigenvalues, its value moves from negative to positive in the process.

Example 6.1.2. Consider $A = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}$ with eigenvalues $\lambda = -4, 6$ and $\det(A) = -24$. Then

$$q_A(x, y) = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^2 - 10xy + y^2$$

whose graph in Figure 6.1 is not convex and has points above and below the xy plane, the domain of q_A.

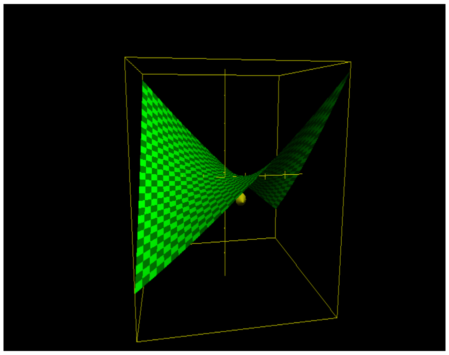


Figure 6.1: $q_A(x, y) = x^2 - 10xy + y^2$

Example 6.1.3. Now consider $A = \begin{bmatrix} 10 & -5 \\ -5 & 2 \end{bmatrix}$ with eigenvalues $\lambda = -0.403, 12$ and $\det(A) = -5$. Then

$$q_A(x, y) = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 10 & -5 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 10x^2 - 10xy + 2y^2.$$

Its graph shown in Figure 6.2 looks more or less convex, but it is not. It has some points below the domain.

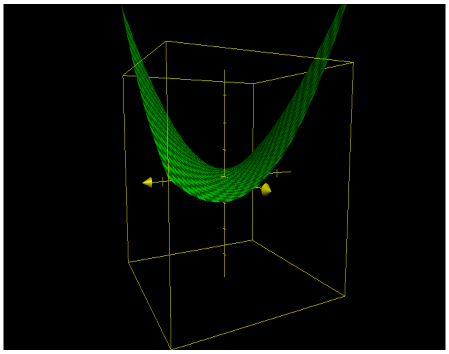


Figure 6.2: $q_A(x, y) = 10x^2 - 10xy + 2y^2$

Example 6.1.4. Next, let $A = \begin{bmatrix} 100 & -5 \\ -5 & 20 \end{bmatrix}$ whose eigenvalues are $\lambda = 19.88, 100.31$ and $\det(A) = 1975$. The graph of

$$q_A(x,y) = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 100 & -5 \\ -5 & 20 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 100x^2 - 10xy + 20y^2$$

seen in Figure 6.3 is clearly convex. Moreover, all points on the graph have nonnegative height.

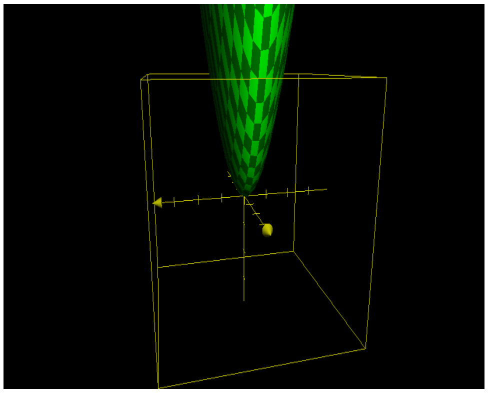


Figure 6.3: $q_A(x,y) = 100x^2 - 10xy + 20y^2$

Example 6.1.5. Finally, let $A = \begin{bmatrix} 500 & -5 \\ -5 & 500 \end{bmatrix}$ with eigenvalues $\lambda = 495, 505$ and $\det(A) = 249975$. The graph of

$$q(x,y) = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 500 & -5 \\ -5 & 500 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 500x^2 - 10xy + 500y^2$$

can be seen in Figure 6.4 This graph is also clearly convex and all points on it have nonnegative height.

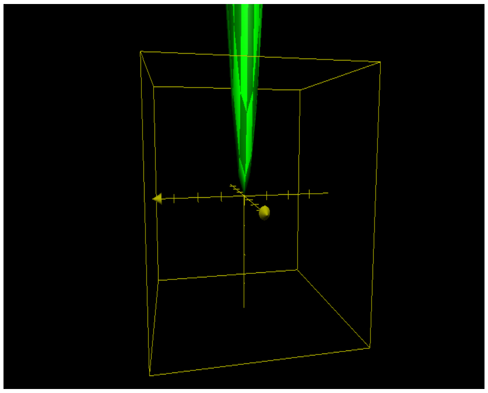


Figure 6.4: $q_A(x,y) = 500x^2 - 10xy + 500y^2$

Definition 6.1.6. Say that a symmetric matrix A is **positive semidefinite (PSD)** if $q_A(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and **positive definite (PD)** if $q_A(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n$ except $\mathbf{x} = \mathbf{0}$.

The matrices in Examples 6.1.4 and 6.1.5 are positive semidefinite. In fact, both are positive definite. It turns out that there are many alternate characterizations of positive semidefinite/definite matrices, each useful for different purposes. We now develop these alternatives.

Definition 6.1.7. A **principal minor** of a matrix $A \in \mathbb{R}^{n\times n}$ is the determinant of the square submatrix of A obtained by deleting the rows and columns of A with the same indices. This means that if we, for example, delete rows $2, 5, 7$, then we should also delete columns $2, 5, 7$. Similarly, a **leading principal minor** or A is the determinant of the square submatrix of A with rows and columns indexed by $1, 2, 3, \ldots, k$ for some k. In this case the indices of the deleted rows and columns are $k+1, \ldots, n$.

Example 6.1.8. The matrix $A = \begin{bmatrix} 1 & 4 & 6 \\ 4 & 2 & 1 \\ 6 & 1 & 6 \end{bmatrix}$ has three leading principal minors:

$$\det \begin{bmatrix} 1 & 4 & 6 \\ 4 & 2 & 1 \\ 6 & 1 & 6 \end{bmatrix}, \quad \det \begin{bmatrix} 1 & 4 \\ 4 & 2 \end{bmatrix}, \quad \det \begin{bmatrix} 1 \end{bmatrix}.$$

Any $n \times n$ matrix has n leading principal minors.

There are 7 principal minors of A corresponding to the 7 subsets of $\{1, 2, 3\}$ (except $\{1, 2, 3\}$ itself) indexing the rows and columns being deleted. Call this index set $\mathcal{I}$ and denote by $D_{\mathcal{I}}(A)$ the corresponding principal minor of A. Here are a few examples:

$$D_{\emptyset} = \det(A), \;\; D_{\{1\}} = \det \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix} = 11, \;\; D_{\{2\}} = \det \begin{bmatrix} 1 & 6 \\ 6 & 6 \end{bmatrix} = -30, \;\; D_{\{1,3\}} = \det \begin{bmatrix} 2 \end{bmatrix} = 2.$$

Here are now various characterizations of a positive semidefinite/definite matrix.

Definition 6.1.9. A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is **positive semidefinite**, denoted $A \succeq 0$, if any of the following equivalent conditions are true:

1. $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$, i.e., the graph of $q_A(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$ is on or above the domain $\mathbb{R}^n$.

2. All eigenvalues of A are non-negative.

3. There exists some matrix B such that $A = B^\top B$.

4. All principal minors of A are non-negative.

Definition 6.1.10. A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is **positive definite**, denoted $A \succ 0$, if any of the following equivalent conditions are true:

1. $\mathbf{x}^\top A\mathbf{x} > 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$.

2. All eigenvalues of A are positive.

3. There exists some matrix B such that $A = B^\top B$ with B having linearly independent columns.

4. All leading principal minors of A are positive.

Note that saying that a matrix is of the form $B^\top B$ is the same as saying that it is of the form $CC^\top$ since we can take $C = B^\top$. Recall that when we studied orthogonal projections, we saw that $A^\top A$ is invertible if and only if the columns of A are independent. In this case, $A^\top A \succ 0$. Note also that all PD matrices are PSD and all PD matrices are invertible since if $A \succ 0$, then $\det(A) > 0$ as it is one of the leading principal minors of A. PSD matrices on the other hand are not all invertible.

We will see later why the conditions given above are equivalent. First let's look at some examples.

Example 6.1.11. The matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ is PSD. Let us check all four conditions:

1. The quadratic form $q_A(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + 4xy + 4y^2 = (x+2y)^2 \geq 0$ for all $(x,y) \in \mathbb{R}^2$.
2. The eigenvalues of A are 0 and 5 which are both non-negative.
3. The matrix A can be factorized as $B^\top B$:
$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}_{B^\top} \underbrace{\begin{bmatrix} 1 & 2 \end{bmatrix}}_{B}.$$
There might be many ways to write such a factorization. For example if $A = B^\top B$ where $B \in \mathbb{R}^{k \times n}$, and Q is an orthonormal matrix in $\mathbb{R}^{k \times k}$, then also $A = B^\top B = B^\top Q^\top Q B = (QB)^\top (QB)$. Since there are infinitely many orthonormal matrices Q in $R^{k \times k}$, there are infinitely many factorizations of A as $B^\top B$ if there is one such factorization.
4. The 1×1 principal minors of A are 1 and 4. The 2×2 principal minor is $\det(A) = 0$. All of these values are non-negative. This matrix is not invertible.

Example 6.1.12. The matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ is PD.

1. The quadratic form
$$q_A(x,y) = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 2x^2 + 2y^2 + 2z^2 + 2xy + 2xz + 2yz = (x+y+z)^2 > 0$$
for all nonzero $(x,y,z) \in \mathbb{R}^3$.
2. The eigenvalues of A are 1, 1, and 4 which are all positive.
3. One can show that
$$A = \underbrace{\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}}_{B^\top} \underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}}_{B}$$
Note that B has linearly independent columns and is thus invertible.
4. The leading principal minors of A are 2, 3 and 4 which are all positive. The matrix A is invertible.

Example 6.1.13. We can construct any number of PSD matrices by choosing a $B \in \mathbb{R}^{k \times n}$ and taking $A = B^\top B$. If $\text{rank}(B) = k$, we know from homework that $\text{rank}(A) = \text{rank}(B^\top B) \leq k$.

In particular, the rank one PSD matrices of size $n \times n$ are exactly those of the form $\mathbf{v}\mathbf{v}^\top$ where $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{v} \neq \mathbf{0}$. Think about why this is true. Any matrix of the form $\mathbf{v}\mathbf{v}^\top$ is PSD and if $\mathbf{v} \neq \mathbf{0}$, then $\text{rank}(\mathbf{v}\mathbf{v}^\top) = 1$. Why is it that all rank one PSD matrices can be written this way? We'll revisit this later.

We now argue why the first three conditions in the definition of a PSD matrix are equivalent. A similar reasoning holds for PD matrices.

Proposition 6.1.14. *A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is PSD if and only if any of the following conditions hold:*

1. $\boldsymbol{x}^\top A\boldsymbol{x} \geq 0$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$*.*

2. All eigenvalues of A *are non-negative.*

3. There exists a matrix $B \in \mathbb{R}^{k\times n}$ *for some* k *such that* $A = B^\top B$*.*

Proof. 1. We first show that $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ if and only if all eigenvalues of A are non-negative. Suppose $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and $A\mathbf{u} = \lambda\mathbf{u}$. Then $\mathbf{u}^\top A\mathbf{u} = \mathbf{u}^\top\lambda\mathbf{u} = \lambda\mathbf{u}^\top\mathbf{u} = \lambda||\mathbf{u}||^2$. Therefore,

$$\mathbf{x}^\top A\mathbf{x} \geq 0\ \forall\mathbf{x} \in \mathbb{R}^n \ \Rightarrow\ \mathbf{u}^\top A\mathbf{u} \geq 0 \ \Rightarrow \lambda\|\mathbf{u}\|^2 \geq 0 \ \Rightarrow\ \lambda \geq 0$$

Now suppose $\lambda_i \geq 0$ for all i, and $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is an orthonormal set of eigenvectors of A with $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$. Since A is symmetric, we have that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ form a basis of $\mathbb{R}^n$. Now we get that

$$0 \leq \lambda_i\|\mathbf{u}_i\|^2 = \mathbf{u}_i^\top A\mathbf{u}_i \ \ \forall\ i$$

If $\mathbf{x} \in \mathbb{R}^n$, then $\mathbf{x} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$ for some constants $c_1, \ldots, c_n \in \mathbb{R}$. Computing $\mathbf{x}^\top A\mathbf{x}$ we get

$$\mathbf{x}^\top A\mathbf{x} = (\sum c_i\mathbf{u}_i^\top)A(\sum c_i\mathbf{u}_i) = \sum_{i=1}^{n} c_i^2\mathbf{u}_i^\top A\mathbf{u}_i + \sum_{i\neq j} c_ic_j\mathbf{u}_i^\top A\mathbf{u}_j = \sum c_i^2\mathbf{u}_i^\top A\mathbf{u}_i \geq 0.$$

The term $\sum_{i\neq j} c_ic_j\mathbf{u}_i^\top A\mathbf{u}_j = 0$ since each of its terms, $c_ic_j\mathbf{u}_i^\top A\mathbf{u}_j = c_ic_j\mathbf{u}_i^\top\lambda_j\mathbf{u}_j = c_ic_j\lambda_j\mathbf{u}_i^\top\mathbf{u}_j = 0$ because $\mathbf{u}_i^\top\mathbf{u}_j = 0$ whenever $i \neq j$.

2. Next, we show that if $A = B^\top B$ then $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Indeed,

$$A = B^\top B \ \Rightarrow\ \mathbf{x}^\top A\mathbf{x} = \mathbf{x}^\top B^\top B\mathbf{x} = (B\mathbf{x})^\top(B\mathbf{x}) = ||B\mathbf{x}||^2 \geq 0 \ \ \forall\mathbf{x}.$$

Conversely, suppose A is symmetric and $\mathbf{x}^\top A\mathbf{x} \geq 0$. Since A is symmetric it has an orthonormal diagonalization, $A = Q\Lambda Q^\top$. Since $\mathbf{x}^\top A\mathbf{x} \geq 0 \geq 0$, from the previous proof we know that all $\lambda_i \geq 0$, and so $\Lambda \geq 0$. Define

$$\sqrt{\Lambda} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix}$$

which is well defined since $\lambda_i \geq 0$ for all i, and set $B = \sqrt{\Lambda}Q^\top$. Since $\sqrt{\Lambda}$ is diagonal, it equals its transpose, and we get that

$$B^\top B = (\sqrt{\Lambda}Q^\top)^\top(\sqrt{\Lambda}Q^\top) = Q\sqrt{\Lambda}^\top\sqrt{\Lambda}Q^\top = Q\sqrt{\Lambda}\sqrt{\Lambda}Q^\top = Q\Lambda Q^\top = A.$$

□

6.2 Gram Matrices

Matrices of the form $B^\top B$ are called *Gram matrices.* Suppose $B \in \mathbb{R}^{k\times n}$ with columns $\mathbf{b}_1, \ldots, \mathbf{b}_n \in \mathbb{R}^k$. Then

$$B^\top B = \begin{bmatrix} \mathbf{b}_1^\top \\ \mathbf{b}_2^\top \\ \vdots \\ \mathbf{b}_n^\top \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^\top\mathbf{b}_1 & \mathbf{b}_1^\top\mathbf{b}_2 & \cdots & \mathbf{b}_1^\top\mathbf{b}_n \\ \mathbf{b}_2^\top\mathbf{b}_1 & \mathbf{b}_2^\top\mathbf{b}_2 & \cdots & \mathbf{b}_2^\top\mathbf{b}_n \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{b}_n^\top\mathbf{b}_1 & \cdots & \cdots & \mathbf{b}_n^\top\mathbf{b}_n \end{bmatrix} = \begin{bmatrix} ||\mathbf{b}_1||^2 & \mathbf{b}_1^\top\mathbf{b}_2 & \cdots & \mathbf{b}_1^\top\mathbf{b}_n \\ \mathbf{b}_2^\top\mathbf{b}_1 & ||\mathbf{b}_2||^2 & \cdots & \mathbf{b}_2^\top\mathbf{b}_n \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{b}_n^\top\mathbf{b}_1 & \cdots & \cdots & ||\mathbf{b}_n||^2 \end{bmatrix}$$

We saw that all Gram matrices are PSD and that all PSD matrices are Gram matrices. In particular, all PSD matrices have the special form you see in the last matrix of the above chain.

Lemma 6.2.1. *If the columns $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ of $B \in \mathbb{R}^{k \times n}$ are linearly independent, then $B^\top B$ is PD.*

Proof. Note that if $\mathbf{b}_1, \ldots, \mathbf{b}_n$ are linearly independent, then $k \geq n$ and $B\mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$. Therefore, if $\mathbf{x} \neq \mathbf{0}$, then $\|B\mathbf{x}\|^2 > 0$ which implies that $(B\mathbf{x})^\top (B\mathbf{x}) = \mathbf{x}^\top B^\top B\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$. Therefore, $B^\top B \succ 0$. □

This gives an alternate proof for why when the columns of B are linearly independent, $B^\top B$ is invertible since in this case, $B^\top B \succ 0$ which means that $\det(B^\top B) \neq 0$.

How about the converse of the above lemma which is one of the definitions of a PD matrix?

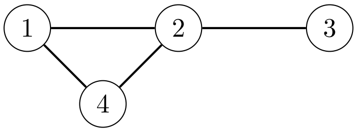
Lemma 6.2.2. *If $A \in \mathbb{R}^{n \times n}$ is PD then $A = B^\top B$ with columns of B linearly independent.*

Proof. Since A is symmetric we saw earlier that we can choose $B = \sqrt{\Lambda} Q^\top \in \mathbb{R}^{n \times n}$ where $A = Q\Lambda Q^\top$ is an orthonormal diagonalization of A. Since $A \succ 0$, all its eigenvalues are positive and so Λ is invertible and hence $\sqrt{\Lambda}$ is also invertible. This means that B is invertible since $\det(B) = \det(\sqrt{\Lambda}) \det(Q^\top) \neq 0$. We conclude that the columns of B are linearly independent. □

6.3 Application: Connectivity of a Graph

Definition 6.3.1. Let G be a graph with n vertices labeled $1, 2, \ldots, n$. The **degree of vertex** i, denoted d_i, is the number of edges incident to vertex i.

Example 6.3.2. In the following graph G, $d_1 = 2, d_2 = 3, d_3 = 1, d_4 = 2$.



Let $D_G \in \mathbb{R}^{n \times n}$ be the diagonal matrix whose i^{th} diagonal entry is d_i. For the graph G above we have

$$D_G = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

Recall that the adjacency matrix of a graph G is the symmetric matrix $A_G = (a_{ij}) \in \mathbb{R}^{n \times n}$ with

$$a_{ij} = \begin{cases} 1 & \text{if } ij \text{ is an edge in } G \\ 0 & \text{otherwise} \end{cases}$$

We can now define the Laplacian of a graph G which is a very important PSD matrix associated to G.

Definition 6.3.3. Let G be a graph with n vertices, diagonal matrix of vertex degrees D_G as above, and adjacency matrix A_G. The **Laplacian** of G is

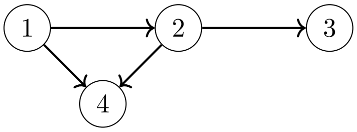
$$L_G = D_G - A_G \in \mathbb{R}^{n \times n}.$$

In our example graph,

$$A_G = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad L_G = D_G - A_G = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix}$$

The Laplacian L_G holds a tremendous amount of information about G. Here are some facts:

- $\lambda = 0$ is always an eigenvalue of L_G with eigenvector $\mathbf{1}$: Note that each row of L_G sums to 0. Therefore, $L_G\mathbf{1} = \mathbf{0}$.
- L_G is always positive semidefinite: To prove this we argue that $L_G = B_G B_G^\top$ where B_G is the **"directed" node-edge incidence matrix** of G computed as follows:
    - Label the rows of B_G by the vertices of the graph G.
    - For each edge of G, pick a direction. If the edge has end points i and j, then label the edge ij if the arrow on the edge points from i to j. In our example, the directions shown below creates the edge labels 12, 14, 23, 24.



    - Label the columns of B_G by the edge labels. Now each entry in B_G has row label k (from the kth vertex) and column label ij (from edge ij directed from i to j). The value of this entry is set to

$$b_{k,ij} = \begin{cases} 1 & k = i \\ -1 & k = j \\ 0 & \text{otherwise} \end{cases}$$

With our graph G, the rows of B_G have labels $1, 2, 3, 4$ and the columns have labels $12, 14, 23, 24$ and

$$B_G = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \end{bmatrix}$$

Put differently, in the column indexed by directed edge ij, we put a 1 in row i and -1 in row j. Now check for yourself that

$$L_G = B_G B_G^\top$$

In our example,

$$L_G = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \end{bmatrix}}_{B_G} \underbrace{\begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}}_{B_G^\top}$$

Therefore , L_G is PSD which means that all its eigenvalues are nonnegative.

Let us summarize our discussion so far.

Lemma 6.3.4. *The Laplacian L_G of a graph G is positive semidefinite and its eigenvalues are*

$$0 = \lambda_1 \le \lambda_2 \le \lambda_3 \le \cdots \le \lambda_n$$

The all-ones vector $\mathbf{1}$ is an eigenvector of L_G with eigenvalue 0.

Since L_G is PSD, the quadratic from $\mathbf{x}^\top L_G \mathbf{x} \ge 0$ for all $\mathbf{x} \in \mathbb{R}^n$. This quadratic form can be understood explicitly.

Proposition 6.3.5. *Let $G = ([n], E)$ be a graph with n vertices and edge set E. If $\boldsymbol{x} = (x_1, \dots, x_n)^\top$,*

$$\boldsymbol{x}^\top L_G \boldsymbol{x} = \sum_{\{ij\} \in E} (x_i - x_j)^2.$$

Proof. Since $L_G = B_G B_G^\top$, $\mathbf{x}^\top L_G \mathbf{x} = \mathbf{x}^\top B_G B_G^\top \mathbf{x}$. Consider $\mathbf{x}^\top B_G$. The column of B_G indexed by edge ij has 1 in position i and -1 in position j. Therefore the dot product of $\mathbf{x}^\top$ with that column is $x_i - x_j$, and the vector $\mathbf{x}^\top B_G$ which is indexed by the edges of G has entry $x_i - x_j$ in position ij. The vector $B_G^\top \mathbf{x}$ is the transpose of $\mathbf{x}^\top B_G$ and so $\mathbf{x}^\top B_G B_G^\top \mathbf{x}$ is $\sum_{ij \in E} (x_i - x_j)^2$. □

Check this claim in our example:

$$\mathbf{x}^\top L_G \mathbf{x} = (\mathbf{x}^\top B_G)(B_G^\top \mathbf{x}) = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \end{bmatrix}}_{B_G} \underbrace{\begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}}_{B_G^\top} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

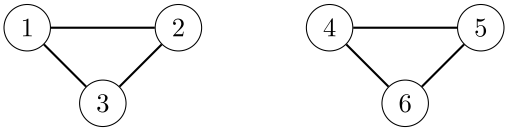
$$= \begin{bmatrix} (x_1 - x_2) & (x_1 - x_4) & (x_2 - x_3) & (x_2 - x_4) \end{bmatrix} \begin{bmatrix} (x_1 - x_2) \\ (x_1 - x_4) \\ (x_2 - x_3) \\ (x_2 - x_4) \end{bmatrix}$$

$$= (x_1 - x_2)^2 + (x_1 - x_4)^2 + (x_2 - x_3)^2 + (x_2 - x_4)^2 = \sum_{\{ij\} \in E} (x_i - x_j)^2$$

We now come to our main application. A graph G is **connected** if there is a way to travel from any vertex in G to any other vertex in G by moving along a sequence of edges in G. The graph in our example is connected. Knowing the connectivity of a graph has many applications. For example, suppose G is a road network where vertices are cities and edges are roads. Then if G is disconnected, there will be a clump of cities that cannot be accessed by road from other cities. For another example, it might be that G is a communication network that is connected at the start, but due to an earthquake becomes disconnected, isolating groups of people. Deciding the connectivity of a graph is one of the most basic questions about a graph. It turns out that the second eigenvalue of L_G decides if G is connected.

Theorem 6.3.6. *Let $G = ([n], E)$ be a graph with Laplacian matrix L_G whose eigenvalues are $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n$. Then G is connected if and only if $\lambda_2 > 0$.*

Before we prove this theorem, let us do an example. Consider the following disconnected graph G:



The Laplacian L_G breaks up into blocks, with each block corresponding to a connected part of G and the zero blocks coming from the lack of connectivity between the two connected components of G.

$$L_G = \left[\begin{array}{ccc|ccc} 2 & -1 & -1 & & & \\ -1 & 2 & -1 & & \mathbf{0} & \\ -1 & -1 & 2 & & & \\ \hline & & & 2 & -1 & -1 \\ & \mathbf{0} & & -1 & 2 & -1 \\ & & & -1 & -1 & 2 \end{array}\right]$$

Now we can explicitly find two linearly independent eigenvectors of L_G with eigenvalue 0 by leveraging the fact that the individual Laplacians of the connected subgraphs of G have $\mathbf{1}$ as an eigenvector of eigenvalue zero. In our example, check that the following two vectors are linearly independent eigenvectors of 0:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

This tells us that $\mathrm{GM}(0) \geq \mathrm{AM}(0) \geq 2$, and hence $0 = \lambda_1 = \lambda_2$.

This idea works in general whenever G is disconnected. Suppose G has $k \geq 2$ connected components. Then L_G will break into k diagonal blocks, one for each connected component of G, and by stacking 0s above and below the all-ones eigenvector of each component, we can build k linearly independent eigenvectors of L_G with eigenvalue 0. This means that $\mathrm{GM}(0) \geq k$ and hence $\lambda_2 = 0$. Let us write a proof of the converse, namely that if G is connected then $\lambda_2 > 0$.

Proof. We will show that if G is connected then $\mathrm{AM}(0) = 1$. Since L_G is symmetric, $\mathrm{GM}(0) = \mathrm{AM}(0)$ and we can conclude that $\lambda_2 \neq \lambda_1 = 0$. Since L_G is PSD, all its eigenvalues are nonnegative and we conclude that $\lambda_2 > 0$.

Let $\mathbf{u}$ be an eigenvector of L_G with eigenvalue 0. Then $\mathbf{u} \neq \mathbf{0}$. The fact that $L_G\mathbf{u} = \mathbf{0}$ implies that

$$0 = \mathbf{u}^\top L_G \mathbf{u} = \sum_{\{ij\} \in E} (u_i - u_j)^2.$$

Since each summand, $(u_i - u_j)^2$ is non-negative, it must be true that $u_i = u_j$ for all $ij \in E$. Since G is connected, there is a way to walk from vertex 1 to any other vertex j along edges of G. Therefore, if $u_1 = c$ then it must be that $u_k = c$ for all vertices k in the path from 1 to j. This argument shows that $\mathbf{u} = c\mathbf{1}$, a multiple of $\mathbf{1}$. If $u_1 = 0$ then by the same argument $\mathbf{u} = \mathbf{0}$ which is a contradiction. Therefore, $u_1 \neq 0$ and we conclude that every eigenvector of 0 is a multiple of $\mathbf{1}$. Hence $\mathrm{AM}(0) = \mathrm{GM}(0) = 1$ and so $\lambda_2 > 0$. □

The second eigenvalue λ_2 of L_G is known as the **spectral gap** of G. It is also called the *Fiedler value* of G and measures the connectivity of the graph G. The larger λ_2 is, the more connected G is. The spectral gap is extremely useful in clustering algorithms which we will see next. For a very nice article on graph connectivity and clustering take a look at https://towardsdatascience.com/spectral-clustering-aba2640c0d5b

6.4 Application: Spectral Clustering

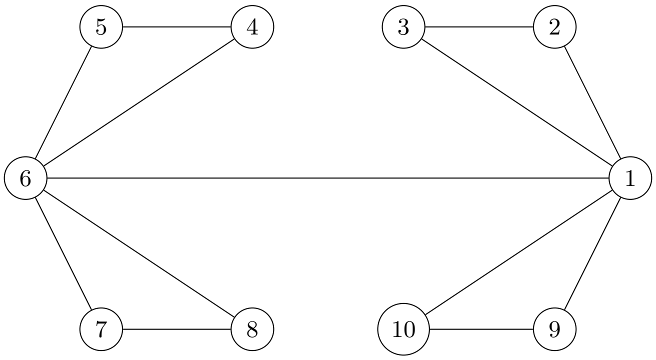
This is Part 1 of spectral clustering which is to be followed by a star problem in homework that completes it. This part used to be a homework problem and is left as a series of exercises to be discussed in class.

Recall the color conventions from star problems: (i) problem statements, (ii) math formulations, (iii) commentary/philosophy, (iv) running example (nothing asked in red needs to be written up), (v) ♠ action items.

The second smallest eigenvalue λ_2 of the Laplacian L_G of a graph G is called the *Fiedler value* of G, and an eigenvector $\mathbf{w}$ of λ_2 is called a *Fiedler vector* of G. We saw that G is connected if and only if $\lambda_2 > 0$. Here is an application of the Fiedler vector $\mathbf{w}$.

An important task in data science is to find *clusters* in a graph. By a cluster we mean a group of vertices that are relatively well connected amongst themselves but not so well connected to the rest of the graph. For example, suppose G is a social network graph with people as vertices and an edge between two people who know each other. Then some natural clusters might be all people who belong to the same church, or soccer club, or do Tai Chi. Someone from church may know someone who does Tai Chi, but perhaps there are only a few such pairs. Knowing clusters in graphs allows one to understand how information or infection might spread in the network. Advertisers use cluster information to target similar ads to people in a given cluster. If you bought a particular knee brace for soccer, then chances are that your soccer friends might also buy it, where as your church friends may not. In this exercise we will use linear algebra to find clusters in a graph.

Running example: The graph below is a reproduction of the example from `https://towardsdatascience.com/spectral-clustering-aba2640c0d5b` with vertices relabeled as $1, \ldots, 10$.



This graph has two obvious clusters in it, the cluster of vertices $\{4, 5, 6, 7, 8\}$ and the cluster of vertices $\{1, 2, 3, 9, 10\}$. There is only one edge between these two groups while vertices in a group have more connections among them.

How do we find clusters in large complicated graphs? To talk about clusters, we use the math terminology of *cuts* in graphs.

Definition 6.4.1. Let G be a graph with vertex set $V = \{1, 2, 3, \ldots, n\}$ and edge set E consisting of pairs of vertices $\{i, j\}$. If $A \subseteq V$ is a subset of vertices, we use the notation $V \backslash A$ for the vertices in V that are not in A. Also, $|A|$ denotes the cardinality of the set A, which is the number of elements in A.

In our example, $V = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and $E = \{\{1, 2\}, \{1, 3\}, \{2, 3\} \ldots\}$. If $A = \{4, 5, 6, 7, 8\}$, then $|A| = 5$, $V \backslash A = \{1, 2, 3, 9, 10\}$ and $|V \backslash A| = 5$.

1. A **cut** in G is a partition of V into two sets A and $V \backslash A$ for some subset of vertices $A \subset V$. This is sometimes called the *cut induced by* A.

2. Let $E(A, V \backslash A)$ be the edges that go between the vertices in A and $V \backslash A$. These are the edges holding A and $V \backslash A$ together in G. In our example, the cut induced by $A = \{4, 5, 6, 7, 8\}$ has $E(A, V \backslash A) = \{\{1, 6\}\}$.

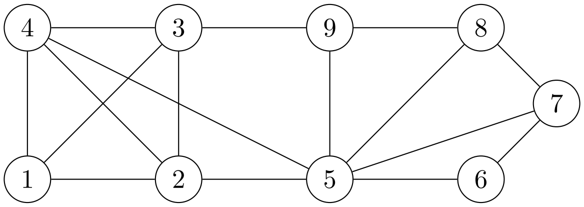
   Then the **density** of the cut induced by A is

   $$\phi(A, V \backslash A) = n \cdot \frac{|E(A, V \backslash A)|}{|A| \cdot |V \backslash A|}$$

   For $A = \{1, 3, 4, 5, 6, 7\}$, $E(A, V \backslash A) = \{\{1, 2\}, \{2, 3\}, \{1, 9\}, \{1, 10\}, \{6, 8\}, \{7, 8\}\}$ and $\phi(A, V \backslash A) = 10 \cdot \frac{6}{6 \cdot 4} = \frac{5}{2}$. Note that there are 6 edges between A and $V \backslash A$ but there could have been $6 \cdot 4 = 24$ between A and $V \backslash A$ (which would have been the case if G was the complete graph K_{10}). So $\frac{6}{6 \cdot 4}$ is the ratio of the number of edges connecting A and $V \backslash A$ in this graph and in the complete graph. The cut density is the product of this ratio and the total number of vertices, 10.

3. Let ϕ_G denote the smallest possible density of a cut in G. We call a cut with density ϕ_G, the *sparsest cut* in G. What is the density of the cut induced by $A = \{4, 5, 6, 7, 8\}$? Do you think there is a sparser cut in our example graph?

The questions below will be based on the following graph H:



Q1 ♠ Calculate the density of the cut in H induced by $A = \{1, 2, 3, 4\}$.

Below is an algorithm that uses the Fiedler vector $\mathbf{w}$ to break a graph into its two main clusters. Since it uses eigenvalues and vectors, the method is called *spectral clustering*.

(a) Sort the components of $\mathbf{w}$ in descending order so that $w_{i_1} \geq w_{i_2} \geq \cdots \geq w_{i_n}$. In our running example, Julia says that $\lambda_2 = 0.2984$ and the Fiedler vector

$$\mathbf{w} = (0.23, 0.33, 0.33, -0.33, -0.33, -0.23, -0.33, -0.33, 0.33, 0.33).$$

One way to sort the components of $\mathbf{w}$ is as

$$w_2 \geq w_3 \geq w_9 \geq w_{10} \geq w_1 \geq w_6 \geq w_4 \geq w_5 \geq w_7 \geq w_8.$$

In our notation, $i_1 = 2, i_2 = 3, i_4 = 9 \ldots$. When two components tie in value, you can break the tie as you wish.

(b) Let $A_k := \{i_1, \ldots, i_k\}$ for $k = 1, \ldots, n-1$. Among the cuts $(A_k, V \backslash A_k)$, output the one with the smallest density. In our example, $A_1 = \{2\}, A_2 = \{2, 3\}, A_3 = \{2, 3, 9\}, A_4 = \{2, 3, 9, 10\}$ etc. The cut $(A_4, V \backslash A_4)$ has density $10\frac{4}{4\cdot 6} = \frac{5}{3}$. For practice, find A_5 and its cut density.

Q2 (a) ♠ Calculate the Fiedler value λ_2 and its eigenvector $\mathbf{w}$ for H using Julia. (I would suggest just finding the Laplacian of H and typing it directly into Julia. Coding graphs into Julia is more involved).

(b) ♠ Sort the components of $\mathbf{w}$ and find all the sets $A_1, \ldots, A_8$.

(c) ♠ Pick one A_i from your list such that the density of the cut it induces has a chance to be sparser than the cut induced by $\{1, 2, 3, 4\}$. Compute the density of this cut.

The following theorem says how well the above algorithm can do.

Theorem 6.4.2. *The following hold for G:*

1. $\phi_G \geq \lambda_2$.

2. *The above algorithm always finds a cut of density at most $4\sqrt{d_G \lambda_2}$ where d_G is the largest degree of a vertex in G.*

Q3 (a) ♠ Argue that this theorem is saying that $\lambda_2 \leq \phi_G \leq 4\sqrt{d_G \lambda_2}$.
(**Hint**: The sparsest cut output by the algorithm may not be the overall sparsest cut in G.)

(b) ♠ What are these bounds for ϕ_G in H?

Call a vector $\mathbf{x}$ *non-constant* if it is not a multiple of $\mathbf{1} = (1, 1, \ldots, 1)$. For $\mathbf{x} = (x_1, \ldots, x_n)$, define the function

$$Q(\mathbf{x}) = n \cdot \frac{\sum_{\{i,j\} \in E(G)} (x_i - x_j)^2}{\sum_{1 \le i < j \le n} (x_i - x_j)^2}$$

Note that the numerator is $\mathbf{x}^\top L_G \mathbf{x}$ and the denominator is the same quadratic function for the complete graph on n vertices, i.e., $\mathbf{x}^\top L_{K_n} \mathbf{x}$.

For a subset $A \subseteq V$, let $\mathbf{c}_A$ be the vector in $\mathbb{R}^n$ with ith coordinate equal to 1 if $i \in A$ and 0 otherwise. This is called the *characteristic vector* of A. In our running example, if we take $A = \{4, 5, 6, 7, 8\}$, then $\mathbf{c}_A = (0, 0, 0, 1, 1, 1, 1, 1, 0, 0)$, and $Q(\mathbf{c}_A) = \frac{2}{5} = \phi(A, V \backslash A)$.

Q4 ♠ Compute the characteristic vector $\mathbf{c}_A$ for $A = \{1, 2, 3, 4\}$ in H, and check that for this A, $Q(\mathbf{c}_A)$ is exactly the density of the cut induced by A.

Q5 (a) ♠ Argue that ϕ_G is the minimum of $Q(\mathbf{x})$ as $\mathbf{x}$ varies over all $\mathbf{c}_A$.

(b) ♠ If you were to use this method to find ϕ_G how many $\mathbf{c}_A$'s would you need to compute in the graph G shown above? Hopefully you see that this is not a very practical way to find ϕ_G.

The star problem in homework is to prove part 1 of the theorem.

6.5 Application: Distance Realization

We end the chapter with one last application of positive semidefinite matrices.

Question 6.5.1. Are there three points $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathbb{R}^2$ such that

$$||\mathbf{p} - \mathbf{q}|| = ||\mathbf{q} - \mathbf{r}|| = 1 \quad \text{and} \quad ||\mathbf{p} - \mathbf{r}|| = 3?$$

The answer is no! The reason is because of the triangle inequality which creates the contradiction

$$||\mathbf{p} - \mathbf{r}|| \le ||\mathbf{p} - \mathbf{q}|| + ||\mathbf{q} - \mathbf{r}|| \implies 3 \le 2$$

For three distances, the triangle inequality is the only restriction, i.e., given $x, y, z \ge 0$ such that

$$x \le y + z, \quad y \le x + z, \quad z \le x + y$$

there are always three points $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathbb{R}^2$ such that $\|\mathbf{p} - \mathbf{q}\| = x, \|p - \mathbf{r}\| = y, \|\mathbf{q} - \mathbf{r}\| = z$.

We could investigate the same sort of question in higher dimensions. For example, are there four points $\mathbf{p}, \mathbf{q}, \mathbf{r}, \mathbf{s} \in \mathbb{R}^3$ with the distances shown in Figure 6.5? The answer is no! However, the triangle inequality presents no problems this time.

The following theorem, known as Schoenberg's theorem, allows us to answer the question in general.

Theorem 6.5.2. **(Schoenberg 1935**) *Given "distances" $d_{ij} \ge 0$ for $i, j = 0, 1, \ldots, n$ and $d_{ii} = 0$ for all i, there exist points $\boldsymbol{p}_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_n \in \mathbb{R}^n$ with $||\boldsymbol{p}_i - \boldsymbol{p}_j|| = d_{ij}$ if and only if the matrix $M = (M_{ij}) \in \mathbb{R}^{n \times n}$ with $M_{ij} = \frac{1}{2}(d_{i0}^2 + d_{0j}^2 - d_{ij}^2)$ is positive semidefinite. If such points $\boldsymbol{p}_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ exist we say that the given distances d_{ij} are* ***realizable****.*

Before proving the theorem, we illustrate its use on the above example with four distances. Let $D \in \mathbb{R}^{(n+1)\times(n+1)}$ be the matrix of distances. The rows and columns of D are indexed $0, 1, \ldots, n$. In our four point example, the distance matrix is

$$D = \begin{bmatrix} 0 & 2 & 3 & 2 \\ 2 & 0 & 2 & 3 \\ 3 & 2 & 0 & 2 \\ 2 & 3 & 2 & 0 \end{bmatrix}$$

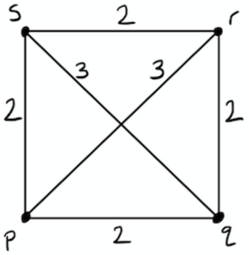


Figure 6.5: Four distances

The matrix $M = (M_{ij})$ in the theorem lives in $\mathbb{R}^{n\times n}$ and has $M_{ij} = \frac{1}{2}(d_{i0}^2 + d_{0j}^2 - d_{ij}^2)$. In our example,

$$M = \frac{1}{2}\begin{bmatrix} 8 & 9 & -1 \\ 9 & 18 & 9 \\ -1 & 9 & 8 \end{bmatrix}$$

A quick computation shows that $\det(M) < 0$ and hence M is not PSD. By Schoenberg's theorem, the given distances are not realizable.

Now let's prove the theorem. We will need to use the cosine theorem from precalculus which says that the length of one side of a triangle can be expressed in terms of the lengths of the other two sides and the cosine of the angle between them. We state that fact using dot products. Please think about why this is the same statement.

Lemma 6.5.3. *Given* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$||\boldsymbol{x} - \boldsymbol{y}||^2 = ||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2 - 2\boldsymbol{x}^\top \boldsymbol{y}$$

Alternatively,

$$\boldsymbol{x}^\top \boldsymbol{y} = \frac{1}{2}(||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2 - ||\boldsymbol{x} - \boldsymbol{y}||^2)$$

Proof. (**of Schoenberg's Theorem**) We first assume that the distances are realizable and aim to show that $M \succeq 0$. If the distances are realizable then there exist points $\mathbf{p}_0, \mathbf{p}_2, \ldots, \mathbf{p}_n \in \mathbb{R}^n$ such that

$$||\mathbf{p}_i - \mathbf{p}_j|| = d_{ij} \;\; \forall \;\; 0 \leq i, j \leq n$$

Since translating the whole configuration of points does not affect the distances among them, we may assume that $\mathbf{p}_0 = \mathbf{0}$. Next, set $\mathbf{x}_i = \mathbf{p}_i - \mathbf{p}_0$ for $i = 1, \ldots, n$. This implies that $||\mathbf{x}_i|| = d_{i0} = d_{0i}$, $||\mathbf{x}_j|| = d_{0j} = d_{j0}$, and $d_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||$. By the cosine theorem, we have

$$\mathbf{x}_i^\top \mathbf{x}_j = \frac{1}{2}(||\mathbf{x}_i||^2 + ||\mathbf{x}_j||^2 - ||\mathbf{x}_i - \mathbf{x}_j||^2) = \frac{1}{2}(d_{i0}^2 + d_{0j}^2 - d_{ij}^2) = M_{ij}$$

This means that

$$M = (M_{ij}) = (\mathbf{x}_i^\top \mathbf{x}_j) = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} = B^\top B$$

which is a Gram matrix. Therefore, M is positive semidefinite.

The other direction follows similarly. If M is positive semidefinite, then $M = B^\top B$ for some matrix $B \in \mathbb{R}^{k\times n}$. Set $\mathbf{p}_0 = \mathbf{0}$ and $\mathbf{p}_i = \mathbf{b}_i$, the i^{th} column of B, for $i = 1, \ldots, n$. Then we have $M_{ij} = \mathbf{p}_i^\top \mathbf{p}_j$. In particular, $M_{ii} = d_{0i}^2 = \|\mathbf{p}_i\|^2$ and so $\|\mathbf{p}_i\| = d_{0i}$ for all i. Then using the cosine theorem,

$$\frac{1}{2}(d_{i0}^2 + d_{0j}^2 - d_{ij}^2) = M_{ij} = \mathbf{p}_i^\top \mathbf{p}_j = \frac{1}{2}(||\mathbf{p}_i||^2 + ||\mathbf{p}_j||^2 - ||\mathbf{p}_i - \mathbf{p}_j||^2) = \frac{1}{2}(d_{i0}^2 + d_{0j}^2 - ||\mathbf{p}_i - \mathbf{p}_j||^2)$$

and we conclude that $\|\mathbf{p}_i - \mathbf{p}_j\| = d_{ij}$ for all i, j. Thus the points $\mathbf{p}_0, \ldots, \mathbf{p}_n$ realize the distances d_{ij}.

Note that this direction tells you how to find the points $\mathbf{p}_i$ and also that they live in $\mathbb{R}^k$ where k is the number of rows in B. To find the smallest dimension in which the points can live, we need to minimize k for which there is a $B \in \mathbb{R}^{k \times n}$ such that $M = B^\top B$. This is the rank of M. So not only does M tell you if the distances are realizable, it also tells you the smallest dimension in which you can find the realization. By factorizing M you also get the points. Very clever!! □

Schoenberg's theorem is part of an area of mathematics called *distance geometry*. There are many beautiful theorems about distances which have a huge number of real world applications. See the article *Six mathematical gems from the history of distance geometry* by Leo Liberti and Carlile Lavor for example: https://onlinelibrary.wiley.com/doi/full/10.1111/itor.12170.

Distance geometry is used in chemistry and biology to model molecules. You might obtain the distances between pairs of atoms in a molecule using NMR (nulear magnetic resonance) which is like an MRI, and your job is to reconstruct a three-dimensional model of how the atoms are configured in the molecule. See for example Figures 6 and 8 in the article *Distance Geometry: Theory, Algorithms and Chemical Applications* by Timothy Havel http://web.mit.edu/tfhavel/www/Public/dg-review.pdf. This sort of mathematics also underlies the very important problem of protein folding. Check out the webpage of the Baker Lab at UW for more on protein folding.

Chapter 7

Polynomial Vector Spaces

In this chapter we will see three vector spaces that are filled with polynomials, and applications of such spaces. This might be the first time you are seeing vector spaces different from $\mathbb{R}^n$, but there are many such examples, and we will see a few more in the rest of the lectures.

7.1 Univariate Polynomials

A univariate polynomial over $\mathbb{R}$ is a polynomial in one variable with coefficients in $\mathbb{R}$. As a running example in this chapter, consider

$$p(x) = x^5 - 20x^4 + 8x^2 - 10.$$

The *monomials* in x are the pure powers of x, namely, $1 = x^0, x, x^2, x^3, x^4, \ldots$ and every univariate polynomial in x is a linear combination over $\mathbb{R}$ of finitely many monomials in x. The polynomial $p(x)$ has *degree* 5, the highest power of x in $p(x)$. The *terms* of $p(x)$ are $x^5, -20x^4, 8x^2$ and -10. Each term is the product of a monomial in x with a coefficient that is a real number. The *leading term* of $p(x)$ is x^5 and its *leading coefficient* is 1. Note that there is no x^3 term and no x term in $p(x)$. The coefficients of $p(x)$ are the coefficients of all terms in $p(x)$ of degree at most 5 including the missing terms. In this example, the coefficients of $p(x)$ are $-10, 0, 8, 0, -20, 1$ written in increasing order of the degree of the monomials they appear with. If we know apriori that $p(x)$ has degree 5, then we can uniquely represent $p(x)$ by its *coefficient vector* $(-10, 0, 8, 0, -20, 1) \in \mathbb{R}^6$.

Definition 7.1.1. The general univariate polynomial over $\mathbb{R}$ of degree d is

$$f(x) = a_d x^d + a_{d-1}x^{d-1} + a_{d-2}x^{d-2} + \cdots + a_1 x + a_0$$

where the coefficients $a_0, a_1, \ldots, a_{d-1}, a_d$ are real numbers and $a_d \neq 0$. It can be uniquely represented by its coefficient vector $(a_0, a_1, a_2, \ldots, a_d) \in \mathbb{R}^{d+1}$.

7.1.1 The infinite-dimensional vector space $\mathbb{R}[x]$

The set of all polynomials in x with real coefficients is denoted as $\mathbb{R}[x]$ and called the *univariate polynomial ring* over $\mathbb{R}$. Let's quickly check that $\mathbb{R}[x]$ is a vector space. If

$$f(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0 \quad \text{and} \quad g(x) = b_t x^t + b_{t-1}x^{t-1} + \cdots + b_1 x + b_0$$

are two polynomials of degrees d and t respectively, then

$$f(x) + g(x) = a_d x^d + a_{d-1}x^{d-1} + a_{d-2}x^{d-2} + \cdots + a_1 x + a_0 + b_t x^t + b_{t-1}x^{t-1} + b_{t-2}x^{t-2} + \cdots + b_1 x + b_0$$

is a polynomial in $\mathbb{R}[x]$ of degree $\max(d,t)$. You can rewrite this sum so that the terms have decreasing degrees if you wish. Also, if $\gamma \in \mathbb{R}$ then

$$\gamma f(x) = \gamma a_d x^d + \gamma a_{d-1}x^{d-1} + \gamma a_{d-2}x^{d-2} + \cdots + \gamma a_1 x + \gamma a_0$$

which is again a polynomial in $\mathbb{R}[x]$ of degree d, unless $\gamma = 0$ in which case, $\gamma f(x) = 0$. Therefore, $\mathbb{R}[x]$ is a vector space over $\mathbb{R}$. A basis of $\mathbb{R}[x]$ is

$$\mathcal{B} = \{1 = x^0, x, x^2, x^3, \ldots, \}$$

called the *monomial basis* of $\mathbb{R}[x]$. Any polynomial is a linear combination of finitely many elements of $\mathcal{B}$ and hence $R[x] = \text{Span}(\mathcal{B})$. Check that the elements of $\mathcal{B}$ are linearly independent. Indeed the only way that a linear combination of monomials is 0 is if all scalars in the combination are 0. This also means that $R[x]$ is an infinite dimensional vector space since $\mathcal{B}$ has infinite size. This might be your first encounter with an infinite-dimensional vector space, so pause to make sure you agree with everything so far.

7.1.2 The finite-dimensional vector space $\mathbb{R}[x]_{\leq d}$

Now suppose we restrict the degree of the polynomials we consider. For example, suppose we only want to look at polynomials of degree at most five, i.e., $d \leq 5$. Then we get a subset of $\mathbb{R}[x]$ denoted as $\mathbb{R}[x]_{\leq d}$. Please check that this is again a vector space since the sum of any two polynomials of degree at most 5 is again a polynomial of degree at most 5 and scaling a polynomial of degree at most 5 creates another polynomial of degree at most 5. Therefore, $\mathbb{R}[x]_{\leq 5}$ is a subspace of $\mathbb{R}[x]$. Further, any polynomial of degree at most 5 is a linear combination of $1, x, x^2, x^3, x^5$, and since $\mathcal{B}_5 := \{1, x, x^2, x^3, x^5\}$ is a subset of $\mathcal{B}$, its elements are linearly independent. Therefore, $\mathcal{B}_5$ is a basis of $\mathbb{R}[x]_{\leq 5}$ and $\dim(\mathbb{R}[x]_{\leq 5}) = 6$. More generally,

Proposition 7.1.2. *For any positive integer d, let $\mathbb{R}[x]_{\leq d}$ be the set of all univariate polynomials of degree at most d. Then $\mathbb{R}[x]_{\leq d}$ is a subspace of $\mathbb{R}[x]$. It has the monomial basis $\mathcal{B}_d := \{1, x, x^2, \ldots, x^d\}$ and* $\dim(\mathbb{R}[x]_{\leq d}) = d + 1$.

As we noted before, there is a bijection between $\mathbb{R}[x]_{\leq d}$ and $\mathbb{R}^{d+1}$ obtained by identifying a polynomial with its coefficient vector. Check that under this identification, addition and scalar multiplication in $\mathbb{R}[x]_{\leq d}$ are the same as addition and scalar multiplication of coefficient vectors. This identification will be very useful when we consider applications. Since $\mathbb{R}[x]$ and $\mathbb{R}[x]_{\leq d}$ are vector spaces they admit linear transformations. You will see in homework that the familiar operation of differentiation is one such linear transformation.

7.1.3 Polynomial Interpolation

Question: Given m points $t_1, t_2, \ldots, t_m \in \mathbb{R}$, is there a polynomial $f(x)$ of degree d that has values $\beta_1, \beta_2, \ldots, \beta_m$ at $t_1, t_2, \ldots, t_m$?

Mathematically, we are looking for coefficients $a_d, a_{d-1}, \ldots, a_0 \in \mathbb{R}$ with $a_d \neq 0$ such that the polynomial $f(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0$ has the evaluations

$$f(t_1) = \beta_1, f(t_2) = \beta_2, \ldots, f(t_m) = \beta_m$$

This question arises in many experimental situations where at time t_i we make a measurement β_i and we want to know if there is a polynomial f of some specified degree d that fits our data in the sense that $f(t_i) = \beta_i$. If such a polynomial exists then you might be able to use it to "explain" the behavior of the phenomenon you are studying at all time points, not just the times at which you made measurements. We'll now use linear algebra to answer the above question.

Here is an algorithm to solve this problem.

1. Set $f(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0$ where $a_d, a_{d-1}, \ldots, a_0$ are unknown (they are variables).
2. Evaluating $f(x)$ at each t_i we get the equations:

$$\begin{aligned} a_d t_1^d + a_{d-1}t_1^{d-1} + \cdots + a_1 t_1 + a_0 =& \beta_1 \\ a_d t_2^d + a_{d-1}t_2^{d-1} + \cdots + a_1 t_2 + a_0 =& \beta_2 \\ a_d t_3^d + a_{d-1}t_3^{d-1} + \cdots + a_1 t_3 + a_0 =& \beta_3 \\ &\vdots \\ a_d t_m^d + a_{d-1}t_m^{d-1} + \cdots + a_1 t_m + a_0 =& \beta_m \end{aligned}$$

3. Express the above equations as $M\mathbf{f} = \mathbf{b}$ where $\mathbf{f}$ is the unknown vector of coefficients of $f(x)$.

$$\underbrace{\begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{d-1} & t_1^d \\ 1 & t_2 & t_2^2 & \cdots & t_2^{d-1} & t_2^d \\ 1 & t_3 & t_3^2 & \cdots & t_3^{d-1} & t_3^d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & t_m & t_m^2 & \cdots & t_m^{d-1} & t_m^d \end{bmatrix}}_{M} \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{d-1} \\ a_d \end{pmatrix}}_{\mathbf{f}} = \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_m \end{pmatrix}}_{\mathbf{b}}$$

4. Note that the matrix M and the vector $\mathbf{b}$ are filled with known numbers. There is a polynomial $f(x)$ with $f(t_i) = \beta_i$ if and only if the system $M\mathbf{f} = \mathbf{b}$ has a solution. We also need this solution to have $a_d \neq 0$ if we want $f(x)$ to have degree d. Solutions with $a_d = 0$ give polynomials of lower degree that have $f(t_i) = \beta_i$.

Comments:

1. The matrix M, which has the very special structure you see above, is called a $m \times (d+1)$ *Vandermonde matrix*. You construct a $m \times (d+1)$ Vandermonde matrix by picking m numbers $t_1, \ldots, t_m$ and making the ith row of M equal to $(1, t_i, t_i^2, \ldots, t_i^d)$.
2. If $m < d+1$ then the system $M\mathbf{f} = \mathbf{b}$ might have infinitely many solutions and if $m > d+1$ it may have no solutions.
3. There is a unique polynomial $f(x)$ with $f(t_i) = \beta_i$ if and only if M is a square invertible matrix. This requires $m = d+1$, i.e., we need values of f at $m = d+1$ points in $\mathbb{R}$.
4. When M is a square $k \times k$ Vandermonde matrix, $\det(M) = (-1)^k \prod_{i<j}(t_i - t_j)$. This is not obvious, but means that M is invertible if and only if $t_i \neq t_j$ for any pair i, j. In other words, we need to measure at distinct times in our experiment.
5. Note that we have identified the polynomial $f(x)$ with its coefficient vector $\mathbf{f}$ in the above calculation. This identification allows polynomials to be used in linear algebra.

In homework we will develop polynomial interpolation further.

7.2 Solving a Univariate Polynomial Equation

Recall that all the values of x for which $p(x) = 0$ are called the *roots* of $p(x)$. By the Fundamental Theorem of Algebra, if degree($p(x)$) $= d$ then $p(x)$ has d roots, some of which may be complex or might occur more than once. For example, the polynomial $f(x) = x^3 - 7x - 6 = (x+2)(x+1)(x-3)$ has three real roots, namely $x = -2, -1, 3$. The graph of $f(x)$ can be seen in Figure 7.1. The graph crosses the x-axis at the roots of the polynomial.

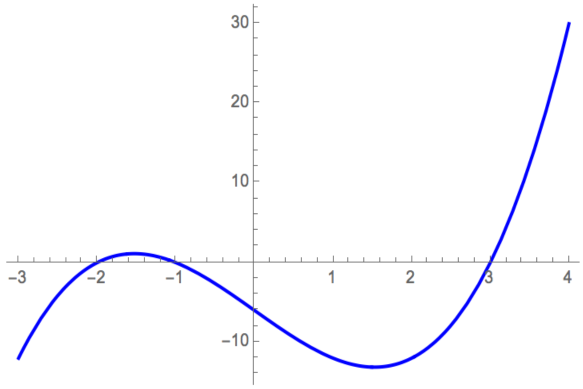


Figure 7.1: The graph of the polynomial $x^3 - 7x - 6 = (x+2)(x+1)(x-3)$.

Example 7.2.1. The general quadratic polynomial is $q(x) = ax^2 + bx + c = 0$ where $a, b, c \in \mathbb{R}$. Its two roots are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The quantity $b^2 - 4ac$ is the *discriminant* of the quadratic polynomial $q(x)$. The quadratic formula shown above tells us the following:

1. If $b^2 - 4ac > 0$, then $q(x)$ has two real roots.
2. If $b^2 - 4ac = 0$ then $q(x)$ has a double real root,
3. and if $b^2 - 4ac < 0$, $q(x)$ has two complex roots of the form $\alpha + i\beta$ and $\alpha - i\beta$.

Try making examples of quadratics that have all these possibilities. □

Idea to solve $p(x) = 0$:

Suppose we could find a matrix A_p such that $p(x)$ is the characteristic polynomial of A_p. Then the roots of $p(x)$ would be the eigenvalues of A_p.

We will see that this is always possible!!

We first give the algorithm and illustrate on an example. Then we will see why the algorithm works.

1. Input: $p(x) = a_d x^d + a_{d-1}x^{d-1} + a_{d-1}x^{d-2} + \cdots + a_1 x + a_0$
2. Set up the following $d \times d$ matrix

$$A_p = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & -a_0/a_d \\ 1 & 0 & 0 & 0 & \cdots & 0 & -a_1/a_d \\ 0 & 1 & 0 & 0 & \cdots & 0 & -a_2/a_d \\ 0 & 0 & 1 & 0 & \cdots & 0 & -a_3/a_d \\ \vdots & & & & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -a_{d-1}/a_d \end{bmatrix}$$

3. The characteristic polynomial of A_p, namely $\det(A_p - \lambda I)$, is either $p(\lambda)$ or $-p(\lambda)$. Therefore, the eigenvalues of A_p are the roots of $p(x)$.

Let's use Julia to find the roots of the polynomial $p(x) = x^5 - 20x^4 + 8x^2 - 10$.

Example 7.2.2. See Julia documentation on polynomials if the following commands are not self evident.

```
julia> using Polynomials

julia> p = Poly([-10,0,8,0,-20,1])
Poly(-10 + 8*x^2 - 20*x^4 + x^5)

julia> roots(p)
5-element Array{Complex{Float64},1}:
 -0.6685161487282535 - 0.4961354572272264im
 -0.6685161487282535 + 0.4961354572272264im
  0.6785047807938136 - 0.5116507106132901im
  0.6785047807938136 + 0.5116507106132901im
  19.980022735868893 + 0.0im
```

The polynomial $p(x)$ has 5 roots as expected. The last one is real. The other 4 are complex and come in conjugate pairs: $a + ib$ and $a - ib$. Recall that if a polynomial with real coefficients has complex roots then the complex roots come in conjugate pairs.

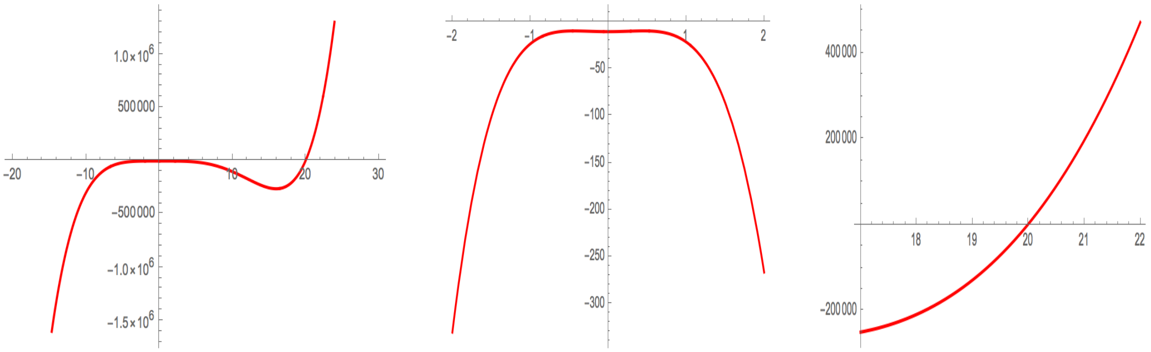


Figure 7.2: On the left we see the graph of $p(x) = x^5 - 20x^4 + 8x^2 - 10$. In the middle it is zoomed in around the origin so you can see that the graph is below the x-axis. On the right the graph is zoomed in near $x = 20$ and you can see there is a real root near 20.

Now we set up the matrix A_p as follows. Call it A in Julia.

```
julia> A = [0 0 0 0 10; 1 0 0 0 0; 0 1 0 0 -8; 0 0 1 0 0; 0 0 0 1 20]
517Array{Int64,2}:
 0  0  0  0  10
 1  0  0  0   0
 0  1  0  0  -8
 0  0  1  0   0
 0  0  0  1  20

julia> using LinearAlgebra
julia> eigvals(A)
5-element Array{Complex{Float64},1}:
 -0.6685161487282542 - 0.4961354572272264im
 -0.6685161487282542 + 0.4961354572272264im
   0.678504780793814 - 0.5116507106132913im
   0.678504780793814 + 0.5116507106132913im
   19.98002273586887 + 0.0im
```

Check that the eigenvalues of A are exactly the roots of $p(x)$.

There are two ways to see that our algorithm works. A first mechanical proof is below. A second proof follows from understanding quotients of polynomial rings which is the content of the next section.

A mechanical proof: Check that the determinant of $A_p - \lambda I$ is exactly $p(\lambda)$ or $-p(\lambda)$. The best strategy is to expand the determinant along the last column and you will see that this will work out. Try it on some examples first and this might help you warm up for the general proof. You might need to know some proof techniques like induction to do this precisely. For right now, just check it on examples and you have a proof by example!

Some comments:

1. The matrix A_p is called the *companion matrix* of the polynomial $p(x)$.

2. We see that every univariate polynomial is a determinant since it is the characteristic polynomial of its companion matrix which is a determinant. It was, and continues to be, an interesting question in mathematics to ask if a polynomial (in many variables) can be expressed as a determinant of special classes of matrices. A positive answer can have good algorithmic consequences.

7.3 Quotients of $\mathbb{R}[x]$ (Optional)

We now introduce a more complicated vector space of polynomials called a quotient space of $\mathbb{R}[x]$. The idea is similar to how we look at integers mod a given integer.

7.3.1 Warm up: Integers mod 5

The set of integers mod 5, denoted as $\mathbb{Z}_5$, is the collection $\{[0],[1],[2],[3],[4]\}$. It is constructed as follows. Take any integer a and suppose $a = 5k + r$ where k is an integer, then we say that a is equal to r mod 5, written as $a \equiv r \bmod 5$, and read as *a is congruent to r* mod 5. Equivalently $a \equiv r \bmod 5$ if and only if $a - r$ is a multiple of 5. The set of all integers congruent to 3 mod 5 is

$$[3] := \{\ldots, -12, -7, -2, 3, 8, 13, 18, 23, \ldots\}$$

Indeed, $8 \equiv 3 \bmod 5$ since $8 = 5 + 3$ and $-2 \equiv 3 \bmod 5$ since $-2 = -1 \cdot 5 + 3$. For any two elements of this set such as 13 and 23, their difference $13 - 23$ is a multiple of 5 We can represent the above set by any one of its elements since all other elements differ from the chosen one by multiples of 5. It is conventional to use the smallest nonnegative integer in the set to represent the set. In the above example this integer is 3 and we say that the set shown is the set of all integers *congruent to* 3 mod 5. We name the set [3] and call it the *congruence class* of 3 mod 5. Note that if we divided any of the elements of the set by 5 we would get the remainder 3. In theory it does not matter which element we choose to represent the set. In our example we could have chosen 13 to represent the set and write the set as [13], but the convention is to use the common remainder of all elements in the set on division by 5.

If r is the remainder obtained by dividing a with 5, then the possible values of r are $0, 1, 2, 3, 4$. Each of these remainders index a congruence class mod 5 in the same way as for 3. There are 5 equivalence classes of integers mod 5 and we have represented them with the remainder on division by 5:

$$\begin{aligned}\{\ldots,-15,-10,-5,0,5,10,15,\ldots\}&=[0]\\ \{\ldots,-14,-9,-4,1,6,11,16,\ldots\}&=[1]\\ \{\ldots,-13,-8,-3,2,7,12,17,\ldots\}&=[2]\\ \{\ldots,-12,-7,-2,3,8,13,18,23,\ldots\}&=[3]\\ \{\ldots,\ldots,-6,-1,4,9,14,19,24,29,\ldots\}&=[4]\end{aligned}$$

The set of congruence classes is $\mathbb{Z}_5$, the integers mod 5. Note that each element represents an infinite collection of integers, and that every integer falls into one of the classes.

7.3.2 Polynomials mod $p(x)$

We can do something similar in $\mathbb{R}[x]$. Pick a polynomial $p(x)$, say $p(x) = x^5 - 20x^4 + 8x^2 - 10$ and create equivalence classes of polynomials by saying that $a(x) \equiv b(x) \bmod p(x)$ if $a(x) - b(x)$ is a multiple of $p(x)$. This is the same as saying that if we divide $a(x)$ by $p(x)$ using long division from high school, then we get the same remainder $r(x)$ as when we divide $b(x)$ by $p(x)$. So to create an equivalence class, we could pick a polynomial and take all other polynomials that differ from it by multiples of $p(x)$. By multiple we now mean polynomials of the form $k(x)p(x)$ where $k(x)$ is also a polynomial. For example, take $a(x) = x^6 + 1$. After long division with $p(x)$ we see that

$$\underbrace{x^6+1}_{a(x)} = \underbrace{(x+20)}_{k(x)}\underbrace{(x^5-20x^4+8x^2-10)}_{p(x)} + \underbrace{400x^4-8x^3-160x^2+10x+201}_{r(x)}.$$

Therefore, $x^6 + 1 \equiv 400x^4 - 8x^3 - 160x^2 + 10x + 201 \bmod p(x)$ since both $x^6 + 1$ and its remainder $400x^4 - 8x^3 - 160x^2 + 10x + 201$ have the same remainder after division by $p(x)$. Any polynomial of the form $(x^6 + 1) + k(x)p(x)$ is equivalent to $x^6 + 1$.

A natural representative for any equivalence class is the common remainder $r(x)$ of all elements in the class after division by $p(x)$. We write $[r(x)]$ for the class represented by $r(x)$. Let $\mathbb{R}[x]/(p(x))$ be the set of all equivalence classes, each represented by a possible remainder after division by $p(x)$. The set of equivalence classes, $\mathbb{R}[x]/(p(x))$, is called the quotient of $\mathbb{R}[x]$ by $p(x)$. Each equivalence class has infinitely many polynomials in it, just like an equivalence class mod 5 had infinitely many integers in it. In the case of integers mod 5 what happened is that all of $\mathbb{Z}$ got partitioned into 5 sets depending on what the remainder on division by 5 was. The same is happening here — all of $\mathbb{R}[x]$ is getting partitioned into infinite sets of polynomials that have the same remainder on division by $p(x)$.

How many equivalence classes are there? Equivalently, how many different remainders are there after division by our given $p(x)$? Note that any polynomial of degree at most 4 cannot be divided by $p(x)$ or equivalently, any polynomial in $\mathbb{R}[x]_{\leq 4}$ is its own remainder when you divide it by $p(x)$ and so all of $\mathbb{R}[x]_{\leq 4}$ are remainders. Already we have infinitely many remainders. Can two of them be in the same equivalence class? If $a(x)$ and $b(x)$ are two different polynomials in $\mathbb{R}[x]_{\leq 4}$, then their difference is also in $\mathbb{R}[x]_{\leq 4}$ and so the difference cannot be a multiple of $p(x)$. This means that $a(x)$ and $b(x)$ belong to different equivalence classes mod $p(x)$. How about a polynomial of degree 5 or more? Such a polynomial can be divided by $p(x)$ and the remainder will be a polynomial in $\mathbb{R}[x]_{\leq 4}$. We conclude that the elements of $\mathbb{R}[x]_{\leq 4}$ represent the different equivalence classes mod $p(x)$. Since $R[x]_{\leq 4}$ has infinitely many elements, there are infinitely many equivalence classes of $\mathbb{R}[x]$ mod $p(x)$. Each element $r(x) \in R[x]_{\leq 4}$ gives rise to the equivalence class $[r(x)] \in \mathbb{R}[x]/(p(x))$ and vice versa. Therefore, there is a bijection between $\mathbb{R}[x]/(p(x))$ and $R[x]_{\leq 4}$ by identifying $[r(x)]$ with $r(x)$.

Recall that if $r_1(x), r_2(x) \in R[x]_{\leq 4}$, then their sum $r_1(x) + r_2(x)$ is also in $R[x]_{\leq 4}$. Does the same rule work in $\mathbb{R}[x]/(p(x))$? i.e., is $[r_1(x)] + [r_2(x)] = [r_1(x) + r_2(x)]$? For this to be true, we would need that when we add two polynomials $f_1(x) \in [r_1(x)]$ (which means that remainder of $f_1(x)$ when divided by $p(x)$ is $r_1(x)$) and $f_2(x) \in [r_2(x)]$ (which means that remainder of $f_2(x)$ when divided by $p(x)$ is $r_2(x)$), the sum polynomial $f_1(x) + f_2(x)$ will have remainder $r_1(x) + r_2(x)$ after division by $p(x)$. Suppose $f_1(x) \in [r_1(x)]$ and $f_2(x) \in [r_2(x)]$. Then $f_1(x) = k_1(x)p(x) + r_1(x)$ and $f_2(x) = k_2(x)p(x) + r_2(x)$. This means that $f_1(x) + f_2(x) = (k_1(x) + k_2(x))p(x) + (r_1(x) + r_2(x))$. Check that $r_1(x) + r_2(x)$ cannot be divided by $p(x)$ since neither $r_1(x)$ nor $r_2(x)$ can be. Therefore, $f_1(x)+f_2(x)$ belongs to the equivalence class $[r_1(x)+r_2(x)]$. So indeed, $[r_1(x)] + [r_2(x)] = [r_1(x) + r_2(x)]$.

Let's check scalar multiplication. Is it true that $\gamma[r(x)] = [\gamma r(x)]$? Again if $f(x) \in [r(x)]$ then $f(x) = k(x)p(x) + r(x)$. Therefore, $\gamma f(x) = \gamma k(x)p(x) + \gamma r(x)$. Since $r(x)$ cannot be divided by $p(x)$, $\gamma r(x)$ cannot be either. Therefore, $\gamma f(x) \in [\gamma r(x)]$, and we have that $\gamma[r(x)] = [\gamma r(x)]$.

These calculations show us that $\mathbb{R}[x]/(p(x))$ is a vector space under the following rules of addition and scalar multiplication which looks exactly like those in $\mathbb{R}[x]_{\leq 4}$.

- **Addition in** $\mathbb{R}[x]/(p(x))$: if $[r_1(x)]$ and $[r_2(x)]$ are two equivalence classes mod $p(x)$, then their sum is the equivalence class $[r_1(x) + r_2(x)]$ mod $p(x)$.

- **Scalar multiplication in** $\mathbb{R}[x]/(p(x))$: if $[r(x)]$ is an equivalence class and $\gamma \in \mathbb{R}$, then $\gamma[r(x)] = [\gamma r(x)]$ mod $p(x)$.

Recall that $R[x]_{\leq 4}$ was a vector space of dimension 5 with basis $\mathcal{B}_4 = \{1, x, x^2, x^3, x^4\}$. The quotient space $\mathbb{R}[x]/(px))$ is also a vector space of dimension 5 with basis $\{[1], [x], [x^2], [x^3], [x^4]\}$. In general we have the following.

Proposition 7.3.1. *Given a polynomial $p(x)$ of degree d,*

1. *The quotient space $\mathbb{R}[x]/(p(x))$ is the set of all equivalence classes of polynomials mod $p(x)$, and its elements are represented by the remainders of polynomials in $\mathbb{R}[x]$ after division by $p(x)$.*

2. *The space $\mathbb{R}[x]/(p(x))$ is a vector space under the addition and scalar multiplication rules given above.*

3. *$\{[1], [x], [x^2], \ldots, [x^{d-1}]\}$ is a basis of $\mathbb{R}[x]/(p(x))$. The elements of $\mathbb{R}[x]/(p(x))$ are linear combinations of $[1], [x], [x^2], \ldots, [x^{d-1}]$.*

4. *$\dim(\mathbb{R}[x]/(p(x)) = d$.*

7.3.3 Proof of the algorithm for solving a polynomial equation

We'll explain the proof on the example $p(x) = x^5 - 20x^4 + 8x^2 - 10$. You can then extrapolate to the general case. Recall that $\mathbb{R}[x]/(p(x))$ consists of all equivalence classes of polynomials mod $p(x)$. The class of a polynomial $f(x)$ is $[r(x)]$ where $r(x)$ is the remainder obtained when dividing $f(x)$ by $p(x)$. The quotient space $\mathbb{R}[x]/(p(x))$ is a vector space of dimension 5 with basis $\{[1], [x], [x^2], [x^3], [x^4]\}$.

Consider the linear transformation $T : \mathbb{R}[x]/(p(x)) \to \mathbb{R}[x]/(p(x))$ that sends $[r(x)] \mapsto [xr(x)]$. For example

$$T([3x^2 + 5x - 1]) = [x(3x^2 + 5x - 1)] = [3x^3 + 5x^2 - x]$$

What happens if $xr(x)$ is a polynomial of degree larger than 4? For example

$$T([x^4 + 5x - 1]) = [x(x^4 + 5x - 1)] = [x^5 + 5x^2 - x]$$

Since $x^5 + 5x^2 - x$ can be divided by $p(x)$ giving the remainder $20x^4 - 3x^2 - x + 10$, we have that

$$[x^5 + 5x^2 - x] = [20x^4 - 3x^2 - x + 10], \quad \text{and} \quad T([x^4 + 5x - 1]) = [20x^4 - 3x^2 - x + 10].$$

You can check that T is a linear transformation by checking the following:

$$T([r_1(x)] + [r_2(x)]) = T([r_1(x)]) + T([r_2(x)]) \quad \text{and} \quad T([\alpha r(x)]) = \alpha T([r(x)]).$$

To compute the matrix A_p of this linear transformation we need to first look at where T sends each of the basis elements $[1], [x], [x^2], [x^3], [x^4]$ of $\mathbb{R}[x]/(p(x))$. The columns of A_p are then the coordinates of $T([1]), T([x]), T([x^2]), T([x^3])$ and $T([x^4])$ in the basis $\{[1], [x], [x^2], [x^3], [x^4]\}$. Let's do the calculation:

$$\begin{aligned}
T([1]) &= [x] = 0\cdot[1] + 1\cdot[x] + 0\cdot[x^2] + 0\cdot[x^3] + 0\cdot[x^4]\\
T([x]) &= [x^2] = 0\cdot[1] + 0\cdot[x] + 1\cdot[x^2] + 0\cdot[x^3] + 0\cdot[x^4]\\
T([x^2]) &= [x^3] = 0\cdot[1] + 0\cdot[x] + 0\cdot[x^2] + 1\cdot[x^3] + 0\cdot[x^4]\\
T([x^3]) &= [x^4] = 0\cdot[1] + 0\cdot[x] + 0\cdot[x^2] + 0\cdot[x^3] + 1\cdot[x^4]\\
T([x^4]) &= [x^5] = 10\cdot[1] + 0\cdot[x] - 8\cdot[x^2] + 0\cdot[x^3] + 20\cdot[x^4] = [20x^4 - 8x^2 + 10]
\end{aligned}$$

Therefore, the matrix representing T in the basis $\{[1], [x], [x^2], [x^3], [x^4]\}$ is exactly

$$A_p = \begin{bmatrix} 0 & 0 & 0 & 0 & 10\\ 1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & -8\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 20 \end{bmatrix}$$

We are almost done. Suppose $\lambda \in \mathbb{R}$ and $[r(x)] \in \mathbb{R}[x]/(p(x))$ is an eigenvalue/eigenvector pair of T, meaning $T([r(x)]) = \lambda[r(x)]$. This means that $[xr(x)] = T([r(x)]) = \lambda[r(x)] = [\lambda r(x)]$. In other words, $(x-\lambda)r(x)$ is a multiple of $p(x)$, or equivalently, there is a polynomial $k(x)$ such that

$$k(x)p(x) = (x-\lambda)r(x).$$

This equality is in $\mathbb{R}[x]$. Let's look at the degrees on the left and right. On the right we have a polynomial of degree at most 5 since the degree of $r(x)$ is at most 4. Since $p(x)$ already has degree 5, it must be that $k(x)$ is a real number. This means $(x-\lambda)r(x)$ is a factorization of $p(x)$ (up to some scalar multiple perhaps) and one of its factors is $(x-\lambda)$. Therefore, λ is a root of $p(x)$. On the other hand, if λ is a root of $p(x)$, then $p(x) = (x-\lambda)r(x)$ for some $r(x)$ of degree 4 and $[xr(x)] = [\lambda r(x)]$ and $[r(x)]$ is an eigenvector of λ for the linear transformation T. Since A_p represents T, we conclude that the eigenvalues of A_p are exactly the roots of $p(x)$.

If you would like a proof in general, then here are the items you need to prove before invoking the argument in the last paragraph above:

1. Suppose $p(x) = a_dx^d + a_{d-1}x^{d-1} + \cdots + a_1x + a_0 \in \mathbb{R}[x]$. We saw that $\mathbb{R}[x]/(p(x))$ is a vector space. Show that
$$T : \mathbb{R}[x]/(p) \to \mathbb{R}[x]/(p) \quad \text{s.t. } [r(x)] \mapsto [xr(x)]$$
is a linear transformation. That is, show that

   (a) $T([r_1(x)] + [r_2(x)]) = T([r_1(x)]) + T([r_2(x)])$, and

   (b) for any real number γ, $T([\gamma r(x)]) = \gamma T([r(x)])$.

   Note: You may verify this in multiple ways depending on the definition of $\mathbb{R}[x]/(p(x))$ that you are using. The two definitions that we have are

   (a) All elements of $\mathbb{R}[x]/(p(x))$ are given by the set of all (unique) remainders of polynomials in $\mathbb{R}[x]$ after division by $p(x)$, i.e. the element $[r(x)]$ denotes the equivalence class of all polynomials that have a remainder of $r(x)$ when divided by $p(x)$.

   (b) All elements of $\mathbb{R}[x]/(p(x))$ are of the form $f(x) + (p(x))$, where $f(x) \in \mathbb{R}[x]$ and $(p(x))$ denotes the subspace of $\mathbb{R}[x]$ containing all (polynomial) multiples of $p(x)$.

2. Find a basis for $\mathbb{R}[x]/(p(x))$ and explain why it is a basis. What is $\dim\Big(\mathbb{R}[x]/(p(x))\Big)$?

3. The companion matrix A_p is the matrix representing the above linear transformation. Compute A_p using the basis you computed in (c). Verify that it agrees with the matrix A_p in Step 2 of the algorithm. (**Hint:** Recall how to compute the matrix of a linear map by looking at what the map does to a basis of the domain vector space.)

Chapter 8

Singular Value Decomposition

The course thus far was focused on eigenvalues and eigenvectors and their applications. This is a theory that only applies to square matrices – it was necessary that the domain and codomain of the linear transformation represented by the matrix were the same so that we could understand when the action of the matrix on a vector was simply scaling. In the presence of a full eigenbasis, the matrix can be diagonalized and diagonalization has many useful consequences. The spectral theorem for symmetric matrices says that all symmetric matrices can be diagonalized, and that the diagonalization is special in the sense of being orthonormal.

We now move on to general (rectangular) matrices and see that every matrix admits an important factorization called a *singular value decomposition* (SVD) which generalizes diagonalization for square matrices. We will study the structure and geometry of this decomposition, followed by glimpses of its numerous applications. While diagonalization of a square matrix was very useful in applications involving powers of the matrix, we will see that the SVD offers fundamental insight into the matrix itself.

8.1 Structure of the SVD

Theorem 8.1.1. *Any matrix $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = r$, admits a factorization of the form*

$$A = U\Sigma V^\top$$

*called a **singular value decomposition (SVD)**. The matrices U, Σ and V have the following properties:*

- *$U = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \cdots & \boldsymbol{u}_m \end{bmatrix} \in \mathbb{R}^{m\times m}$ is orthonormal, i.e., $UU^\top = U^\top U = I_m$. The columns $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ form an orthonormal basis for $\operatorname{Col}(A)$ and $\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m$ form an orthonormal basis for $\operatorname{Null}(A^\top)$. Recall that $\operatorname{Col}(A)$ and $\operatorname{Null}(A^\top)$ are orthogonal complements. The vectors $\boldsymbol{u}_i$ are the **left singular vectors** of A.*

- *$V = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r & \boldsymbol{v}_{r+1} & \cdots & \boldsymbol{v}_n \end{bmatrix} \in \mathbb{R}^{n\times n}$ is also orthonormal, i.e., $VV^\top = V^\top V = I_n$. The columns $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ form an orthonormal basis for $\operatorname{Row}(A)$ and $\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n$ form an orthonormal basis for $\operatorname{Null}(A)$. Recall that $\operatorname{Row}(A)$ and $\operatorname{Null}(A)$ are orthogonal complements. The vectors $\boldsymbol{v}_i$ are the **right singular vectors** of A.*

- $\Sigma \in \mathbb{R}^{m\times n}$ *is a nonnegative matrix of the same shape as* A *and non-zero entries only on the diagonal:*

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

  Its first r diagonal entries are positive and arranged in non-increasing order, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0$. *These positive entries are the* **singular values** *of* A.

The SVD can be summarized as follows:

$$A = \underbrace{\left[\begin{array}{c|c} \underbrace{\mathbf{u}_1 \ \cdots \ \mathbf{u}_r}_{\text{Col}(A)} & \underbrace{\mathbf{u}_{r+1} \ \cdots \ \mathbf{u}_m}_{\text{Null}(A^\top)} \end{array}\right]}_{m\times m} \underbrace{\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}}_{m\times n} \underbrace{\left[\begin{array}{cc} \mathbf{v}_1^\top & \\ \vdots & \text{Row}(A) \\ \mathbf{v}_r^\top & \\ \hline \mathbf{v}_{r+1}^\top & \\ \vdots & \text{Null}(A) \\ \mathbf{v}_n^\top & \end{array}\right]}_{n\times n}$$

We will not prove the above theorem in this course. Instead we will focus on what the SVD means and how to use it. Let's first see some examples.

Example 8.1.2. Consider the rank two matrix

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}.$$

Its singular value decomposition is:

$$\begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \underbrace{\frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{45} & 0 \\ 0 & \sqrt{5} \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}}_{V^\top}.$$

Check all the claims in the theorem.

Example 8.1.3. We now use Julia to compute the SVD of the following 3×3 matrix whose rank is 2:

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

```
julia> B = [1 1 1; -1 0 2; 0 1 3]
317Array{Int64,2}:
  1  1  1
 -1  0  2
  0  1  3

julia> svd(B)
```

```
SVD{Float64,Float64,Array{Float64,2}}
U factor:
317Array{Float64,2}:
 -0.30519    0.757315  -0.57735
 -0.503259  -0.64296   -0.57735
 -0.808449   0.114355   0.57735
singular values:
3-element Array{Float64,1}:
  3.904484344750072
  1.6598198702273692
 -0.0
Vt factor:
317Array{Float64,2}:
 0.0507284  -0.285221  -0.957119
 0.84363     0.525159  -0.111784
 0.534522   -0.801784   0.267261
```

Truncating the above numbers at 5 decimal places, the singular value decompostion of B is

$$\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix} = \begin{bmatrix} -0.30519 & 0.75731 & -0.57735 \\ -0.50325 & -0.64296 & -0.57735 \\ -0.80844 & 0.11435 & 0.57735 \end{bmatrix} \begin{bmatrix} 3.90448 & 0 & \\ 0 & 1.65981 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.05072 & -0.28522 & -0.95711 \\ 0.84363 & 0.52515 & -0.11178 \\ 0.53452 & -0.80178 & 0.26726 \end{bmatrix}$$

Notice that B has only two singular values and $\operatorname{rank}(B) = 2$.

We now make sense of the components of the SVD. Suppose $A = U\Sigma V^\top$ as in Theorem 8.1.1. Consider the matrix $A^\top A \in \mathbb{R}^{n\times n}$ which is symmetric (in fact positive semidefinite). Check that

$$A^\top A = V\Sigma^\top \underbrace{U^\top U}_{I_m} \Sigma V^\top = V\Sigma^\top \Sigma V^\top$$

and

$$\Sigma^\top \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

Since V is orthonormal and $\Sigma^\top\Sigma$ is diagonal, it must be that $V\Sigma^\top\Sigma V^\top$ is an orthonormal diagonalization of $A^\top A$. This allows us to make the following observations:

1. $\sigma_1^2, \dots, \sigma_n^2$ are the eigenvalues of $A^\top A$ in non-increasing order. They are indeed nonnegative since $A^\top A$ is PSD. Since $\sigma_i = 0$ if and only if $i = r+1, \dots, n$, the eigenvalue 0 of $A^\top A$ has arithmetic, and hence also geometric, multiplicity $n - r$. This means that $\operatorname{rank}(A^\top A) = \operatorname{rank}(A) = r$.

2. The columns of V consist of n orthonormal eigenvectors of $A^\top A$, and $A^\top A\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i$ for $i = 1, \dots, n$. In particular, $A^\top A\mathbf{v}_i = 0$ for all $i = r+1, \dots, n$ and hence $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ form an orthonormal basis for $\operatorname{Null}(A^\top A)$. If $A\mathbf{x} = \mathbf{0}$ then $A^\top A\mathbf{x} = \mathbf{0}$ and so $\operatorname{Null}(A) \subseteq \operatorname{Null}(A^\top A)$. But since $\operatorname{rank}(A^\top A) = \operatorname{rank}(A)$, we have that $\operatorname{Null}(A) = \operatorname{Null}(A^\top A)$. Therefore, $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ form an orthonormal basis for $\operatorname{Null}(A)$. Since the entire collection $\mathbf{v}_1, \dots, \mathbf{v}_n$ is orthonormal, $\mathbf{v}_1, \dots, \mathbf{v}_r$ lives in $\operatorname{Row}(A) = \operatorname{Null}(A)^\perp$. Moreover, $\mathbf{v}_1, \dots, \mathbf{v}_r$ forms an orthonormal basis for $\operatorname{Row}(A)$.

We summarize the main findings in a lemma.

Lemma 8.1.4. *Suppose $A \in \mathbb{R}^{m\times n}$ and $\text{rank}(A) = r$.*

1. *The **singular values** of A are the positive square roots of the r positive eigenvalues of $A^\top A$.*
2. *The **right singular vectors** of A are an orthonormal basis of eigenvectors of $A^\top A$. The first r of them form an orthonormal basis of $\text{Row}(A)$ and the last $n - r$ of them form an orthonormal basis of $\text{Null}(A)$.*

Consider again the equation $A = U\Sigma V^\top$. Multiplying both sides by V we get that

$$AV = U\Sigma V^\top V = U\Sigma$$

since $V^\top V = I_n$. Next note that $U\Sigma$ is equal to

$$\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m\end{bmatrix}\begin{bmatrix}\sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & \\ & & & & & 0\end{bmatrix} = \begin{bmatrix}\sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0}\end{bmatrix}$$

Therefore,

$$A\mathbf{v}_i = \sigma_i\mathbf{u}_i \ \ \forall\ i = 1, \dots, r \quad \text{and} \quad A\mathbf{v}_i = \mathbf{0} \ \ \forall\ \ i = r+1, \dots, n$$

In other words, the first r left singular vectors of A are

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sigma_i} \quad \text{for all } \ i = 1, \dots, r$$

Using this equation we can check that these first r $\mathbf{u}_i$'s are mutually orthogonal:

$$\mathbf{u}_i^\top\mathbf{u}_j = \left(\frac{A\mathbf{v}_i}{\sigma_i}\right)^\top\left(\frac{A\mathbf{v}_j}{\sigma_j}\right) = \frac{\mathbf{v}_i^\top(A^\top A\mathbf{v}_j)}{\sigma_i\sigma_j} = \frac{\mathbf{v}_i^\top(\sigma_j^2\mathbf{v}_j)}{\sigma_i\sigma_j} = \frac{\sigma_j\mathbf{v}_i^\top\mathbf{v}_j}{\sigma_i} = 0$$

The last equality follows from the fact that the columns of V are mutually orthogonal. We can also see that these first r $\mathbf{u}_i$'s are unit vectors:

$$||\mathbf{u}_i||^2 = \mathbf{u}_i^\top\mathbf{u}_j = \left(\frac{A\mathbf{v}_i}{\sigma_i}\right)^\top\left(\frac{A\mathbf{v}_i}{\sigma_i}\right) = \frac{\mathbf{v}_i^\top A^\top A\mathbf{v}_i}{\sigma_i^2} = \mathbf{v}_i^\top\mathbf{v}_i = ||\mathbf{v}_i||^2 = 1$$

hence $||\mathbf{u}_i|| = 1$.

Since a matrix and its transpose have the same eigenvalues, the non-zero eigenvalues of $AA^\top$ are also $\sigma_1^2, \dots, \sigma_r^2$. Using the fact that $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$ for $i = 1, \dots, r$, we get that

$$AA^\top\mathbf{u}_i = AA^\top\left(\frac{A\mathbf{v}_i}{\sigma_i}\right) = \frac{A\Big(A^\top A\mathbf{v}_i\Big)}{\sigma_i} = \frac{A\sigma_i^2\mathbf{v}_i}{\sigma_i} = \sigma_i A\mathbf{v}_i = \sigma_i^2\mathbf{u}_i$$

and hence the first r left singular vectors $\mathbf{u}_i$ are eigenvectors of $AA^\top$ with eigenvalues σ_i^2.

The last $m - r$ singular vectors $\mathbf{u}_{r+1}, \dots, \mathbf{u}_m$ are chosen to be an orthonormal basis of $\text{Null}(A^\top)$. In fact, if we had done the above calculation starting with the PSD matrix $AA^\top$ instead of $A^\top A$. Then by the exact same reasoning as before we would get the following results:

Lemma 8.1.5. *Suppose* $A \in \mathbb{R}^{m\times n}$ *and* $\operatorname{rank}(A) = r$.

1. *The **singular values** of* A *are the positive square roots of the* r *positive eigenvalues of* $AA^\top$.
2. *The* **left singular vectors** *of* A *are an orthonormal basis of eigenvectors of* $AA^\top$. *The first* r *of them form an orthonormal basis of* $\operatorname{Col}(A)$ *and the last* $m-r$ *of them form an orthonormal basis of* $\operatorname{Null}(A^\top)$.

We have now seen that the singular values of A are square roots of the non-zero eigenvalues of both $A^\top A$ and $AA^\top$. If A is an $m \times n$ matrix, with $m \neq n$, then $A^\top A$ and $AA^\top$ have different sizes, hence different numbers of eigenvalues. This poses potential confusion, but we note that they both have the same number of **non-zero** eigenvalues equal to $\operatorname{rank}(A)$. There will potentially be a different multiplicity of 0 eigenvalues for $A^\top A$ and $AA^\top$, but this does not pose any problems to the mechanics of the SVD.

The above discussion allows us to compute an SVD as we illustrate below.

Example 8.1.6. Consider again the rank 2 matrix

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$$

We first compute the singular values and right singular vectors:

$$A^\top A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \quad \text{with} \quad \lambda_1 = 45 = \sigma_1^2, \lambda_2 = 5 = \sigma_2^2 \quad \text{and} \quad \mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

This implies that

$$V = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sqrt{45} & 0 \\ 0 & \sqrt{5} \end{bmatrix}$$

We now use the equation $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ to find the $\mathbf{u}_i$. We have

$$A\mathbf{v}_1 = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} \\ 9/\sqrt{2} \end{bmatrix} = \frac{3}{\sqrt{2}}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sqrt{45}\frac{1}{\sqrt{10}}\begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

so $\mathbf{u}_1 = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

$$A\mathbf{v}_2 = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} -3/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \sqrt{5}\frac{1}{\sqrt{10}}\begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

so $\mathbf{u}_2 = \frac{1}{\sqrt{10}}\begin{bmatrix} -3 \\ 1 \end{bmatrix}$. It is important to note that since $\operatorname{rank}(A) = 2$ we did not have to compute any vectors from $\operatorname{Null}(A)$ or $\operatorname{Null}(A^\top)$. In more general scenarios, bases for these subspaces will have to be computed and normalized in order to obtain the matrices V and U.

We now can conclude that

$$U = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}$$

and the singular value decomposition of A is

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{45} & 0 \\ 0 & \sqrt{5} \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

In practice one usually finds the rank of a matrix by computing the SVD.

1. Compute the singular value decomposition (in Julia).
2. If some singular values are VERY close to 0, then they are 0 (this happens because of rounding errors in the computer).
3. $\operatorname{rank}(A) =$ the number of nonzero singular values of A.

8.2 The Reduced SVD

Let us look again at the SVD of $A \in \mathbb{R}^{m\times n}$ and see what parts of the decomposition really matter:

$$\begin{aligned}
A &= \left[\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_r \quad \middle| \quad \mathbf{u}_{r+1} \quad \cdots \quad \mathbf{u}_m\right] \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \left[\begin{array}{c} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_r \\ \hline v_{r+1} \\ \vdots \\ \mathbf{v}_n \end{array}\right] \\
&= \left[\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \middle| \quad \mathbf{0} \quad \cdots \quad \mathbf{0}\right] \left[\begin{array}{c} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_r \\ \hline v_{r+1} \\ \vdots \\ \mathbf{v}_n \end{array}\right] \\
&= \left[\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r\right] \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_r \end{bmatrix} \\
&= \left[\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_r\right] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ & \ddots & & 0 \\ & \cdots & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_r \end{bmatrix}
\end{aligned}$$

Definition 8.2.1. The **reduced singular value decomposition** of $A \in \mathbb{R}^{m\times n}$, $\operatorname{rank}(A) = r$ is the truncated factorization

$$A = \underbrace{\left[\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_r\right]}_{m\times r} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}}_{r\times r} \underbrace{\left[\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_r\right]^\top}_{r\times n} = \hat{U}\hat{\Sigma}\hat{V}^\top.$$

Note that in contrast to the (full) singular value decomposition, the reduced version has $\hat{\Sigma}$ as a square diagonal matrix and $\hat{U}$ and $\hat{V}$ as rectangular matrices in general.

Example 8.2.2. Let A be the following matrix

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

The only eigenvalue of A is 0 with $\mathrm{AM}(0) = 4$ and $\mathrm{GM}(0) = 1$. Moreover, $\operatorname{rank}(A) = 3$ and A is not invertible. We compute that

$$A^\top A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$$

with eigenvalues $0, 1, 4, 9$ with and hence $\sigma_1 = 3, \sigma_2 = 2, \sigma_3 = 1$, and $\sigma_4 = 0$. The respective eigenvectors of $A^\top A$ are

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and so

$$V^\top = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Next, we use the fact that $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ for $i = 1, 2, 3$ to find the first three columns of U. They are $\mathbf{u}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, and $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. The last step is to find the one orthonormal basis vector for $\text{Null}(A^\top)$, which by inspection is $\mathbf{u}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. This rounds out the computation, giving us the matrix U and the hence the full singular value decomposition of A is

$$A = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}}_{V^\top}$$

The reduced singular value decomposition eliminated the 0 singular values and any singular vectors coming from null spaces. The reduced SVD of A is

$$A = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\hat{U}} \underbrace{\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\hat{\Sigma}} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{\hat{V^\top}}$$

Julia typically outputs the reduced SVD of a matrix as we see in the following example of a 3×5 matrix.

Example 8.2.3. `julia> using LinearAlgebra`

```
julia> A = [1 2 3 4 5; 1 17 20 0 3; 2 5 8 -1 -4]
317Array{Int64,2}:
 1   2   3   4   5
 1  17  20   0   3
 2   5   8  -1  -4
```

```
julia> svd(A)
SVD{Float64,Float64,Array{Float64,2}}
U factor:
317Array{Float64,2}:
 -0.14392    0.775189    0.615117
 -0.935544   0.0960262  -0.339905
 -0.322558  -0.624389    0.711404

singular values:
3-element Array{Float64,1}:
 28.23022323551315
  7.678580767418608
  2.8449768841455674

Vt factor:
317Array{Float64,2}:
 -0.0610899  -0.630703    -0.769498   -0.0089663  -0.0792058
 -0.0491708   0.00792853  -0.0975465   0.485134    0.867553
  0.596848   -0.348382     0.25957     0.614791   -0.277593
```

The above output is the reduced SVD of A. In this example, $A \in \mathbb{R}^{3\times 5}$. From the above SVD we see that $\operatorname{rank}(A) = 3$ since there are 3 singular values. Therefore we expect $\hat{U} \in \mathbb{R}^{3\times 3}$, $\Sigma \in \mathbb{R}^{3\times 3}$ and $\hat{V}^\top \in \mathbb{R}^{3\times 5}$ as we see in the above output. In the full SVD Σ would have been 3×5 and $V^\top$ would have been 5×5.

8.3 Three Immediate Consequences

The SVD has many remarkable consequences. Here we list three that are immediate.

8.3.1 Rank One Decomposition

Every matrix can be written in multiple ways as the sum of rank one matrices. For example, if $A \in \mathbb{R}^{m\times n}$ you can create m rank one matrices R_i by retaining row i of A and replacing all other entries of A by 0. Then $A = R_1 + \cdots + R_m$ is a decomposition of A into rank one matrices. You could make a similar decomposition by retaining columns of A or even just individual entries of A. It is not clear what such decompositions are good for. The SVD creates a very special rank one decomposition that has numerous uses. The SVD allows every matrix to be written as a positive linear combination of rank one matrices. The number of rank one matrices in the decomposition is exactly the rank of the matrix.

Theorem 8.3.1. *Let $A \in \mathbb{R}^{m\times n}$ be a matrix of rank r and reduced SVD*

$$A = \underbrace{\begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \end{bmatrix}}_{m\times r} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}}_{r\times r} \underbrace{\begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r \end{bmatrix}^\top}_{r\times n}.$$

Then $A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^\top + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^\top + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^\top$ which is a **rank one decomposition** *of A into r rank one matrices made up of outer products of singular vectors of A.*

Example 8.3.2. Check that

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = 3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$$

The rank one decomposition of a matrix has numerous uses. We will see in class how to use it for image compression.

8.3.2 Polar Decomposition

Let $A \in \mathbb{R}^{n\times n}$ be a *square* matrix. Observe that we can use the SVD to write A as

$$A = U\Sigma V^\top = UV^\top V\Sigma V^\top = (UV^\top)(V\Sigma V^\top).$$

Check that $UV^\top$ is orthogonal and $V\Sigma V^\top$ is PSD. We can conclude that any square matrix is the product of an othogonal matrix and a PSD matrix. This factorization is a **polar decomposition of** A.

8.3.3 Change of Bases

The SVD of $A \in \mathbb{R}^{m\times n}$ is like a diagonalization of A. It allows the action of A to be reduced to the action of Σ if $A = U\Sigma V^\top$ is the SVD of A.

The linear transformation represented by A is:

$$A : \mathbb{R}^n \to \mathbb{R}^m, \quad \mathbf{x} \mapsto A\mathbf{x}$$

Suppose $A\mathbf{x} = \mathbf{b}$. Then from the SVD we have that

$$\mathbf{b} = A\mathbf{x} = U\Sigma V^\top \mathbf{x} = U(\Sigma(V^\top \mathbf{x})).$$

We analyze $U(\Sigma(V^\top\mathbf{x}))$ step by step from the right to left. In the first step we compute $V^\top\mathbf{x}$ and interpret it. Since V is orthonormal, its columns form an orthonormal basis of $\mathbb{R}^n$. The coordinates of $\mathbf{x} \in \mathbb{R}^n$ in the basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is

$$\mathbf{y} := V^{-1}\mathbf{x} = V^\top\mathbf{x}.$$

Next, we apply Σ to $V^\top\mathbf{x}$ to get $\mathbf{b}' := \Sigma(V^\top\mathbf{x}) = \Sigma\mathbf{y} \in \mathbb{R}^m$. This action scales the first r entries of $\mathbf{y}$ separately by the singular values $\sigma_1, \dots, \sigma_r$ and the remaining entries are 0. In particular this action is very easy to understand.

Finally we need to understand $\mathbf{b} = U\Sigma V^\top\mathbf{x} = U\Sigma\mathbf{y} = U\mathbf{b}'$. Since U is orthonormal, its columns form an orthonormal basis of $\mathbb{R}^m$. The coordinates of $\mathbf{b} = A\mathbf{x}$ in the U basis is

$$U^\top\mathbf{b} = U^\top(U\Sigma V^\top\mathbf{x}) = \Sigma V^\top\mathbf{x} = \mathbf{b}'.$$

Said another way if we think of $\mathbf{b}' = \Sigma\mathbf{y}$ as written in the U basis, then the coordinates of $\mathbf{b}'$ in the standard basis of $\mathbb{R}^m$ is $\mathbf{b}$. The following theorem summarizes this discussion.

Theorem 8.3.3. *Let $A = U\Sigma V^\top \in \mathbb{R}^{m\times n}$ be the SVD of A. If we use $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ and $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_m\}$ as orthonormal bases of the domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$ of A, respectively, then A behaves like the matrix Σ.*

The action of A can be broken up into three pieces as in the following pictures:

$$\underbrace{\mathbf{x}}_{\text{(In standard basis of } \mathbb{R}^n)} \quad \xrightarrow{V^\top} \quad \underbrace{\mathbf{y} := V^\top\mathbf{x}}_{\text{(In } V \text{ basis, still in } \mathbb{R}^n)} \quad \xrightarrow{\Sigma} \quad \underbrace{\mathbf{b}' := \Sigma\mathbf{y} = \Sigma V^\top\mathbf{x}}_{\text{(In } U \text{ basis, now in } \mathbb{R}^m)} \quad \xrightarrow{U} \quad \underbrace{U\mathbf{b}' = U\Sigma V^\top\mathbf{x} = \mathbf{b}}_{\text{(Back to standard basis, in } \mathbb{R}^m)}$$

We conclude that after changing the basis of the domain $\mathbb{R}^n$ to the V-basis and the basis of the codomain $\mathbb{R}^m$ to the U basis, the action of A is simply the action of the matrix Σ. Remember that the SVD exists for **all** matrices. So this is a vast generalization of the result we saw in Chapter 1 that said that every diagonalizable square matrix behaves like its eigenvalue matrix Λ with respect to an eigenbasis.

$$\begin{tikzcd}
\begin{array}{c}\text{standard}\\ \text{basis}\\ e_1,\dots,e_n\end{array} & \begin{array}{c}\mathbb{R}^n\\ x\end{array} \arrow[r, "A"] \arrow[d, "V^T"] & \begin{array}{c}\mathbb{R}^m\\ b = Ax\end{array} & \begin{array}{c}\text{standard}\\ \text{basis}\\ e_1,\dots,e_m\end{array} \\
\begin{array}{c}V\ \text{basis}\\ \{v_1,\dots,v_n\}\end{array} & \begin{array}{c}\mathbb{R}^n\\ y := V^T x\end{array} \arrow[r, "\Sigma"] & \begin{array}{c}\mathbb{R}^m\\ \Sigma y = b'\end{array} \arrow[u, "U"] & \begin{array}{c}U\ \text{basis}\\ \{u_1,\dots,u_m\}\end{array}
\end{tikzcd}$$

8.4 Geometry of the SVD

Next we look at the geometry of the SVD. The key property of a linear transformation that you have seen so far is that it sends a plane (of any dimension) in the domain to a plane in the codomain of the transformation. For example, a two-dimensional plane can go to another two-dimensional plane or a line or a point. In this section we will look at what a linear transformation does to the unit sphere and we will see that all the components of the SVD pop out from this action.

The *sphere* of radius 1 in $\mathbb{R}^n$ (called the *unit sphere*) is

$$\mathbb{S}^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_n^2 = 1\}.$$

For example in $\mathbb{R}^2$, $\mathbb{S}^1 = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ is the unit circle and in $\mathbb{R}^3$, $\mathbb{S}^2 = \{(x_1, x_2, x_3) \in \mathbb{R}^2 : x_1^2 + x_2^2 + x_3^2 = 1\}$ is what we think of normally as the sphere of radius 1.

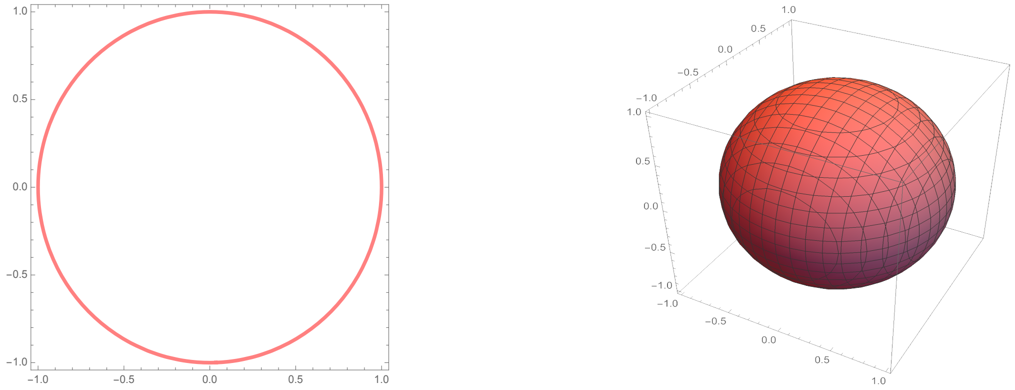


Figure 8.1: Unit circle in $\mathbb{R}^2$ and the unit sphere in $\mathbb{R}^3$

In $\mathbb{R}^2$ we are also familiar with an ellipse which has the equation

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$

for some $a, b > 0$. The quantities a and b are the lengths of the *semiaxes* of this ellipse. The higher dimensional analog of an ellipse is called a **hyperellipse** and has the equation

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \cdots + \frac{x_n^2}{a_n^2} = 1$$

The numbers $a_1, \ldots, a_n$ denote the lengths of the semiaxes of the hyperellipse.

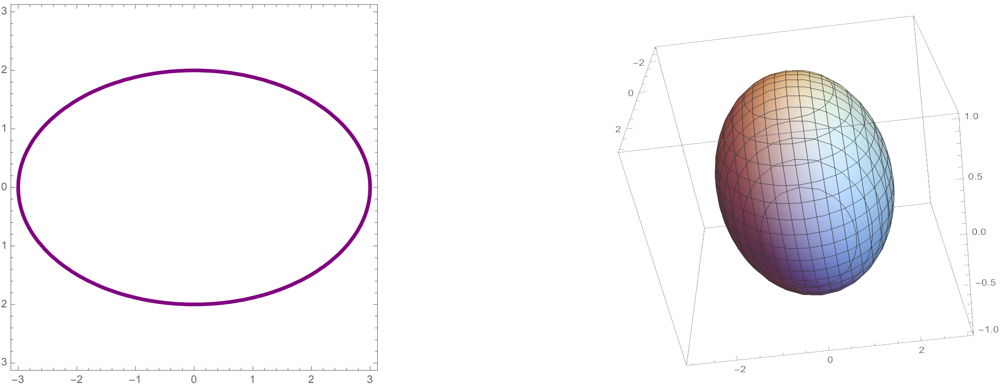


Figure 8.2: Eliipse in $\mathbb{R}^2$ with semiaxes of lengths 2 and 3 and hyperellipse in $\mathbb{R}^3$ with semiaxes of lengths $1, 2, 3$.

Let's look at the image of $\mathbb{S}^{n-1}$ under the linear transformation represented by $A \in \mathbb{R}^{m\times n}$. The key facts we will see are summarized in the following theorem. See Figures 8.3 and 8.4 for illustrations.

Theorem 8.4.1. *Let $A = U\Sigma V^\top$ be the SVD of $A \in \mathbb{R}^{m\times n}$ and let* $\operatorname{rank}(A) = r$. *Then*

1. *The image of $\mathbb{S}^{n-1}$ under A is a r-dimensional hyperellipse in $\mathbb{R}^m$, which we denote as $A(\mathbb{S}^{n-1})$.*

2. *The lengths of the semiaxes of the hyperellipse $A(\mathbb{S}^{n-1})$ are the singular values $\sigma_1, \ldots, \sigma_r$ of A. (In particular they are positive numbers.)*

3. *The first r left singular vectors of A, $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$, are the unit vectors along the semiaxes of $A(\mathbb{S}^{n-1})$.*

4. *The first r right singular vectors of A, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$, are the preimages in $\mathbb{S}^{n-1}$ of the semiaxes, $\sigma_1\boldsymbol{u}_1, \ldots, \sigma_r\boldsymbol{u}_r$, of the hyperellipse. In other words, $A\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i$ for $i = 1, \ldots, r$.*

We now see why the above theorem is true. Recall that the SVD allows us to think of the action of $A = U\Sigma V^\top$ as the composition of three linear transformations:

$$\begin{array}{ccccccc} \mathbb{R}^n & \longrightarrow & \mathbb{R}^n & \longrightarrow & \mathbb{R}^m & \longrightarrow & \mathbb{R}^m \\ \mathbf{x} & \mapsto & V^\top\mathbf{x} & \mapsto & \Sigma V^\top\mathbf{x} & \mapsto & U\Sigma V^\top\mathbf{x} \end{array}$$

The first and third transformations are given by orthonormal matrices $V^\top$ and U. So let's first look at what an orthonormal matrix does to the sphere $\mathbb{S}^{n-1}$.

Proposition 8.4.2. *If $Q \in \mathbb{R}^{n\times n}$ is an orthonormal matrix, then the linear transformation represented by Q preserves dot products, and hence also angles between vectors, and lengths of vectors.*

Proof. Suppose $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $Q \in \mathbb{R}^{n\times n}$ is an orthonormal matrix. Then

$$(Q\mathbf{v})^\top(Q\mathbf{w}) = \mathbf{v}^\top Q^\top Q\mathbf{w} = \mathbf{v}^\top I\mathbf{w} = \mathbf{v}^\top\mathbf{w}.$$

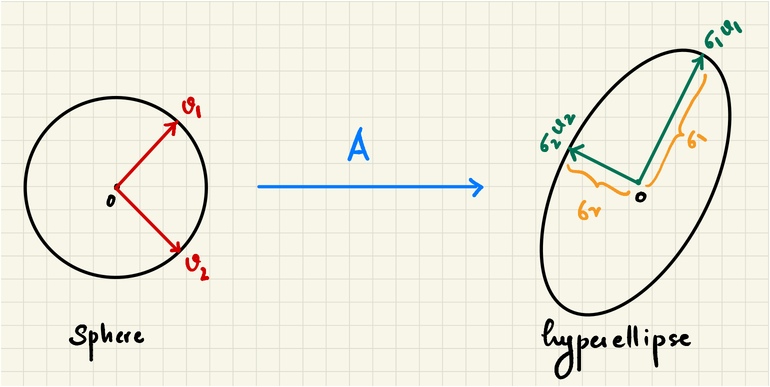


Figure 8.3: A sphere goes to a hyperellipse under the action of A.

In particular, Q preserves lengths of vectors since

$$||Q\mathbf{v}||^2 = (Q\mathbf{v})^\top(Q\mathbf{v}) = \mathbf{v}^\top\mathbf{v} = ||\mathbf{v}||^2.$$

Since $\mathbf{v}^\top\mathbf{w} = \|v\|\|w\|\cos(\theta)$ where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$, we also get that θ is still the angle between $Q\mathbf{v}$ and $Q\mathbf{w}$. □

This property shows that an orthonormal matrix sends the unit sphere to itself, and an orthonormal basis of $\mathbb{R}^n$ to another orthonormal basis of $\mathbb{R}^n$.

Corollary 8.4.3. *If $Q \in \mathbb{R}^{n\times n}$ is an orthonormal matrix, then $Q(\mathbb{S}^{n-1}) = \mathbb{S}^{n-1}$. If $\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ is an orthonormal basis of $\mathbb{R}^n$ then so is $\{Q\boldsymbol{b}_1, \dots, Q\boldsymbol{b}_n\}$.*

Recall that if $Q \in \mathbb{R}^{n\times n}$ is orthonormal, then $\det(Q) = \pm 1$. If $\det(Q) = 1$ then Q is in fact a rotation matrix. If $\det(Q) = -1$ then it is sometimes a reflection, but the general interpretation is more complicated. Check for instance that the matrix representing reflection across the x-axis in $\mathbb{R}^2$ is an orthonormal matrix with determinant -1. For the purposes of intiution you can think of Q as a rotation matrix for what we are going to do next, but this is not strictly correct.

Let's now see what the three linear transformations given by the SVD of A are doing geometrically. We will assume that $m \geq n$. The discussion below is summarized in Figure 8.5.

- The first transformation is $V^\top : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ and $V^\top$ is orthonormal. This map sends $\mathbb{S}^{n-1}$ to itself, i.e., $V^\top\mathbb{S}^{n-1} = \mathbb{S}^{n-1}$. The standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ in the codomain $\mathbb{R}^n$ came from the orthonormal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ in $\mathbb{R}^n$ since $\mathbf{v}_i$ is sent to $V^\top\mathbf{v}_i = \mathbf{e}_i \in \mathbb{R}^n$ by $V^\top$. A vector $\mathbf{y}$ in the codomain came from $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{y} = V^\top\mathbf{x}$. This means that $\mathbf{y}$ gives the coordinates of $\mathbf{x}$ in the basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$.

- Next we have the linear transformation $\Sigma : \mathbb{R}^n \to \mathbb{R}^m$ which simply scales coordinates, namely

  $$\mathbf{y} = (y_1, \dots, y_n)^\top \in \mathbb{R}^n \mapsto \Sigma\mathbf{y} = (\sigma_1 y_1, \dots, \sigma_r y_r, 0, \dots, 0)^\top \in \mathbb{R}^m$$

  In particular, $\mathbf{e}_i \in \mathbb{R}^n \mapsto \Sigma\mathbf{e}_i = \sigma_i\mathbf{e}'_i \in \mathbb{R}^m$ where $\mathbf{e}'_1, \dots, \mathbf{e}'_m$ are the standard basis vectors in $\mathbb{R}^m$. Therefore, each standard basis vector $\mathbf{e}_i \in \mathbb{R}^n$ for $i = 1, \dots, r$ becomes the scaled standard basis vector $\sigma_i\mathbf{e}'_i \in \mathbb{R}^m$, while the rest of the standard basis vectors in $\mathbb{R}^n$ get mapped to the origin in $\mathbb{R}^m$. Geometrically, this sends the sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$ to the r-dimensional hyperellipse $\Sigma V^\top\mathbb{S}^{n-1} \subset \mathbb{R}^m$ in which the first r radii $\mathbf{e}_1, \dots, \mathbf{e}_r$ of the sphere $\mathbb{S}^{n-1}$ map to the semiaxes $\sigma_1\mathbf{e}'_1, \dots, \sigma_r\mathbf{e}'_r$ of the r-dimensional hyperellipse, and the rest of the radii $\mathbf{e}_i, i = r+1, \dots, n$, map to the origin. The semiaxes of this hyperellipse are in the directions of the first r standard basis vectors of $\mathbb{R}^m$.

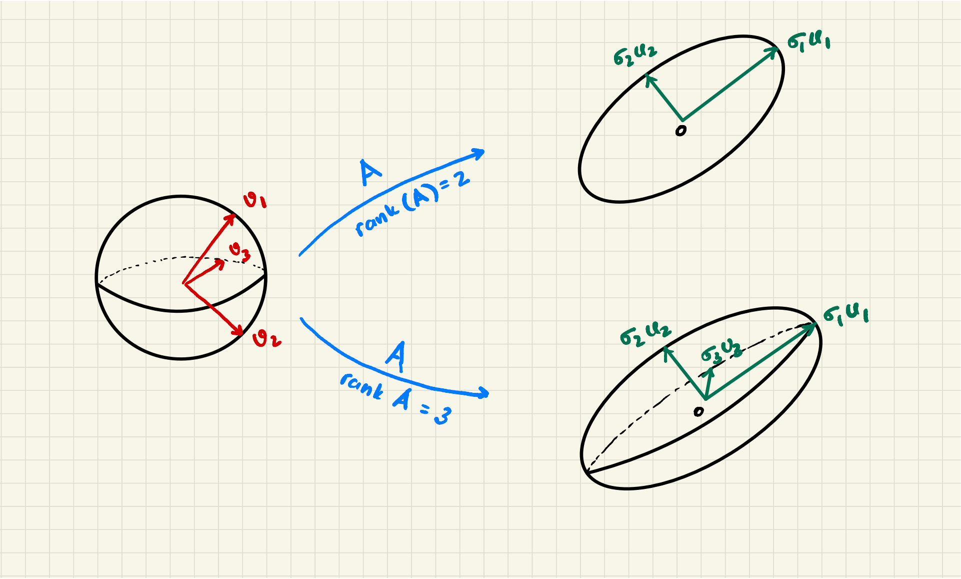


Figure 8.4: A sphere goes to hyperellipses of different dimension depending on the rank of A.

- Finally we apply the orthonormal matrix $U \in \mathbb{R}^{m\times m}$ to the r-dimensional hyperellipse $\Sigma V^\top \mathbb{S}^{n-1} \subset \mathbb{R}^m$. This sends a semiaxis $\sigma_i \mathbf{e}_i'$ to $U(\sigma_i \mathbf{e}_i') = \sigma_i U \mathbf{e}_i = \sigma_i \mathbf{u}_i \in \mathbb{R}^m$. Since angles and lengths are preserved by an orthonormal transformation, the whole hyperellipse $\Sigma V^\top \mathbb{S}^{n-1}$ maps to another r-dimensional hyperellipse $U\Sigma V^\top \mathbb{S}^{n-1}$ whose semiaxes are $\sigma_1 \mathbf{u}_1, \ldots, \sigma_r \mathbf{u}_r$. If U were are rotation, then we would be simply rotating $\Sigma V^\top \mathbb{S}^{n-1}$ so that the unit vector $\mathbf{e}_1, \ldots, \mathbf{e}_r \in \mathbb{R}^m$ rotate to $\mathbf{u}_1, \ldots, \mathbf{u}_r \in \mathbb{R}^m$. A formal proof of why a hyperellipse maps to a hyperellipse under an orthonormal transformation is omitted.

8.5 Application: Low Rank Approximations of a Matrix

Just like we have the 2-norm of a vector $\mathbf{x} \in \mathbb{R}^n$ defined as $\sqrt{x_1^2 + \cdots + x_n^2}$, matrices have a 2-norm. In fact, there are many norms for both vectors and matrices, but in this class we only consider the 2-norm.

Definition 8.5.1. The **2- norm** of $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, also known as the **spectral norm** of A, is

$$||A||_2 = \max_{\mathbf{x}\neq \mathbf{0}} \left\{ \frac{||A\mathbf{x}||_2}{||\mathbf{x}||_2} \right\}.$$

Since for any real number α, $||\alpha \mathbf{x}|| = |\alpha| ||\mathbf{x}||$, we get that

$$\frac{||A\alpha\mathbf{x}||_2}{||\alpha\mathbf{x}||_2} = \frac{|\alpha| ||A\mathbf{x}||_2}{|\alpha| ||\mathbf{x}||_2} = \frac{||A\mathbf{x}||_2}{||\mathbf{x}||_2}.$$

This means that we can always take $\mathbf{x}$ to have unit length and define

$$||A||_2 = \max_{||\mathbf{x}||_2 = 1} \left\{ ||A\mathbf{x}||_2 \right\}.$$

In words, the 2-norm of the matrix A is the maximum stretch that A applies to a unit vector.

From the geometry of the SVD we see that the the maximum stretch A applies to a unit vector in $\mathbb{R}^n$ is exactly σ_1, the largest singular value of A:

$$\|A\|_2 = \sigma_1$$

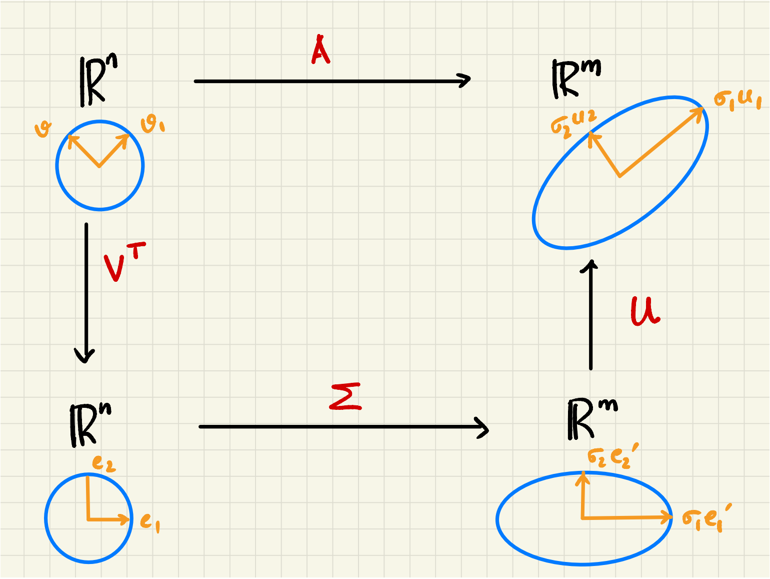


Figure 8.5: The SVD geometrically.

Recall that the SVD of A (with $\text{rank}(A) = r$) allows one to write A as a sum of r rank 1 matrices:

$$A = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_r^\top \end{bmatrix} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^\top$$

This rank one decomposition allows us to find low rank approximations to A. A good approximation of A must be a matrix that is "close" to A and we measure closeness with the matrix norm. In many applications, especially in engineering, one needs to find a low rank matrix that is close to the matrix we have. We will see some applications of this in a bit. How do we find the closest rank k matrix to A for some some $k < r$?

Definition 8.5.2. Let $\text{rank}(A) = r$. For $k < r$, define the **rank k approximation to A** as the k^{th} partial sum

$$A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \cdots \sigma_k \mathbf{u}_k \mathbf{v}_k^\top = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

Check for yourself that $\text{rank}(A_k) = k$. We now see that A_k is the "best" approximation of A by a matrix of rank k.

Theorem 8.5.3. (**Eckart-Young Theorem**) *The closest rank k matrix to A (in 2-norm) is the matrix*

$$A_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top.$$

More formally, if B is any $m \times n$ matrix of rank k, then

$$||A - B||_2 \geq ||A - A_k|| = \sigma_{k+1}.$$

We do not prove this theorem in this course but it is easy to see that $||A - A_k||_2 = \sigma_{k+1}$.

$$A - A_k = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top - \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = \sum_{i=k+1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = \begin{bmatrix} \mathbf{u}_{k+1} & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sigma_{k+1} & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_{k+1}^\top \\ \vdots \\ \mathbf{v}_r^\top \end{bmatrix}$$

This is the SVD of $A - A_k$ and so $||A - A_k||_2$ is the largest singular value of this matrix, which is σ_{k+1}.

We illustrate the above results on Example 8.2.2 from before.

Example 8.5.4. Recall that

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\hat{U}} \underbrace{\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{\hat{\Sigma}} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{\hat{V^\top}}$$

$$= 3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$$

Since the first column of A is $\mathbf{0}$ and the other columns are linearly independent, Col(A) is three-dimensional and hence A sends $\mathbb{R}^4$ to $\mathbb{R}^3$. In particular, $(x_1, x_2, x_3, x_4)^\top$ is sent by A to $(x_2, 2x_3, 3x_4, 0)^\top$. Therefore, we can draw the image of A in $\mathbb{R}^3$ by projecting onto the first three coordinates of a vector in $\mathbb{R}^4$.

By the Eckart-Young theorem, the closest rank 1 matrix to A is

$$A_1 = 3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since rank$(A_1) = 1$, the image of the unit sphere in $\mathbb{R}^4$ under A_1 is a line segment (1-dimensional hyperellipse) with semiaxis $3(0, 0, 1, 0)^\top$. Since we are projecting onto the first three coordinates, we can think of this as the line segment in $\mathbb{R}^3$ with semiaxis $3\mathbf{e}_3$.

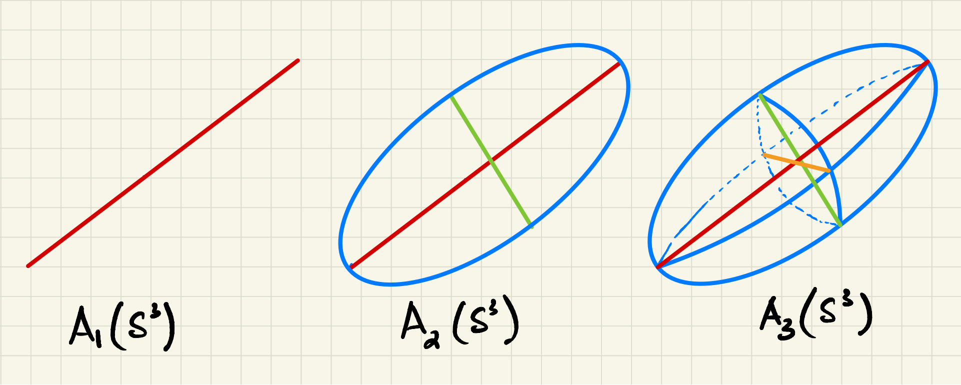


Figure 8.6: The low rank approximations of A build up the image hyperellipse of A one axis at a time from the longest to the shortest.

The closest rank 2 matrix to A is

$$A_2 = 3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This matrix sends the unit sphere in $\mathbb{R}^4$ to a 2-dimensional hyperellipse with semiaxes $3(0,0,1,0)^\top$ and $2(0,1,0,0)^\top$. Again projecting onto the first three coordinates, this is an ellipse with semiaxes $3\mathbf{e}_3$ and $2\mathbf{e}_2$.

The full matrix A is a rank 3 linear transformation from $\mathbb{R}^4$ to $\mathbb{R}^4$ and the image of the unit sphere in $\mathbb{R}^4$ under A is the 3-dimensional hyperellipse with semiaxes $3(0,0,1,0)^\top, 2(0,1,0,0)^\top$ and $(1,0,0,0)$. Projecting onto the first three coordinates we get the last picture in Figure 8.6.

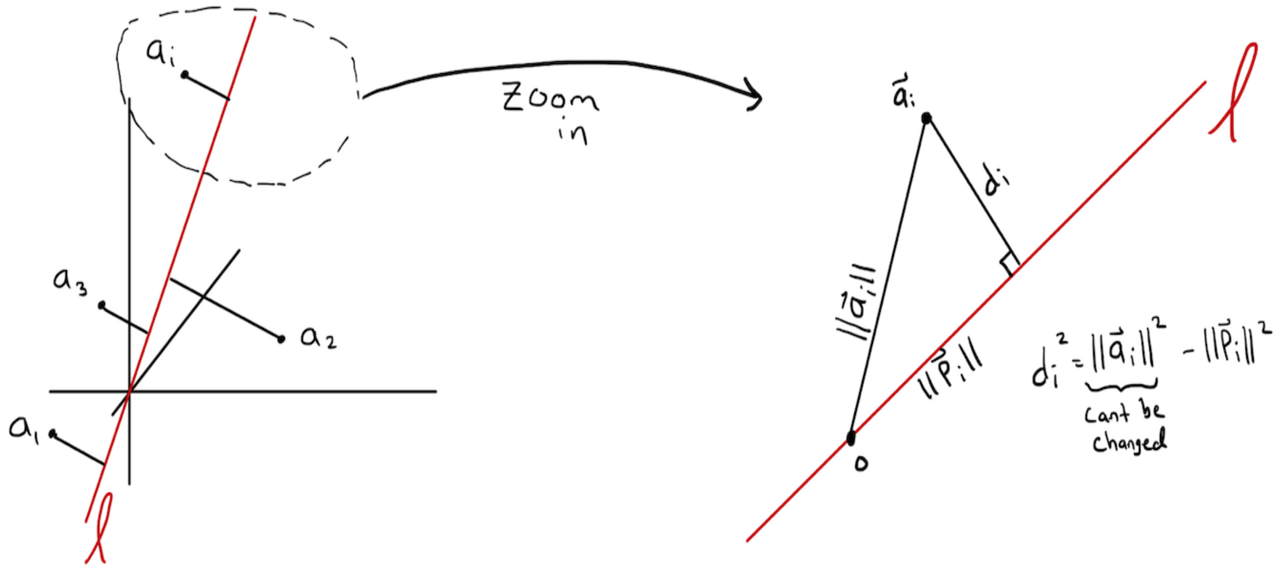
Geometrically, note that as we take higher and higher rank approximations of A from the rank one decomposition given by the SVD we are building up the image hyperellipse $A(\mathbb{S}^{n-1})$ one axis at a time, starting with the longest axis and working our way up in increasing order of axis length. We can also think of this process as finding better and better approximations of the image hyperellipse by lower dimensional ellipses.

8.6 Best Fit k-Planes

In linear regression we learned how to find the best fit line to a collection of points in $\mathbb{R}^2$ where by best fit we meant a line that minimized the sum of squares of the vertical distances from the points to the line. In this section we will see how the SVD can be used to find best fit planes to data points where now we will be minimizing the sum of squares of perpendicular distances from the points to the plane.

Question 8.6.1. Suppose we have m data points $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^n$. What is the "best fit k-dimensional subspace" to these data points?

We will assume that the m data points are the rows of a matrix $A \in \mathbb{R}^{m \times n}$. When $k = 1$ we are looking for the best fit line ℓ to the given data points, i.e., the line through the origin that minimizes the sum of squared orthogonal distances of $\mathbf{a}_1, \dots, \mathbf{a}_m$ to the line.



By Pythagorus' Theorem, the squared distance of $\mathbf{a}_i$ from ℓ is $d_i^2 = ||\mathbf{a}_i||^2 - ||\mathbf{p}_i||^2$, where $\mathbf{p}_i = \text{proj}_l \mathbf{a}_i$ is the (orthogonal) projection of $\mathbf{a}_i$ onto ℓ. From this we can conclude that in order to *minimize* $\sum d_i^2$, we need to *maximize* $\sum ||\mathbf{p}_i||^2$, since the norms $||\mathbf{a}_i||^2$ are fixed.

Question 8.6.2. Which line through the origin maximizes the value $\sum ||\mathbf{p}_i||^2$?

If $\mathbf{v}$ is a unit vector on the best fit line ℓ and θ is the angle between $\mathbf{a}_i$ and $\mathbf{v}$ then

$$\mathbf{a}_i^\top \mathbf{v} = ||\mathbf{a}_i|| ||\mathbf{v}|| \cos\theta = ||\mathbf{a}_i|| \cos\theta$$

which shows that $\|\mathbf{p}_i\| = |\mathbf{a}_i^\top \mathbf{v}|$, and $||\mathbf{p}_i||^2 = |\mathbf{a}_i^\top \mathbf{v}|^2$. Recalling that the $\mathbf{a}_i$ were the rows of the matrix A, we have

$$A\mathbf{v} = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix} \mathbf{v} = \begin{bmatrix} \mathbf{a}_1^\top \mathbf{v} \\ \vdots \\ \mathbf{a}_n^\top \mathbf{v} \end{bmatrix}$$

and hence,

$$\sum_{i=1}^{m} ||\mathbf{p}_i||^2 = ||A\mathbf{v}||^2$$

This translates our original question to a new one.

Question 8.6.3. Which unit vector $\mathbf{v} \in \mathbb{R}^n$ maximizes the value $||A\mathbf{v}||^2$?

We just saw the answer to this question in the previous section. Recalling that

$$||A||_2 = \max_{||\mathbf{x}||_2=1} \Big\{ ||A\mathbf{x}||_2 \Big\} = \sigma_1$$

and that the maximum is attained by the first right singular vector $\mathbf{v}_1$ of A, we conclude that $\mathbf{v}_1$ maximizes $||A\mathbf{v}||^2$. We summarize our finding in the following proposition.

Proposition 8.6.4. *If $A = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_m^\top \end{bmatrix} \in \mathbb{R}^{m \times n}$, then the best fit line through the origin to the rows of A is $\ell = \mathrm{Span}\{\boldsymbol{v}_1\}$, where $\boldsymbol{v}_1$ is the first right singular vector of A.*

Finding best fit planes to $\mathbf{a}_1, \ldots, \mathbf{a}_m$ proceeds similarly. We leave out the details of the proof, but state the generalization of the calculation we just did.

Theorem 8.6.5. *Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$ be the rows of the matrix $A \in \mathbb{R}^{m \times n}$ and suppose $\mathrm{rank}(A) = r$. For $1 \leq k \leq r$, let $V_k = \mathrm{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$ where $\boldsymbol{v}_i$ denotes the ith right singular vector of A. Then the best fit k-dimensional subspace to the rows of A is V_k. By best fit we mean the plane that minimizes the sum of squares of orthogonal distances from the points to the plane.*

By transposing A, we can obtain a similar theorem for the columns of A. Observe that if $A = U\Sigma V^\top$ then $A^\top = V\Sigma^\top U^\top$. Since the rows of $A^\top$ are the columns of A, we can directly apply the previous theorem to conclude the following.

Theorem 8.6.6. *Let $U_k = \mathrm{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\} \subseteq \mathbb{R}^m$ where $\boldsymbol{u}_i$ denotes the ith left singular vector of A. Then the best fit k-subspace to the columns of A is U_k.*

You should pause for a moment and let this sink it. We have seen many reasons why the singular value decomposition of A is important. The singular values alone tell us a tremendous amount of information, but we now have more. Mainly that the left and right singular vectors give the best fit k-dimensional subspaces to the columns (resp. rows) of A. Before we illustrate these results on Example 8.2.2, we state another important result that links the best fit k-planes to the rank one decomposition of A from the SVD.

Proposition 8.6.7. *Suppose $A = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ and $A_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$. The rows of A_k are the projections of the rows of A onto the subspace $V_k = \mathrm{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$.*

Proof. Let $\mathbf{a}$ be a row of A. Since the right singular vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ form an orthonormal basis for V_k, from homework we know that the projection of $\mathbf{a}$ onto V_k is

$$\mathrm{proj}_{V_k}(\mathbf{a}) = \mathbf{v}_1\mathbf{v}_1^\top \mathbf{a} + \mathbf{v}_2\mathbf{v}_2^\top \mathbf{a} + \cdots + \mathbf{v}_k\mathbf{v}_k^\top \mathbf{a}$$

Written as a row vector,

$$\mathrm{proj}_{V_k}(\mathbf{a})^\top = (\mathbf{a}^\top\mathbf{v}_1)\mathbf{v}_1^\top + (\mathbf{a}^\top\mathbf{v}_2)\mathbf{v}_2^\top + \cdots + (\mathbf{a}^\top\mathbf{v}_k)\mathbf{v}_k^\top.$$

Therefore, the matrix whose *rows* are the projections of the rows of A (the $\mathbf{a}_i$) onto V_k is:

$$\begin{bmatrix} \sum_{i=1}^k (\mathbf{a}_1^\top\mathbf{v}_i)\mathbf{v}_i^\top \\ \sum_{i=1}^k (\mathbf{a}_2^\top\mathbf{v}_i)\mathbf{v}_i^\top \\ \vdots \\ \sum_{i=1}^k (\mathbf{a}_m^\top\mathbf{v}_i)\mathbf{v}_i^\top \end{bmatrix} = \sum_{i=1}^k \Big(A\mathbf{v}_i\Big)\mathbf{v}_i^\top = \sum_{i=1}^k \Big(\sigma_i\mathbf{u}_i\Big)\mathbf{v}_i^\top = \sum_{i=1}^k \sigma_i\mathbf{u}_i\mathbf{v}_i^\top = A_k.$$

□

Example 8.6.8. Recall that

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\hat{U}} \underbrace{\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\hat{\Sigma}} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{\widehat{V^\top}}$$

$$= 3\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$$

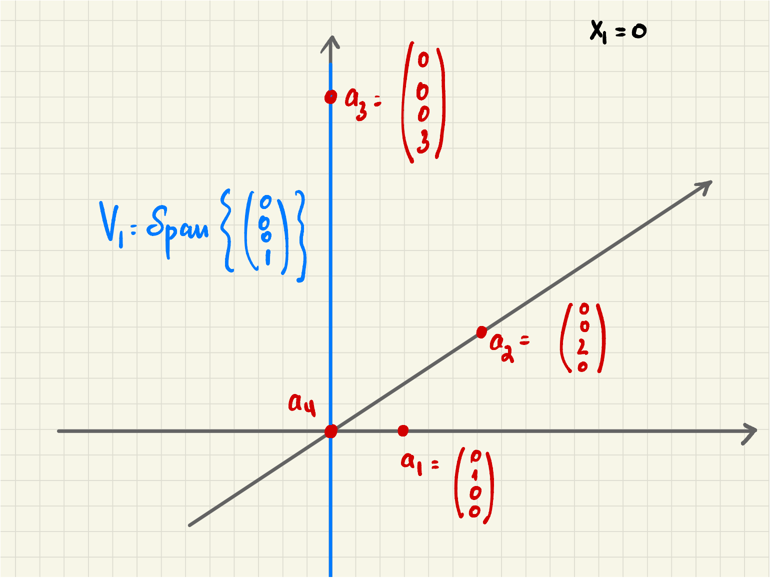


Figure 8.7: V_1 is the best fit line to the rows of A.

All rows of A have first coordinate 0 and so we can draw them in $\mathbb{R}^3$ by dropping the first coordinate. By the above theorem, the best fit line, plane and 3-dimensional plane to the 4 data points are:

- $V_1 = \mathrm{Span}\{(0,0,0,1)\}$

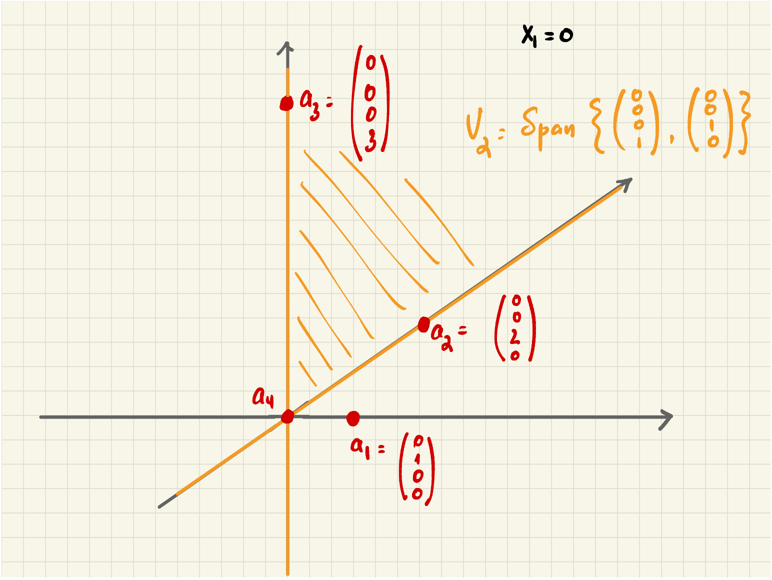

$x_1 = 0$
$a_3 = \begin{pmatrix}0\\0\\0\\3\end{pmatrix}$
$V_2 = \text{Span}\left\{\begin{pmatrix}0\\0\\0\\1\end{pmatrix}, \begin{pmatrix}0\\0\\1\\0\end{pmatrix}\right\}$
$a_2 = \begin{pmatrix}0\\0\\2\\0\end{pmatrix}$
a_4
$a_1 = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}$

Figure 8.8: V_2 is the best fit plane to the rows of A.

- $V_2 = \text{Span}\{(0,0,0,1),(0,0,1,0)\}$
- $V_3 = \text{Span}\{(0,0,0,1),(0,0,1,0),(0,1,0,0)\}$

The matrix

$$A_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top = 3 \begin{bmatrix}0\\0\\1\\0\end{bmatrix} \begin{bmatrix}0 & 0 & 0 & 1\end{bmatrix} = \begin{bmatrix}0 & 0 & 0 & 0\\0 & 0 & 0 & 0\\0 & 0 & 0 & 3\\0 & 0 & 0 & 0\end{bmatrix}$$

Check that the projections of $\mathbf{a}_1, \ldots, \mathbf{a}_4$ on V_1 are exactly the rows of A_1. Indeed, $\mathbf{a}_1$ and $\mathbf{a}_2$ project to $\mathbf{0}$, whereas $\mathbf{a}_3, \mathbf{a}_4$ project to themselves because they are already on V_1. See Figure 8.7.

The matrix

$$A_2 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^\top = 3 \begin{bmatrix}0\\0\\1\\0\end{bmatrix} \begin{bmatrix}0 & 0 & 0 & 1\end{bmatrix} + 2 \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \begin{bmatrix}0 & 0 & 1 & 0\end{bmatrix} = \begin{bmatrix}0 & 0 & 0 & 0\\0 & 0 & 2 & 0\\0 & 0 & 0 & 3\\0 & 0 & 0 & 0\end{bmatrix}.$$

Projections of the rows of A on V_2 are the following: $\mathbf{a}_1$ projects to $\mathbf{0}$, $\mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ project to themselves because they lie in V_2. Again, the rows of A_2 are the projections of $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ onto V_2. See Figure 8.8.

8.7 Application: Principal Component Analysis (PCA)

The theory we have developed in the last section is the basis of PCA, which identifies the principal directions of variance of data. Suppose our data consists of points $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ and $A_0 \in \mathbb{R}^{m \times n}$ is the matrix whose columns are $\mathbf{a}_1, \ldots, \mathbf{a}_n$. The idea of PCA is to project this data into a lower dimensional space (dimension reduction) while capturing as much of the variation present in the data. The principal components are uncorrelated, and ordered so that the first few retain most of the variation present in the original data set.

The underlying method is to find the best-fit planes to the data as we have done in the previous section. Recall that if the data is sitting in the columns of a matrix, then the best-fit k-plane to the data is $U_k = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ where $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are the left singular vectors of the matrix.

Since the best-fit planes all pass through the origin in $\mathbb{R}^m$, it makes sense to first center the data so that its mean is at the origin.

Step 1: Let $A \in \mathbb{R}^{m\times n}$ be the matrix obtained by subtracting the mean of each row of A_0 from all components of that row. Then $A\mathbf{1} = \mathbf{0}$ which means that the average of the columns of A (data point) is the origin.

Step 2: Use the SVD of A to find $\mathbf{u}_1, \dots, \mathbf{u}_r$ where $r = \text{rank}(A)$. The vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$ are called the **principal components** of the data points. They capture orthogonal directions in which the data varies starting from the most prominent direction of variance, then the next, the next and so on.

Recall the following results as well:

- $U_k = \text{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is the best fit k-plane to the columns of A.
- The columns of A_k are the projections of the the data onto U_k.

In statistics, $\mathbf{u}_1, \dots, \mathbf{u}_m$ are computed as the eigenvectors of $\frac{AA^\top}{n-1}$ which is called the *sample covariance matrix*. We won't get into the statistics behind PCA in this class, so we'll just focus on the SVD of A as we have learned in this class. Note that if m is small compared to n, then $AA^\top$ is a small matrix, of size $m \times m$.

Example 8.7.1. (from Strang Section 7.3) Suppose we have math scores and history scores of six students. Our data points are of the form (x_i, y_i) where x_i denotes the math score of student i and y_i denotes the history score of student i. Centering our data we get

$$A = \begin{bmatrix} 3 & -4 & 7 & 1 & -4 & -3 \\ 7 & -6 & 8 & -1 & -1 & -7 \end{bmatrix}$$

```
julia> using LinearAlgebra

julia> A = [3 -4 7 1 -4 -3; 7 -6 8 -1 -1 -7]
217Array{Int64,2}:
 3  -4  7   1  -4  -3
 7  -6  8  -1  -1  -7

julia> svd(A)
SVD{Float64,Float64,Array{Float64,2}}
U factor:
217Array{Float64,2}:
 -0.560629  -0.828067
 -0.828067   0.560629
singular values:
2-element Array{Float64,1}:
 16.870954927874198
  3.9205713641811695
Vt factor:
217Array{Float64,2}:
 -0.443268   0.427416   -0.625272   0.015852  0.182004   0.443268
  0.367344  -0.0131368  -0.334502  -0.354208  0.701847  -0.367344
```

This tells us that the principal components are:

$$\mathbf{u}_1 = \begin{pmatrix} -0.5606 \\ -0.8280 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} -0.8280 \\ 0.5606 \end{pmatrix}$$

and the maximum variation of the data is along $U_1 = \text{Span}\{\mathbf{u}_1\}$ and the next most variation in a direction orthogonal to U_1 is along $\text{Span}\{\mathbf{u}_2\}$.

Let's check if we would have gotten the same answer if we had worked with the sample covariance matrix.

```
julia> transpose(A)
617Transpose{Int64,Array{Int64,2}}:
  3   7
 -4  -6
  7   8
  1  -1
 -4  -1
 -3  -7

julia> B = A*transpose(A)
217Array{Int64,2}:
 100  125
 125  200

julia> eigvals(B)
2-element Array{Float64,1}:
  15.370879821637395
 284.62912017836265

julia> eigvecs(B)
217Array{Float64,2}:
 -0.828067  0.560629
  0.560629  0.828067

julia> C = B/5
217Array{Float64,2}:
 20.0  25.0
 25.0  40.0

julia> eigvals(C)
2-element Array{Float64,1}:
  3.0741759643274786
 56.92582403567252

julia> eigvecs(C)
217Array{Float64,2}:
 -0.828067  0.560629
  0.560629  0.828067
```

In the above calculations, note that the eigenvectors of $AA^\top$ and $\frac{AA^\top}{5}$ are the same and they are precisely $\mathbf{u}_1$ and $\mathbf{u}_2$. The order is flipped since the order in which the eigenvalues are written in Julia is from the smallest to largest.

See the video on PCA by Steve Brunton to see some large data sets in action: https://www.youtube.com/watch?v=VqjJ5YYt78Y. He also explains all the normalization that is done in statistics.

Here is another example that illustrates the heart of PCA. Suppose we look at car sales in the United States and have a data set showing the preferences of n customers for m types of cars. The data is arranged as the columns of a matrix A where columns represent people and rows represent cars. Column j tells you the probability that customer j will buy car i. We expect car sales to be driven by a few dominant factors that are independent like age, income, family size etc, say r factors. For each factor we have a vector that tells you the preference for a type of car when only that factor is taken into account. These vectors form the columns of a matrix U of size $m \times r$. The SVD writes A as $A = U(\Sigma V^\top)$ where the jth column of $\Sigma V^\top$ gives the weights for each factor for customer j. See Figure 8.9 for a visualization of these matrices.

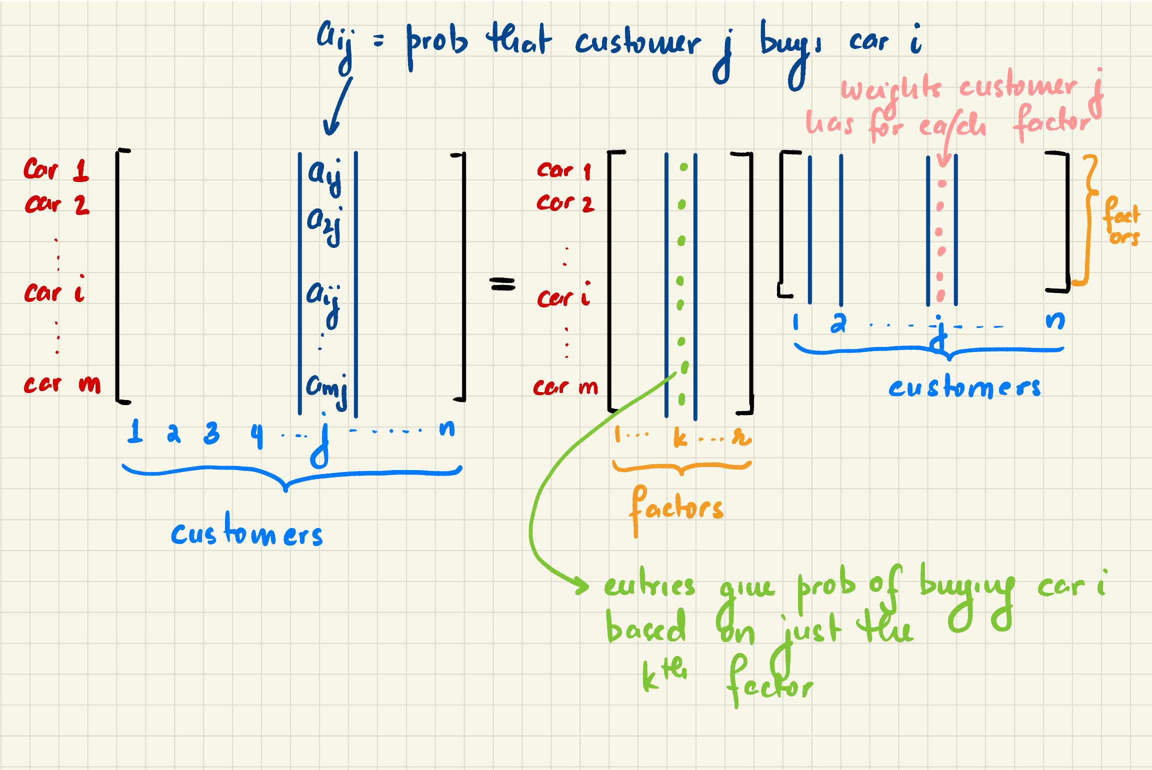


Figure 8.9: The SVD identifies the important (independent) factors that determine the data.

The Netflix problem falls in this general class of problems. It's more complicated than what we have seen so far, but has the same spirit. See the following video by Steve Brunton https://www.youtube.com/watch?v=sooj-_bXWgk. There are many many other examples of PCA. See some of the other videos by Brunton https://www.youtube.com/watch?v=gXbThCXjZFM&list=PLMrJAkhIeNNSVjnsviglFoY2nXildDCcv

Chapter 9

Error Correcting Codes

In this chapter we will see a new type of vector space, namely a vector space over a *finite field.* Such vector spaces lead to important applications in areas such as coding theory.

9.1 Vector spaces over finite fields

Recall that $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$, the set of integers mod 5, is the set of all remainders obtained from all integers after division by 5. All integers have a representative in $\mathbb{Z}_5$, namely the remainder you get when you divide the integer by 5. By this definition, 0 represents all integer multiples of 5. We say that two integers a and b are *congruent mod* 5 if they have the same remainder on division by 5 or equivalently, $a - b$ is a multiple of 5. For example, 1 and 6 are congruent mod 5, written as $1 \equiv 6 \bmod 5$. Similarly you could have $\mathbb{Z}_k$ for any integer k. Note that $\mathbb{Z}_1 = \mathbb{Z}$.

In this chapter we focus on $\mathbb{Z}_2 = \{0, 1\}$, the integers mod 2, or *binary numbers.* How do we add and multiply in $\mathbb{Z}_2$? The only thing to remember is that we should do all operations as we normally would but every time you get a number different from 0 or 1 you should replace it by its representative in $\mathbb{Z}_2$.

Addition:

$$0 + 0 = 0,\ \ 0 + 1 = 1 + 0 = 1,\ \ 1 + 1 = 0.$$

The reason why $1 + 1 = 0$ is because normally $1 + 1 = 2$ and the representative of 2 in $\mathbb{Z}_2$ is 0.

Multiplication: We can only use elements from $\mathbb{Z}_2$ and we should stay in $\mathbb{Z}_2$:

$$0 \cdot 0 = 0,\ \ 0 \cdot 1 = 0,\ \ 1 \cdot 1 = 1$$

The multiplicative inverse of 1 is 1 since $1 \cdot 1 = 1$.

All other rules of addition and multiplication are the same as in $\mathbb{R}$: addition and multiplication are both commutative and associative and so on.

A set that has all these properties is called a *field.* We are being vague about "all these properties". The set of real numbers $\mathbb{R}$, the set of rational numbers $\mathbb{Q}$ and the set of complex numbers $\mathbb{C}$ are all fields. $\mathbb{Z}_2$ is called the finite field of two elements. In general, for any prime number p, $\mathbb{Z}_p$ is the finite field of p elements. In other words $\mathbb{Z}_p$ has the properties of $\mathbb{R}$ and $\mathbb{C}$. The word "finite" is used since the field $\mathbb{Z}_p$ has only finitely many (i.e., p) elements.

Just as with $\mathbb{R}$ and $\mathbb{C}$ we can have vector spaces over $\mathbb{Z}_2$. By this we mean sets that satisfy all the rules of being a vector space and where linear combinations have coefficients from $\mathbb{Z}_2$ and scalar multiplication only uses elements of $\mathbb{Z}_2$.

Example 9.1.1. What is $(\mathbb{Z}_2)^3$? Remember $\mathbb{R}^3$ is the set of all triples of real numbers (a, b, c). So $(\mathbb{Z}_2)^3$ must be all triples of elements from $\mathbb{Z}_2$.

$$(\mathbb{Z}_2)^3 = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}$$

Note that these are precisely the corners of a special cube of side length 1 (called the *unit cube*) in $\mathbb{R}^3$.

Check that $(\mathbb{Z}_2)^3$ is a vector space over $\mathbb{Z}_2$ just like $\mathbb{R}^3$ is a vector space over $\mathbb{R}$:

$$\text{If } \mathbf{a}, \mathbf{b} \in (\mathbb{Z}_2)^3 \text{ and } \alpha, \beta \in \mathbb{Z}_2 \text{ then } \alpha\mathbf{a} + \beta\mathbf{b} \in (\mathbb{Z}_2)^3.$$

For example, if $\mathbf{a} = (1,0,0), \mathbf{b} = (1,0,1), \alpha = 0, \beta = 1$ then $0(1,0,0) + 1(1,0,1) = (0,0,0) + (1,0,1) = (1,0,1)$ which is still in $(\mathbb{Z}_2)^3$. Try a few more examples.

$(\mathbb{Z}_2)^4 = \{(0,0,0,0),(1,0,0,0),(0,1,0,0),\ldots,(1,1,0,0),\ldots,(0,1,1,1),(1,1,1,1)\}$ which are precisely all the corners of the unit cube in $\mathbb{R}^4$. It has $2^4 = 16$ elements. This is again a vector space over $\mathbb{Z}_2$.

In general $(\mathbb{Z}_2)^n$ is a vector space over $\mathbb{Z}_2$ with 2^n elements.

Exercise 9.1.2. What is a basis for $(\mathbb{Z}_2)^3$? In general, $(\mathbb{Z}_2)^n$? What are the dimensions of these vector spaces?

Since $(\mathbb{Z}_2)^n$ is a vector space, it can have *subspaces.* These would be subsets of $(\mathbb{Z}_2)^n$ that themselves form a vector space over $\mathbb{Z}_2$. We will see subspaces in the next section.

9.2 Hamming Code

The following material is taken from the book *Thirty Three Miniatures* by J. Matouşek.

Suppose we wish to transmit a message as a string $\mathbf{v}$ of 0s and 1s. The transmission channel could introduce errors. For example you might send the string $\mathbf{v} = 1011$ but the string your buddy receives is $\mathbf{w} = 1001$, which has one error. We assume that the probability of many errors is small, say the probability of two errors is very small, but there is a chance of one error. In general, the probability of k errors might be very small, but there is a significant chance of $k-1$ or less errors. Error correcting codes will pad your original message with extra digits that can help you correct errors. Below we see how this works.

Example 9.2.1. Suppose the probability of two errors is very small but one error is quite possible. Then if we wish to send the string 1011, we could make the rule that we will triple every digit and send 111000111111. Then if your buddy receives 110000111111, they will know that there is an error in the first digit and the message really is 1011. Of course there might be more errors but since the chance of two errors is small, this assessment seems reasonable. The question is, do you really need to triple every digit? Can you be more economical?

One of the best known error correcting codes is the **Hamming code** which was discovered in the 1950s.

Example 9.2.2. Here is what a Hamming code would do with a string $\mathbf{v} = abcd$ where $a, b, c, d \in \{0,1\}$. It would send $\mathbf{w} = abcdefg$ where

$$e = a + b + c \bmod 2, \;\; f = a + b + d \bmod 2, \;\; g = a + c + d \bmod 2$$

So if $\mathbf{v} = 1011$, the code would send $\mathbf{w} = 1011001$. We will see that this can correct one error.

Now here is where $\mathbb{Z}_2$ and $(\mathbb{Z}_2)^n$ come into the picture. Recall that the elements of $(\mathbb{Z}_2)^n$ are precisely all strings of length n from the *alphabet* $\{0,1\} = \mathbb{Z}_2$. An element of $(\mathbb{Z}_2)^n$ is called a *string* or *word* of length n (or with n-bits) from the alphabet $\mathbb{Z}_2$.

Definition 9.2.3.
1. A **code** of length n is any subset of $(\mathbb{Z}_2)^n$.
2. A **linear code** of length n is a subspace of $(\mathbb{Z}_2)^n$.

Example 9.2.4. Consider all 7-bit strings that the Hamming code in the previous example will produce starting with all 4-bit strings:

$$C = \{0000000, 0001011, 0010101, 0011110, 0100110, 0101101, 0110011, 0111000, 1000111,$$
$$1001100, 1010010, 1011001, 1100001, 1101010, 1110100, 1111111\}$$

C is a code since it is a subset of $(\mathbb{Z}_2)^7$.

We will now see that the code C is in fact a linear code. To prove this we need to argue that C is a subspace of $(\mathbb{Z}_2)^7$. How do we prove such a thing?

Remember that there are two ways of representing a subspace. We can either find a generating set (or basis) or write the set as the solutions to a finite number of linear equations. All this also works over $\mathbb{Z}_2$ as long as all calculations are done mod 2. Let's look at these methods for a subspace C of $(\mathbb{Z}_2)^n$.

1. **By basis**: Suppose G is a $k \times n$ matrix whose rows from a basis of C. Then all elements of C are linear combinations of the rows of G and so we can write C as

$$C = \{\mathbf{y}^\top G : \mathbf{y} \in (\mathbb{Z}_2)^k\}$$

We call G a *generator matrix* of C.

Example 9.2.5. In our example C, we can take

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

Check that every one of the 16 elements in C is a linear combination of the four rows of G and there are no more linear combinations. This is one way to see that C is a linear code but this is laborious.

2. **By linear equations**: Recall that the Hamming code satisfies the linear equations:

$$a + b + c \equiv e \bmod 2, \;\; a + b + d \equiv f \bmod 2, \;\; a + c + d \equiv g \bmod 2$$

If we bring all variables to the left then these equations read:

$$a + b + c - e \equiv 0 \bmod 2, \;\; a + b + d - f \equiv 0 \bmod 2, \;\; a + c + d - g \equiv 0 \bmod 2$$

However, in $\mathbb{Z}_2$, the negative of an element is itself. Therefore, $-e = e, -f = f, -g = g$ and the equations are

$$a + b + c + e \equiv 0 \bmod 2, \;\; a + b + d + f \equiv 0 \bmod 2, \;\; a + c + d + g \equiv 0 \bmod 2$$

Equivalently, C is the nullspace of the following matrix:

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

Check that the 16 elements in C satisfy $P\mathbf{x} = \mathbf{0}$. Can you see that there are no more solutions? This is a lot easier to check than the checking if all elements of C are combinations of the rows of G. The matrix P is called the *parity check matrix* of the code C. Note that the columns of P are exactly the nonzero elements of $(\mathbb{Z}_2)^3$.

To finish we need some coding theory terminology which will help us see that the Hamming code above can correct one error.

Definition 9.2.6. 1. The **Hamming distance** of two words $\mathbf{u}, \mathbf{v}$ in $(\mathbb{Z}_2)^n$ is the number of places in which they are different. Mathematcially:

$$d(\mathbf{u}, \mathbf{v}) = |\{i : u_i \neq v_i, \text{ for } i = 1, \ldots, n\}|$$

For example, $d(1011, 1001) = 1$ since the two words differ only in position 3, while $d(1011, 1000) = 2$.

It is useful to think of the Hamming distance geometrically as the smallest number of edges of the unit cube that you need to walk on to go from $\mathbf{u}$ to $\mathbf{v}$. The way to "walk" from a corner of the unit cube to another, is to start with the initial corner and move successively to a neighboring corner, which is a word that differs from the given word in exactly one digit. For example, we can "walk" from 1101 to 1000 by successively making the moves 1011 → 1001 → 1000. You could have also done 1011 → 1010 → 1000. There are other longer routes between 1011 and 1000 through the corners of the cube but the smallest number of steps needed is unique and this number of steps is the Hamming distance between 1101 and 1000, namely $d(1011, 1000) = 2$.

2. The **minimum distance** of a code C is the smallest distance between any two words in C. Mathematically,

$$d(C) := \min\{d(\mathbf{u}, \mathbf{v}) \,:\, \mathbf{u}, \mathbf{v} \in C,\ \ \mathbf{u} \neq \mathbf{v}\}$$

In our example C, $d(C) = 3$. Check that any two code words differ in at least 3 bits and there are pairs with Hamming distance exactly 3.

3. A code $C \subseteq (\mathbb{Z}_2)^n$ **corrects** t **errors** if for every $\mathbf{u} \in (\mathbb{Z}_2)^n$ there is at most one $\mathbf{v} \in C$ such that $d(\mathbf{u}, \mathbf{v}) \leq t$.

For example, our code C corrects one error if for every $\mathbf{u} \in (\mathbb{Z}_2)^7$ there is at most one string in C at distance 1 or 0 from $\mathbf{u} \in (\mathbb{Z}_2)^7$. Our example is indeed a one-error correcting code. Please check on a few examples. We will prove this shortly.

Theorem 9.2.7. *A code C corrects t errors if and only if $d(C) \geq 2t + 1$.*

We will prove the following special case which should help you see how to prove the general case. Also, everything below is about 1-correcting codes. The arguments generalize.

Theorem 9.2.8. *A code C corrects 1 error if and only if $d(C) \geq 3$.*

Proof. Suppose $d(C) \leq 2$. Then there are two code words $\mathbf{u}, \mathbf{v} \in C$ such that $d(\mathbf{u}, \mathbf{v}) \leq 2$. This means that either $\mathbf{u}$ and $\mathbf{v}$ are neighboring vertices of the unit cube or there is way to walk along the edges of the unit cube in $\mathbb{R}^n$ from $\mathbf{u}$ to $\mathbf{v}$ via a word $\mathbf{w}$ which is also a corner of the unit cube. In the first case, there is a code word within distance 1 from $\mathbf{u}$ and in the second case, the word $\mathbf{w}$ has two code words within distance 1 from it. Either way, C cannot correct 1 error by definition. This proves that if C corrects 1 error then $d(C) \geq 3$.

For the converse, suppose C is not 1-correcting. Then there is some $\mathbf{w} \in (\mathbb{Z}_2)^n$ such that there are two or more code words within distance 1 of it. Suppose $\mathbf{u}, \mathbf{v} \in C$ are two of these code words. Then by the same argument as above, we can walk from $\mathbf{u}$ to $\mathbf{v}$ via $\mathbf{w}$ in two steps. This in turn means that $d(C) \leq 2$. □

We now introduce the generalized Hamming code which is the family of codes in which our example code lives. We will see that they are all have $d(C) = 3$ and are hence 1-correcting codes.

Definition 9.2.9. Fix a positive integer l. The **generalized Hamming code** (for l) is the linear code in $(\mathbb{Z}_2)^n$ where $n = 2^l - 1$ with parity matrix P whose columns are all the nonzero elements of $(\mathbb{Z}_2)^l$. In particular, generalized Hamming codes are linear codes since they are solutions of $P\mathbf{x} = 0$.

Example 9.2.10. In our code C, $l = 3$. The code words are in $(\mathbb{Z}_2)^7$ and $7 = 2^3 - 1$. The parity matrix P has all the nonzero elements of $(\mathbb{Z}_2)^3$ as columns.

Theorem 9.2.11. *The generalized Hamming code C has $d(C) = 3$ and thus is a 1-error correcting code.*

Proof. We first note that for any linear code C,

$$d(C) = \min\{d(\mathbf{0}, \mathbf{u}) \,:\, \mathbf{u} \in C,\ \ \mathbf{u} \neq \mathbf{0}\}.$$

In other words, to compute the smallest distance between two code words, it is enough to compute the smallest distance between $\mathbf{0}$ and any code word. Suppose not. Then there are two nonzero code words $\mathbf{u}, \mathbf{w}$

whose distance is $d(C)$. Now consider $\mathbf{0} = \mathbf{u} - \mathbf{u}$ and $\mathbf{w} = \mathbf{v} - \mathbf{u}$. Since C is a subspace, $\mathbf{0} = \mathbf{u} - \mathbf{u}$ and $\mathbf{w} = \mathbf{v} - \mathbf{u}$ are also in C and distances don't change under subtraction, so $d(C) = d(\mathbf{0}, \mathbf{w})$.

To prove our theorem, we need to show that $d(C) \geq 3$ which by the above is same as showing that $d(\mathbf{0}, \mathbf{w}) \geq 3$ for every nonzero $\mathbf{w} \in C$. This is in turn is same as showing that every nonzero $\mathbf{w} \in C$ has at least 3 nonzero bits. The parity matrix now helps. We can show that no word $\mathbf{w}$ with at most 2 nonzero bits satisfies $P\mathbf{w} = 0$. If $\mathbf{w}$ had only one nonzero bit then $P\mathbf{w} = 0$ if and only if a column of P is $\mathbf{0}$, but this is not allowed in the definition of P. If $\mathbf{w}$ had two nonzero bits and $P\mathbf{w} = 0$ then two columns of P are the same which is also not true. Therefore we are done. □

Thus we see that our running example code C is 1-error correcting. To finish, let's see how one encodes and decodes messages using the Hamming code.

9.2.1 Encoding and Decoding Messages using the Hamming Code

How can we encode and decode given a 1-correcting linear code $C \subseteq (\mathbb{Z}_2)^n$ with generator matrix G of size $k \times n$ and parity check matrix P of size $(n-k) \times n$?

Encoding: Given a word $\mathbf{v} \in (\mathbb{Z}_2)^k$ we encode it as $\mathbf{w} = \mathbf{v}^\top G \in C$.

In homework you will argue that we can always choose G to look like $G = \begin{bmatrix} I_k & A \end{bmatrix}$. Therefore, the first k block of $\mathbf{w} = \mathbf{v}^\top G$ is just $\mathbf{v}$. More generally, for any G, if there is no error, then we can recover $\mathbf{v}$ by solving $\mathbf{w} = \mathbf{v}^\top G$ which has a unique solution since the rows of G are linearly independent. Check that with the G we had for our Hamming code in $(\mathbb{Z}_2)^7$, 1011 would be sent to $(1, 0, 1, 1)G = 1011001$ as we had before.

Decoding: Suppose we send the code word $\mathbf{w}$ and receive $\mathbf{w}'$. If at most one error has occurred we have $\mathbf{w}' = \mathbf{w}$ or $\mathbf{w}' = \mathbf{w} + \mathbf{e}_i$ for some $i \in \{1, 2, \ldots, n\}$. If $\mathbf{w}' = \mathbf{w}$ then $P\mathbf{w}' = 0$. If $\mathbf{w}' = \mathbf{w} + \mathbf{e}_i$ then $P\mathbf{w}' = P\mathbf{w} + P\mathbf{e}_i = P\mathbf{e}_i$ which is the ith column of P. Thus if there was at most one error, we can immediately tell if an error occurred and we see which bit was wrong, namely the ith bit was wrong and there is a unique correction.

In homework you will see that if you have a basis of any subspace stored as the rows of $G = \begin{bmatrix} I_k & A \end{bmatrix}$, then the subspace is the set of solutions of $\begin{bmatrix} -A^\top & I_{n-k} \end{bmatrix} \mathbf{x} = 0$.

Chapter 10

Vector Spaces over $\mathbb{C}$

In this final chapter we will see vector spaces over $\mathbb{C}$, the set of complex numbers. Many of the results we have seen so far for vectors and matrices over $\mathbb{R}$ will carry over to $\mathbb{C}$ once we have the correct definitions. The first step is understanding the basics of $\mathbb{C}$ itself. After introducing complex numbers, we look at complex vectors, then complex matrices and then the complex spectral theorem.

10.1 Complex Numbers

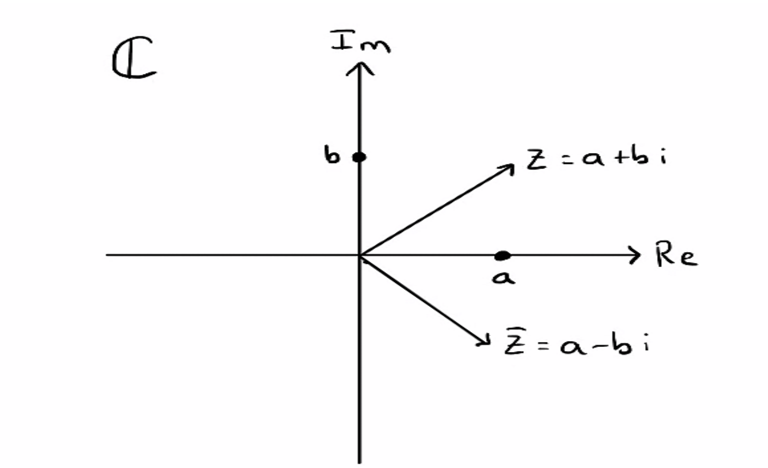
Recall that a complex number $z \in \mathbb{C}$ has the form $z = a + bi$ where $a, b \in \mathbb{R}$. The imaginary $i = \sqrt{-1}$ is defined to be $i^2 = -1$ or equivalently, the solution to $x^2 = -1$. We call a the *real part* of z, denoted $\text{Re}(z)$ and we call b the *imaginary part* of z, denoted $\text{Im}(z)$. We can add and multiply complex numbers:

$$(a + bi) + (c + di) = (a + c) + (b + d)i \quad \text{and} \quad (a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

Every complex number has a complex conjugate that we have seen before.

Definition 10.1.1. If $z = a + bi$ then the **complex conjugate** of z, denoted $\overline{z}$, is the complex number $\overline{z} = a - bi$.

Note that if $z \in \mathbb{R}$ then $\overline{z} = z$. We can visualize complex numbers (and their conjugates) in the complex plane with imaginary and real axes as follows:



The numbers $z = a + bi$ and $\overline{z} = a - bi$ work together as follows:

- $z + \overline{z} = (a + bi) + (a - bi) = 2a = 2\text{Re}(z) \in \mathbb{R}$
- $z\overline{z} = (a + bi)(a - bi) = a^2 + b^2 \in \mathbb{R}$
- For $z, w \in \mathbb{C}$, $\overline{zw} = \overline{z}\,\overline{w}$ and $\overline{z + w} = \overline{z} + \overline{w}$

The length (or **modulus**) of the complex number z is the real number

$$|z| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}.$$

It is the Euclidean length of the vector denoting z, in the complex plane. Also,

$$\frac{1}{z} = \frac{1}{a + bi} = \frac{a - bi}{(a + bi)(a - bi)} = \frac{a - bi}{a^2 + b^2} = \frac{\overline{z}}{z\overline{z}} = \frac{\overline{z}}{|z|^2} \in \mathbb{C}$$

and if $a^2 + b^2 = |z| = 1$ then $\frac{1}{z} = \overline{z}$.

Taking powers of complex numbers can be computationally difficult based on what we have so far. The **polar form** of a complex number greatly reduces this difficulty and will be important to learn. Given any $z \in \mathbb{C}$, setting $r = |z|$, we can write

$$z = |z|(\cos\theta + i\sin\theta) = r(\cos\theta + i\sin\theta) = re^{i\theta}$$

Example 10.1.2. Let $z = 3 - 2i$. We can see that $|z| = \sqrt{9 + 4} = \sqrt{13}$, hence

$$z = \sqrt{13}(\frac{3}{\sqrt{13}} - \frac{2}{\sqrt{13}}i)$$

This means that $\cos\theta = \frac{3}{\sqrt{13}}$ and $\sin\theta = \frac{-2}{\sqrt{13}}$ hence we can find θ via $\theta = \cos^{-1}\left(\frac{3}{\sqrt{13}}\right) = \sin^{-1}\left(\frac{-2}{\sqrt{13}}\right)$

Using series expansions, we get that $\cos\theta + i\sin\theta = e^{i\theta}$. Hence, if $z = r(\cos\theta + i\sin\theta)$ then $z = re^{i\theta}$ and

$$z^n = (re^{i\theta})^n = r^n e^{in\theta} = r^n(\cos n\theta + i\sin n\theta)$$

Moreover, if $|z| = 1$, then $r = 1$ and $z^n = e^{in\theta}$. That is, z^n is just z rotated by θ, n times. The polar form tells us the all important fact that when you multiply two complex numbers you **multiply the lengths and add the angles**.

Now that we have a grasp on the elements of $\mathbb{C}$, we can look at vector spaces over $\mathbb{C}$.

10.2 Complex Vectors and the Vector Space $\mathbb{C}^n$

Define a complex vector to be

$$\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^n$$

where $z_j = a_j + ib_j \in C$ for all j. Verify that $\mathbb{C}^n$ is a vector space over $\mathbb{C}$ with respect to addition and scalar multiplication where the set of scalars is $\mathbb{C}$:

$$\textbf{Addition:} \qquad \mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}, \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{C}^n \;\Rightarrow\; \mathbf{z} + \mathbf{w} = \begin{bmatrix} z_1 + w_1 \\ \vdots \\ z_n + w_n \end{bmatrix} \in \mathbb{C}^n$$

Scalar multiplication: $\quad a+bi \in \mathbb{C},\ z \in \mathbb{C}^n \ \Rightarrow\ (a+bi)\mathbf{z} = \begin{bmatrix}(a+bi)z_1\\ \vdots \\ (a+bi)z_n\end{bmatrix} \in \mathbb{C}^n$

The analog of the transpose of a real vector is the following.

Definition 10.2.1. Given $\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^n$, the **conjugate transpose** of $\mathbf{z}$ is

$$\overline{\mathbf{z}}^\top = \begin{bmatrix} \overline{z_1} & \cdots & \overline{z_n} \end{bmatrix}$$

Given any complex number z_i, recall that $\overline{z_i}z_i \in \mathbb{R}$. We can expand on this idea and use conjugate transposes to define norms of complex vectors. Given $\mathbf{z} \in \mathbb{C}^n$ check that $\overline{\mathbf{z}}^\top\mathbf{z} \in \mathbb{R}$. We define the **norm** of $\mathbf{z}$ to be the real number

$$||\mathbf{z}|| = \sqrt{\overline{\mathbf{z}}^\top \mathbf{z}}$$

A common notation for the conjugate transpose is $\mathbf{z}^* = \overline{\mathbf{z}}^\top$. We will adopt this notation from now on.

Note that if $z \in \mathbb{C}$ then $z^* = \overline{z}$. Therefore, the formula for the length of $z \in \mathbb{C}$ which is $|z| = \sqrt{\overline{z}z}$ is a special case of the formula for the length of a vector $\mathbf{z} \in \mathbb{C}^n$ which is $||\mathbf{z}|| = \sqrt{\overline{\mathbf{z}}^\top\mathbf{z}}$.

Example 10.2.2. Let $\mathbf{z} = \begin{bmatrix} 2-i \\ 3+5i \end{bmatrix}$ and $\mathbf{z}^* = \begin{bmatrix} 2+i & 3-5i \end{bmatrix}$. Then

$$\mathbf{z}^*\mathbf{z} = \begin{bmatrix} 2+i & 3-5i \end{bmatrix} \begin{bmatrix} 2-i \\ 3+5i \end{bmatrix} = |2-i|^2 + |3+5i|^2 = 5 + 34$$

and hence $\|\mathbf{z}\| = \sqrt{39}$.

Using conjugate transposes we can also define an inner product on $\mathbb{C}^n$.

Definition 10.2.3. The **Hermitian inner product** of $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{C}^n$ is

$$\mathbf{v}^*\mathbf{u} = \begin{bmatrix} \overline{v}_1 & \cdots & \overline{v}_n \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \overline{v}_1 u_1 + \cdots + \overline{v}_n u_n$$

- We say that $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\mathbf{v}^*\mathbf{u} = 0$, or equivalently, if $\mathbf{u}^*\mathbf{v} = 0$.
- We say that $\mathbf{u}$ and $\mathbf{v}$ are **orthonormal** if $\mathbf{u}^*\mathbf{v} = 0$ and $||\mathbf{u}|| = ||\mathbf{v}|| = 1$.

Example 10.2.4. The vectors $\mathbf{u} = \begin{bmatrix} 1 \\ i \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} i \\ 1 \end{bmatrix}$ in $\mathbb{C}^2$ are orthogonal since

$$\mathbf{u}^*\mathbf{v} = \begin{bmatrix} 1 & -i \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = 0$$

They are not orthonormal since their moduli are not 1.

Check for yourself that if you just took the transpose of a complex vector and computed $\mathbf{z}^\top\mathbf{z}$, you may not get a nonnegative real number. For example, if $z = 1+i$ then $z^\top z = (1+i)^\top(1+i) = 2i$ which is not a real number. If $\mathbf{z} = \begin{bmatrix} 1 \\ i \end{bmatrix}$ then $\mathbf{z}^\top\mathbf{z} = 0$ which cannot be the length of a nonzero vector. Therefore, it is necessary to take the conjugate transpose as opposed to just the transpose when we work with complex numbers and vectors.

10.3 $\mathbb{C}^{m\times n}$ and the Complex Spectral Theorem

We can now define a complex matrix as $A = (z_{ij}) \in \mathbb{C}^{m\times n}$ with $z_{ij} \in \mathbb{C}$ for all i, j. The **conjugate transpose** of A is obtained by transposing A and then conjugating all entries, i.e.,

$$A^* = (\overline{z}_{ji}) \in \mathbb{C}^{n\times m}$$

Example 10.3.1. Let $A = \begin{bmatrix} 1 & i \\ 0 & 1+i \end{bmatrix}$ then the conjugate transpose is $A^* = \begin{bmatrix} 1 & 0 \\ -i & 1-i \end{bmatrix}$.

We also mention that the usual properties of transpose carry over into the complex setting, namely,

$$(A\mathbf{u})^*\mathbf{v} = \mathbf{u}^*(A^*\mathbf{v}) \quad \text{and} \quad (AB)^* = B^*A^*$$

We can now define the complex analog of a symmetric matrix which appears in numerous applications.

Definition 10.3.2. A matrix $A \in \mathbb{C}^{n\times n}$ is **Hermitian** if $A = A^*$.

A consequence of this definition is that every real symmetric matrix is Hermitian.

Example 10.3.3. If $A = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix}$ then $A^* = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix} = A$, and hence A is Hermitian.

As we had with symmetric matrices, Hermitian matrices have three important properties that combine to give the spectral theorem in the complex setting.

Proposition 10.3.4. *If A is Hermitian and $\boldsymbol{z} \in \mathbb{C}^n$ then $\boldsymbol{z}^*A\boldsymbol{z} \in \mathbb{R}$.*

Proof. Recall that a complex number z is real if and only if $\overline{z} = z$. Taking the conjugate of $\mathbf{z}^*A\mathbf{z}$ and using the fact that $A = A^*$, we see that

$$(\mathbf{z}^*A\mathbf{z})^* = \mathbf{z}^*A^*(\mathbf{z}^*)^* = \mathbf{z}^*A^*\mathbf{z} = \mathbf{z}^*A\mathbf{z}.$$

Therefore, $\mathbf{z}^*A\mathbf{z} \in \mathbb{R}$. □

Proposition 10.3.5. *Every eigenvalue of a Hermitian matrix is real.*

Proof. Assume that $A^* = A$ and $A\mathbf{z} = \lambda\mathbf{z}$ with $\lambda \in \mathbb{C}$. Then from Proposition 10.3.4 we know that

$$\mathbf{z}^*A\mathbf{z} = \mathbf{z}^*\lambda\mathbf{z} = \lambda\mathbf{z}^*\mathbf{z} = \lambda||\mathbf{z}||^2 \in \mathbb{R}$$

Since $||\mathbf{z}||^2 \in \mathbb{R}$, we must have $\lambda \in \mathbb{R}$ as well. □

Example 10.3.6. Continuing the example from above with $A = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix}$ we have that

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} 2-\lambda & 3-3i \\ 3+3i & 5-\lambda \end{bmatrix}\right) = \lambda^2 - 7\lambda - 8 = (\lambda - 8)(\lambda + 1)$$

and so the eigenvalues of A are 8 and -1 which are real.

Proposition 10.3.7. *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian and assume that $A\boldsymbol{z} = \lambda\boldsymbol{z}, A\boldsymbol{y} = \beta\boldsymbol{y}$ with $\lambda \neq \beta$ and $\boldsymbol{z}, \boldsymbol{y} \in \mathbb{C}^n$. Then $\boldsymbol{y}^*\boldsymbol{z} = 0$. That is, eigenvectors of a Hermitian matrix corresponding to different eigenvalues are always orthogonal.*

Proof. First, observe that

$$\mathbf{y}^* A\mathbf{z} = \mathbf{y}^* \lambda \mathbf{z} = \lambda \mathbf{y}^* \mathbf{z}$$

Furthermore, since $A\mathbf{y} = \beta \mathbf{y}$ and $\beta \in \mathbb{R}$, we have that

$$(A\mathbf{y})^* = (\beta \mathbf{y})^* = \beta \mathbf{y}^* \implies \mathbf{y}^* A^* = \beta \mathbf{y}^*$$

Multiplying both sides of this equation by $\mathbf{z}$ on the right (and using the fact that A is Hermitian) we can conclude that

$$\mathbf{y}^* A^* \mathbf{z} = \mathbf{y}^* A\mathbf{z} = \beta \mathbf{y}^* \mathbf{z}$$

Now we have two expressions for $\mathbf{y}^* A\mathbf{z}$, and equating them we get

$$\lambda \mathbf{y}^* \mathbf{z} = \beta \mathbf{y}^* \mathbf{z} \implies (\lambda - \beta)\mathbf{y}^* \mathbf{z} = \mathbf{0} \implies \mathbf{y}^* \mathbf{z} = \mathbf{0}$$

since $\lambda \neq \beta$. □

Example 10.3.8. Carrying on with the same matrix from the previous two examples we have

$$A = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix}$$

with eigenvalues $\lambda = 8$ and $\lambda = -1$. The corresponding eigenvectors are

$$\mathbf{z} = \begin{bmatrix} 1 \\ 1+i \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 1-i \\ -1 \end{bmatrix}$$

respectively. Computing their inner product we can see that

$$\mathbf{y}^* \mathbf{z} = \begin{bmatrix} 1+i & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1+i \end{bmatrix} = (1+i) - (1+i) = 0$$

Furthermore, we can divide these vectors by their norms to obtain orthonormal vectors

$$\frac{\mathbf{z}}{||\mathbf{z}||} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1+i \end{bmatrix} \quad \text{and} \quad \frac{\mathbf{y}}{||\mathbf{y}||} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1-i \\ 1 \end{bmatrix}$$

Since these vectors live in $\mathbb{C}^2$ we can conclude that they are actually an orthonormal basis of $\mathbb{C}^2$ (with respect to the Hermitian inner product). This phenomenon always happens with Hermitian matrices. We have an orthonormal basis of eigenvectors, therefore, we can diagonalize A by writing

$$A = Q \begin{bmatrix} 8 & 0 \\ 0 & -1 \end{bmatrix} Q^*$$

where

$$Q = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1-i \\ 1+i & -1 \end{bmatrix}$$

This matrix is special in that its conjugate transpose equals its inverse. This is a consequence of the fact that its columns form an orthonormal basis.

Definition 10.3.9. A complex square matrix Q with the property that $Q^*Q = QQ^* = I$ is called a **unitary** matrix. It is the complex analog of an orthonormal real matrix.

We can now combine all these ideas to arrive at the complex spectral theorem.

Theorem 10.3.10. *If $A \in \mathbb{C}^{n\times n}$ is Hermitian then*

- *All eigenvalues of A are real.*
- *$\mathbb{C}^n$ has an orthonormal basis of eigenvectors of A.*
- *If Q is the eigenvector matrix for A, then Q is unitary.*
- *The matrix A is* ***unitarily diagonalizable****, i.e. there exists a diagonal matrix $\Lambda \in \mathbb{R}^{n\times n}$ and a unitary matrix Q such that*
$$A = Q\Lambda Q^*$$

Example 10.3.11. Writing out the unitary diagonalization for the matrix $A = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix}$ we get

$$\begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1-i \\ 1+i & -1 \end{bmatrix}\begin{bmatrix} 8 & 0 \\ 0 & -1 \end{bmatrix}\frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1-i \\ 1+i & -1 \end{bmatrix}$$

Unitary matrices have special properties.

Theorem 10.3.12. *Let $Q \in \mathbb{C}^{n\times n}$ be unitary. If $\lambda \in \mathbb{C}$ is an eigenvalue of Q then $|\lambda| = 1$. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are eigenvectors of Q corresponding to different eigenvalues then $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal.*

Proof. Suppose $\mathbf{x} \in \mathbb{C}^n$. Then $\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x}\rangle = \mathbf{x}^*\mathbf{x} = \mathbf{x}^*Q^*Q\mathbf{x} = (Q\mathbf{x})^*(Q\mathbf{x}) = \langle Q\mathbf{x}, Q\mathbf{x}\rangle = \|Q\mathbf{x}\|^2$. Recall that orthonormal real matrices also had this property. If now $Q\mathbf{u} = \lambda\mathbf{u}$ for a non-zero $\mathbf{u}$, then $\|Q\mathbf{u}\| = |\lambda|\|\mathbf{u}\|$. However, since $\|Q\mathbf{u}\| = \|\mathbf{u}\|$ it must be that $|\lambda| = 1$. Recall that if $|\lambda| = 1$ then $\overline{\lambda} = \frac{1}{\lambda}$.

Now suppose $Q\mathbf{u} = \lambda\mathbf{u}$ and $Q\mathbf{v} = \mu\mathbf{v}$ where $\lambda \neq \mu$. Then

$$\langle \mathbf{u}, \mathbf{v}\rangle = \langle Q\mathbf{u}, Q\mathbf{v}\rangle = \langle \lambda\mathbf{u}, \mu\mathbf{v}\rangle = \overline{\mu}\lambda\langle \mathbf{u}, \mathbf{v}\rangle = \frac{\lambda}{\mu}\langle \mathbf{u}, \mathbf{v}\rangle.$$

Since $\lambda \neq \mu$, $\frac{\lambda}{\mu} \neq 1$. Therefore it must be that $\langle \mathbf{u}, \mathbf{v}\rangle = 0$. □

Any matrix of the form B^*B where $B \in \mathbb{C}^{k\times n}$ is Hermitian since $(B^*B)^* = B^*B$.

Definition 10.3.13. A Hermitian matrix $A \in \mathbb{C}^{n\times n}$ is **positive semidefinite** if $A = B^*B$ for some $B \in \mathbb{C}^{k\times n}$.

Example 10.3.14. Let $B = \begin{bmatrix} 1 & 2+i \end{bmatrix}$. Then

$$B^*B = \begin{bmatrix} 1 \\ 2-i \end{bmatrix}\begin{bmatrix} 1 & 2+i \end{bmatrix} = \begin{bmatrix} 1 & 2+i \\ 2-i & 5 \end{bmatrix}$$

is positive semidefinite.

As for real PSD matrices, all eigenvalues of a Hermitian PSD matrix are nonnegative real numbers.

Proposition 10.3.15. *If $A \in \mathbb{C}^{n\times n}$ is Hermitian and PSD then all its eigenvalues are nonnegative real numbers.*

Proof. Suppose A is Hermitian PSD. Then there exists $B \in \mathbb{C}^{k\times n}$ such that $A = B^*B$. Suppose $A\mathbf{z} = \lambda\mathbf{z}$ where $\mathbf{z} \neq \mathbf{0}$. Since A is Hermitian we know that $\lambda \in \mathbb{R}$ and since $A = B^*B$ we have that $(B^*B)\mathbf{z} = \lambda\mathbf{z}$. Multiplying both sides of this last equation by $\mathbf{z}^*$ we get

$$\mathbf{z}^*(B^*B)\mathbf{z} = \lambda\mathbf{z}^*\mathbf{z} \;\Rightarrow\; (B\mathbf{z})^*(B\mathbf{z}) = \lambda\|\mathbf{z}\|^2 \;\Rightarrow\; \|B\mathbf{z}\|^2 = \lambda\|\mathbf{z}\|^2.$$

This implies that $\lambda = \frac{\|B\mathbf{z}\|^2}{\|z\|^2} \geq 0$. □

There is **much** more to discover about the world of complex matrices. Many of the nice theorems and properties we have seen for real matrices have their complex analogs, and often times the statements in the complex setting simply involve interchanging the word *transpose*, with *conjugate transpose*, *symmetric* with *Hermitian*, and *orthogonal* with *unitary*.

10.4 Application: Fourier Analysis

Let V be the set of all functions $f : \mathbb{R} \to \mathbb{C}$ that are periodic with period 2π. This means that $f(x) = f(x+2\pi)$ for all $x \in \mathbb{R}$. Recall that we can add and scalar multiply functions by the rule

$$(f+g)(x) = f(x) + g(x), \quad (\gamma f)(x) = \gamma f(x) \ \forall \ \gamma \in \mathbb{C}$$

Under these rules, V is a vector space over $\mathbb{C}$. We can also assume that functions in V are sufficiently nice in the sense of being bounded and continuous without changing that V is a complex vector space.

The functions $e^{inx} = \cos(nx) + i\sin(nx)$ for $n = 0, \pm1, \pm2, \pm3, \ldots$ lie in V. Recall that we can interpret e^{inx} as a unit vector in the complex plane, which as a function of x, rotates around the unit circle making n revolutions as x varies between 0 to 2π. So for different integer values of n we have infinitely many unit vectors rotating around the unit circle in the complex plane at different speeds. If n is a negative integer, the rotation is clockwise, while if n is a positive integer the rotation is counter clockwise.

The vector space V has an inner product defined as

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

The norm of a function f under this inner product is

$$\|f\|^2 = \int_{-\pi}^{\pi} f(x)\overline{f(x)}dx = \int_{-\pi}^{\pi} |f(x)|^2 dx$$

In particular, $\|e^{inx}\|^2 = \int_{-\pi}^{\pi} e^{inx}e^{-inx}dx = \int_{-\pi}^{\pi} dx = 2\pi$, and so $\|e^{inx}\| = \sqrt{2\pi}$. Check that

$$\langle e^{inx}, e^{imx} \rangle = \int_{-\pi}^{\pi} e^{inx}e^{-imx}dx = \int_{-\pi}^{\pi} e^{i(n-m)x}dx = \begin{cases} 0 & \text{if } n \neq m \\ 2\pi & \text{if } n = m \end{cases}$$

From these calculations we conclude that the infinite set of functions

$$\phi_n = \frac{1}{\sqrt{2\pi}}e^{inx}, \quad n \in \mathbb{Z}$$

are orthonormal. It turns out that the vector ϕ_n also span the vector space V. Therefore, any $f \in V$ can be written uniquely as

$$f = \sum_{n\in\mathbb{Z}} \hat{f}_n e^{inx} = \sum_{n\in\mathbb{Z}} \sqrt{2\pi}\hat{f}_n \phi_n$$

which is called the **Fourier series** of f. The coefficients $\{\hat{f}_n\}$ are the **Fourier coefficients** of f and they tell you how much of each function ϕ_n is in f. We conclude that V is an infinite-dimensional vector space with orthonormal basis $\{\phi_n : n \in \mathbb{Z}\}$.

Since $\{\phi_n\}$ is an orthonormal basis of V, we can compute the coefficients $\hat{f}_n$ easily:

$$\sqrt{2\pi}\hat{f}_n = \langle f, \phi_n \rangle = \int_{-\pi}^{\pi} f(x)\frac{1}{\sqrt{2\pi}}e^{-inx}dx$$

which implies that

$$\hat{f}_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx$$

If f is real valued, meaning $f : \mathbb{R} \to \mathbb{R}$ then the Fourier series of f is an infinite sum of sines and cosines of increasing frequency. Fourier analysis is the basis of signal processing allowing us to understand the components of a periodic signal which then gives us the ability to modify the signal or compress it by removing all components with really small Fourier coefficients.

Here are some examples of Fourier series. Assume that all functions below are periodic with period 2π:

1. $f(x) = x \;\;\Rightarrow\;\; f(x) = 2\sum_1^\infty \frac{(-1)^{n+1}}{n} \sin nx$
2. $f(x) = |\sin(x)| \;\;\Rightarrow\;\; f(x) = \frac{2}{\pi} - \frac{4}{\pi}\sum_1^\infty \frac{\cos 2nx}{4n^2-1}$

The 3blue1brown video on Fourier series gives a nice animation of how one can see any function in V as a sum of scaled rotating unit vectors in the complex plane. It also has nice illustrations of how sine/cosine waves make up a periodic function as above.

10.4.1 The Discrete Fourier Transform (DFT)

Sometimes we do not know the function f but instead have a set of measurements which are values of f at discrete points $x_0, \ldots, x_{n-1} \in \mathbb{R}$. This is the case if we observe f through an experiment. Let $f = (f_0, f_1, \ldots, f_{n-1})^\top \in \mathbb{C}^n$ denote the vector of observed data where $f_i = f(x_i)$. It is still the case the f is a sum of weighted frequency functions as above where the weights denoted by $\hat{f}_i$ are the Fourier coefficients. The formula for $\hat{f}_k$ in this discrete setting is

$$\hat{f}_k = \sum_{j=0}^{n-1} f_j e^{-i\frac{2\pi}{n}kj} = \sum_{j=0}^{n-1} f_j (e^{-i\frac{2\pi}{n}})^{kj}$$

Let $\omega_n := e^{-i\frac{2\pi}{n}} = \cos(\frac{2\pi}{n}) - i\sin(\frac{2\pi}{n})$. Then note that $(\omega_n)^n = e^{-i2\pi} = \cos(2\pi) - i\sin(2\pi) = 1$ which means that ω_n is a nth root of 1, meaning that it is a solution to the equation $x^n = 1$. We can also think of ω_n as a unit vector in the complex plane and the powers $\omega_n, \omega_n^2, \omega_n^3, \ldots, \omega_n^n = 1$ are n unit vectors spaced evenly around the unit circle in the complex plane with the angle between any two consecutive vectors being $\frac{2\pi}{n}$. In the rest of this section it would be useful to have ω_8 in mind. The powers $\omega_8, \omega_8^2, \ldots, \omega_8^7, 1$ are 8 unit vectors in the complex plane each 45^0 from the next.

With the notation for ω_n, notice that we can rewrite the expression for $\hat{f}_k$ as a dot product of vectors. More generally, setting $\hat{f} = (\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_{n-1})^\top$ we have that

$$\begin{pmatrix} \hat{f}_0 \\ \hat{f}_1 \\ \hat{f}_2 \\ \vdots \\ \hat{f}_{n-1} \end{pmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega_n & \omega_n^2 & \omega_n^3 & \cdots & \omega_n^{n-1} \\ 1 & \omega_n^2 & \omega_n^4 & \omega_n^6 & \cdots & \omega_n^{2(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \omega_n^{n-1} & \omega_n^{2(n-1)} & \omega_n^{3(n-1)} & \cdots & \omega_n^{(n-1)^2} \end{bmatrix}}_{F_n} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{n-1} \end{pmatrix}$$

The matrix F_n is called the **discrete Fourier transform**. Note that it is a Vandermonde matrix and is hence invertible. It turns out that $\frac{1}{\sqrt{n}}F_n$ is unitary which makes it easy to compute F_n^{-1}. This allows us to compute f from $\hat{f}$ by inverting F_n in the formula $\hat{f} = F_n f$. The matrix F_n^{-1} is the **inverse discrete Fourier transform**.

The discrete Fourier transform is behind a huge part of our everyday life such as audio, video and image processing. It is the basis of signal processing in general. It is how jpeg works for example. See the videos by Steve Brunton for several applications of the DFT.

10.4.2 The Fast Fourier Transform (FFT)

Since the DFT is so very essential to our everyday life, it is crucial that it can be computed efficiently. The Fast Fourier Transform is an algorithm that computes the DFT efficiently making applications actually possible. Check that the multiplication $F_n f$ involves $O(n^2)$ calculations. The FFT does this multiplication in $O(n \log n)$ steps which makes all the difference in being to implement the DFT. When n is large, $n \log n$ is more or less like n. The FFT was named one of the top 10 algorithms of the 20th century by Science magazine, some will argue it is **the** top algorithm of the last century.

The idea behind the FFT is to decompose F_n as a product of sparse matrices so that the matrix vector multiplication $F_n f$ can be done very fast, in $O(n \log n)$ time. Let's see how this works on an example.

Example 10.4.1. Let $n = 8$. Setting $w := \omega_8$ we get

$$F_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & w & w^2 & w^3 & w^4 & w^5 & w^6 & w^7 \\ 1 & w^2 & w^4 & w^6 & w^8 & w^{10} & w^{12} & w^{14} \\ 1 & w^3 & w^6 & w^9 & w^{12} & w^{15} & w^{18} & w^{21} \\ 1 & w^4 & w^8 & w^{12} & w^{16} & w^{20} & w^{24} & w^{28} \\ 1 & w^5 & w^{10} & w^{15} & w^{20} & w^{25} & w^{30} & w^{35} \\ 1 & w^6 & w^{12} & w^{18} & w^{24} & w^{30} & w^{36} & w^{42} \\ 1 & w^7 & w^{14} & w^{21} & w^{28} & w^{35} & w^{42} & w^{49} \end{bmatrix}$$

The first step is to permute the columns of F_8 so that columns $1, 3, 5, 7$ are first followed by columns $2, 4, 6, 8$. This can be achieved by multiplying F_8 on the right by the following permutation matrix:

$$P_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Computing $F_8 P_8$ and rewriting selectively using $w^8 = 1, w^4 = -1, w^6 = -w^2, w^7 = -w^3$ etc, we get

$$F_8 P_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & w^2 & w^4 & w^6 & w & w^3 & w^5 & w^7 \\ 1 & w^4 & w^8 & w^{12} & w^2 & w^6 & w^{10} & w^{14} \\ 1 & w^6 & w^{12} & w^{18} & w^3 & w^9 & w^{15} & w^{21} \\ 1 & w^8 & w^{16} & w^{24} & w^4 & w^{12} & w^{20} & w^{28} \\ 1 & w^{10} & w^{20} & w^{30} & w^5 & w^{15} & w^{25} & w^{35} \\ 1 & w^{12} & w^{24} & w^{36} & w^6 & w^{18} & w^{30} & w^{42} \\ 1 & w^{14} & w^{28} & w^{42} & w^7 & w^{21} & w^{35} & w^{49} \end{bmatrix} = \left[\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & w^2 & w^4 & w^6 & w & w^3 & w^5 & w^7 \\ 1 & w^4 & w^8 & w^{12} & w^2 & w^6 & w^{10} & w^{14} \\ 1 & w^6 & w^{12} & w^{18} & w^3 & w^9 & w^{15} & w^{21} \\ \hline 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & w^2 & w^4 & w^6 & -w & -w^3 & -w^5 & -w^7 \\ 1 & w^4 & w^8 & w^{12} & -w^2 & -w^6 & -w^{10} & -w^{14} \\ 1 & w^6 & w^{12} & w^{18} & -w^3 & -w^9 & -w^{15} & -w^{21} \end{array}\right]$$

Setting

$$D_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & w & 0 & 0 \\ 0 & 0 & w^2 & 0 \\ 0 & 0 & 0 & w^3 \end{bmatrix}$$

we see that

$$F_8 P_8 = \begin{bmatrix} I_4 & D_4 \\ I_4 & -D_4 \end{bmatrix} \begin{bmatrix} F_4 & 0 \\ 0 & F_4 \end{bmatrix}$$

Recursively, we can rewrite F_4 and F_2 in the same way as

$$F_4 P_4 = \begin{bmatrix} I_2 & D_2 \\ I_2 & -D_2 \end{bmatrix} \begin{bmatrix} F_2 & 0 \\ 0 & F_2 \end{bmatrix} \quad \text{and} \quad F_2 P_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Assembling all the pieces back together we get a simple expression for F_8 as a product of sparse matrices. This is how the FFT works.