Linear Algebra Refresher

Review, Amplification, Examples

Linear Algebra Refresher

Review, Amplification, Examples

Thomas R. Shemanske
Dartmouth College

Version: October 11, 2023

Preface

Linear algebra is an elegant subject and remarkable tool whose influence reaches well beyond uses in pure and applied mathematics. Certainly all math majors as well as majors from a growning number of STEM fields have taken a first course in linear algebra. On the other hand, as with any tool or collection of knowledge, without frequent use, one's facility with the material wanes.

These notes are not intended as a first or second course in linear algebra, though they assume the reader has seen the material in a basic linear algebra course, covered for example in [1], [2], or [3].

These notes will undertake a review of many basic topics from a typical first course, often taking the opportunity to interleave more advanced concepts with simpler ones when convenient. It will refresh the reader's memory of definitions, structural results, core examples, and provide some computational tools to help the reader come to a deeper appreciation of the ideas first met perhaps a long time ago.

An additional resource of which to take advantage is Robert Beezer's A First Course in Linear Algebra[1] which explores the use of computation in far more depth than is done here. Computations in these notes uses Sage (sagemath.org[2]) which is a free, open source, software system for advanced mathematics. Sage can be used either on your own computer, a local server, or on SageMathCloud (https://cocalc.com[3]).

Thomas R. Shemanske
Hanover, NH
2020

[1] `linear.ups.edu`
[2] `sagemath.org`
[3] `cocalc.com`

Contents

Preface iv

1 A quick review of a first course 1

1.1 Vector spaces and linear maps . . . 1
1.2 Measuring injectivity and surjectivity . . . 5
1.3 Rank and Nullity . . . 7
1.4 Coordinates and Matrices . . . 11
1.5 Eigenvalues, eigenvectors, diagonalization . . . 16
1.6 Minimal and characteristic polynomials . . . 22
1.7 Some Sage examples . . . 24
1.8 Exercises (with solutions) . . . 27

2 Vector space constructions 42

2.1 Sums and Direct Sums . . . 42
2.2 Quotient Spaces . . . 44
2.3 Linear Maps out of quotients . . . 46

3 Inner Product Spaces 48

3.1 Inner Product Spaces . . . 48
3.2 Orthogonality and applications . . . 51
3.3 Orthogonal Projections and Least Squares Approximations . . . 59
3.4 Diagonalization of matrices in Inner Product Spaces . . . 73
3.5 Adjoint Maps and properties . . . 79
3.6 Singular Value Decomposition . . . 83
3.7 Exercises (with solutions) . . . 103

4 Definitions and Examples 112

4.1 Basic Definitions and Examples . . . 112

Back Matter

References and Suggested Readings **118**

Chapter 1

A quick review of a first course

As stated in the preface, these notes presume the reader has seen the material in a basic linear algebra course such as in [1], [2], or [3] at some point in their past. This first chapter is a rapid summary of the theory up to inner product spaces. It includes vector spaces and linear maps, span and linear independence, coordinates and the matrix of a linear map, the notions of rank and nulllity, and the basic results surrounding diagonalization. It ends with some examples showing how to leverage Sage to answer questions surrounding these basic ideas.

1.1 Vector spaces and linear maps

In simplest terms, linear algebra is the study of vector spaces and linear maps between them. But what does that really mean? One overriding goal in Mathematics is to classify objects into distinct "types", and also to characterize the manner in which complicated structures are constructed from simpler ones. For example, in linear algebra the notion of when two vector spaces are the same "type" (i.e., are indistinguishable as vector spaces) is captured by the notion of isomorphism. In terms of structure, the notions of bases and direct sums play a crucial role.

1.1.1 Some familiar examples of vector spaces

While most of the examples and applications we shall consider are vector spaces over the field of real or complex numbers, for the examples below, we let F denote any field. First recall the definition of a vector space [click the link to toggle the definition].

- For an integer $n \geq 1$, the set $V = F^n$, of n-tuples of numbers in F viewed as column vectors with n entries, is a vector space over F.

- For integers $m, n \geq 1$, the vector space of $m \times n$ matrices with entries from F is denoted $M_{m\times n}(F)$. Column vectors in F^m are the matrices in $M_{m\times 1}(F)$, while row vectors in F^n are matrices in $M_{1\times n}(F)$.

- For an integer $n \geq 1$, we denote by $P_n(F)$ the vector space of polynomials of degree at most n having coefficients in F.
- The vector space of all polynomials with coefficients in F is often denoted as $P(F)$ in many linear algebra texts, though in more advanced courses (say abstract algebra) the more typical notation is $F[x]$ (with x the indeterminant), a notation we shall use here.

1.1.2 Linear independent and spanning sets

Let V be a vector space over a field F. For a subset $S \subseteq V$, we have the fundamental notions of linear independence, linear dependence, span, basis, and dimension. We remind the reader that even when dealing with infinite dimensional vector spaces, linear combinations involve only a finite number of summands.

Checkpoint 1.1.1 Let W be a subspace of a vector space V, and S a subset of W. Show that $\mathrm{Span}(S) \subseteq W$.

Hint. Since W is itself a vector space, it is closed under vector addition and scalar multiplication.

There are many important theorems which relate the above notions and which can be found in all standard books. We summarize some of these here.

For the remainder of this section we restrict to a vector space V of finite dimension n.

Theorem 1.1.2 Constructing bases. *Let V have finite dimension n, and let $S \subset V$.*

- *If S is linearly independent, then $\#S \leq n$, and if $\#S < n$, then S can be extended to a basis for V, that is there is a finite subset T of V, so that $S \cup T$ is a basis for V.*
- *If $\#S > n$, then S is linearly dependent, and there is a subset $S_0 \subset S$ which is linearly independent and for which $\mathrm{Span}(S_0) = \mathrm{Span}(S)$.*
- *In more colloquial terms, any linearly independent subset of V can be extended to a basis for V, and any spanning set can be reduced to produce a basis.*

As a consequence of the above, we have another important theorem.

Theorem 1.1.3 *Let V be a vector space with finite dimension n. Then*

- *Any set of n linearly independent vectors in V is a basis for V.*
- *Any set of n vectors in V which span V is a basis for V.*

Proof. The proofs are straightforward from the above since if a set of n linearly independent vectors in V did not span, you could add a vector to the set of n and obtain an independent set with $n+1$ elements. Similarly, if n elements spanned V but were not independent, you could eliminate one giving a basis with too few elements. ■

Fundamental to the proofs of these theorems is the following:

Theorem 1.1.4 *If $S \subset V$ is a linearly independent set, and $v \in V \setminus \mathrm{Span}(S)$, then $S \cup \{v\}$ is linearly independent.*

Exercises

1. Let A be an $m \times n$ matrix. Its **row space** is the span of the rows of A and so is a subspace of F^n. Its **column space** is the span of its columns and so is a subspace of F^m.

Can any given column of a matrix always be used as part of a basis for the column space?

Hint. Under what conditions is a set with one vector a linearly independent subset of the vector space?

Answer. Any column of a matrix which is not the column of all zeros can be used as part of the basis of the column space since the single nonzero column is a linearly independent set.

2. Suppose the first two columns of a matrix are nonzero. What is an easy way to check that both columns can be part of a basis for the column space?

Hint. What does the notion of linear dependence reduce to in the case of two vectors?

Answer. Two columns which are not multiples of one another may be used as part of the basis for the column space.

3. Do you think there is an easy way to determine if the first three nonzero columns of a matrix can be part of a basis for the column space?

Hint. Easy may be in the eye of the beholder.

Answer. Not typically by inspection. Given the first two columns are linearly independent, one needs to know the third is not a linear combination of the first two. In Section 1.3 we provide answers using either elementary column operations, or perhaps surprisingly elementary row operations.

1.1.3 Defining a linear map.

Starting from the definition of a linear map, one proves by induction that a linear map takes linear combinations of vectors in the domain to the same linear combination of the corresponding vectors in the codomain. More precisely we have

Proposition 1.1.5 Linear maps preserve structure. *Let V, W be vector spaces over a field F, and $T : V \to W$ a linear map. Then for every finite collection of vectors $v_1, \ldots, v_r \in V$, and scalars $a_1, \ldots a_r \in F$ we have*

$$T(a_1v_1 + \cdots + a_rv_r) = a_1T(v_1) + \cdots + a_rT(v_r). \tag{1.1.1}$$

If the goal is to define a linear map $T : V \to W$, one must define $T(v)$ for all vectors $v \in V$, so it is ideal to know how to represent a given vector as a linear combination of others. In particular this leads to the notion of a basis for a vector space. Recall some standard bases for familiar vector spaces.

So now let's suppose V is a finite-dimensional vector space over a field F with basis $\mathcal{B} = \{v_1, \ldots, v_n\}$, and W is a completely arbitrary vector space over F. To define a linear map $T : V \to W$ it is certainly necessary to define the values $T(v_1), \ldots, T(v_n)$. The important point is that this is all that needs to be done!

Theorem 1.1.6 Uniquely defined linear maps. *Let V be a finite-dimensional vector space over a field F with basis $\mathcal{B} = \{v_1, \ldots, v_n\}$. Let W be any vector space over F, and let $w_1, \ldots, w_n$ be arbitrarily chosen vectors in W. Then there is a unique linear map $T : V \to W$ which satisfies $T(v_i) = w_i$, for $i = 1, \ldots n$.*

Indeed Proposition 1.1.5 tells us that if such a linear map T exists, then

$$T(a_1v_1 + \cdots + a_nv_n) = a_1w_1 + \cdots + a_nw_n. \tag{1.1.2}$$

Since $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a basis for V, every element of V has a unique expression of the form $a_1v_1 + \cdots + a_nv_n$, so the map T is defined for every vector in V, and it is easy to determine from its definition that T is indeed a linear map.

Next recall the definition of the span of a set of vectors, and gain some facility by doing the following exercises.

Exercises

1. Let $T : V \to W$ be a linear map between vector spaces, and $\{v_1, \ldots, v_r\} \subseteq V$. Show that

$$T(\mathrm{Span}(\{v_1, \ldots, v_r\})) = \mathrm{Span}\{T(v_1), \ldots, T(v_r)\}.$$

Hint 1. When you want to show that two sets, say X and Y are equal, you must show $X \subseteq Y$ and $Y \subseteq X$. And to show that (for example) $X \subseteq Y$, you need only show that for each choice of $x \in X$, that $x \in Y$.

Hint 2. So if $w \in T(\mathrm{Span}(\{v_1, \ldots, v_r\}))$, then $w = T(a_1v_1 + \cdots + a_rv_r)$ for some choice of scalars $a_1, \ldots, a_r$.

2. Let $V = P_2(\mathbb{R})$ be the vector space of all polynomials of degree at most two with real coefficients. We know that both sets $\{1, x, x^2\}$ and $\{2, 3x, 2 + 3x + 4x^2\}$ are bases for V.

By Theorem 1.1.6, there are uniquely determined linear maps $S, T : V \to V$ defined by

$$\begin{aligned} &T(1) = 0, \quad T(x) = 1, \quad T(x^2) = 2x.\\ &S(2) = 0, \quad S(3x) = 3, \quad S(2 + 3x + 4x^2) = 3 + 8x. \end{aligned}$$

Show that the maps S and T are the same.

Hint 1. Why is it enough to show that $S(1) = 0$, $S(x) = 1$, and $S(x^2) = 2x$?

Hint 2. How does the linearity of S play a role?

1.2 Measuring injectivity and surjectivity

1.2.1 Injective and surjective linear maps: assessment and implications.

Given a linear map $T : V \to W$ (between vector spaces V, W), we know the function-theoretic definitions of injective and surjective. Let's first give an alternate characterization of these primitives, and then explore how linearity informs and refines our knowledge.

Given a function $f : X \to Y$ between sets X and Y, and an element $y \in Y$, the **inverse image of** y is the set of elements of X which map onto y via f, that is

$$f^{-1}(y) = \{x \in X \mid f(x) = y\}.$$

Thus an equivalent way in which to say that a function f is **surjective** is if for every $y \in Y$, the inverse image, $f^{-1}(y)$ is non-empty, and an equivalent way to say that a function is **injective** is to say for every $y \in Y$, the inverse image, $f^{-1}(y)$ is either empty or consists of a single element.

For a linear map $T : V \to W$, the inverse image of the $\mathbf{0}_W$ in W plays a special role and is given name recognition:

Definition 1.2.1 The **kernel** or **nullspace** of T is defined as

$$\ker(T) = \mathrm{Null}(T) = T^{-1}(\mathbf{0}_W) = \{v \in V \mid T(v) = \mathbf{0}_W\}.$$

$\diamondsuit$

One recalls that since $T(\mathbf{0}_V) = \mathbf{0}_W$, we always have $\mathbf{0}_V \in \ker(T)$, and indeed the kernel (null space) is a subspace of V.

Now using that for a linear map T, $T(v) = T(v')$ if and only if $T(v-v') = \mathbf{0}_W$, one easily deduces the familiar proposition below. Since it should be clear from context, we shall henceforth simply write $\mathbf{0}$, leaving to the reader to understand the space to which we are referring.

Proposition 1.2.2 *A linear map $T : V \to W$ is injective if and only if* $\ker(T) = \{\mathbf{0}\}$.

The significance of this proposition is that rather than checking that $T^{-1}(w)$ consists of at most one element for every $w \in W$ (as for a generic function), for linear maps it is enough to check for the single element $w = \mathbf{0}$. The kernel also says something about the image of a linear map. Suppose $T(v_0) = w$. Then $T(v) = w$ if and only if $v = v_0 + k$, where $k \in \ker(T)$. Said another way

$$T^{-1}(w) = \{v_0 + k \mid k \in \ker(T)\} = v_0 + \ker(T). \tag{1.2.1}$$

Now that we have reminded ourselves of the definitions and basic properties, we explore how bases dovetail with the notion of injective and surjective linear maps.

Proposition 1.2.3 Linear maps and bases. *Let $T : V \to W$ be a linear map between vector spaces and suppose that V is finite-dimensional with basis $\mathcal{B} = \{v_1, \ldots, v_n\}$. Then*

1. *T is injective if and only if $\{T(v_1), \ldots, T(v_n)\}$ is a linearly independent subset of W.*

2. *T is surjective if and only if $\{T(v_1), \ldots, T(v_n)\}$ is a spanning set for W.*

Proof of (1). First suppose that T is injective and to proceed by contradiction that $\{T(v_1), \ldots, T(v_n)\}$ is linearly dependent. Then there exist scalars $a_1, \ldots, a_n$ not all zero, so that

$$a_1T(v_1) + \cdots + a_nT(v_n) = \mathbf{0}.$$

By (1.1.1)

$$T(a_1v_1 + \cdots + a_nv_n) = a_1T(v_1) + \cdots + a_rT(v_r) = \mathbf{0},$$

which means that $(a_1v_1+\cdots+a_nv_n) \in \ker(T)$. Since $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a linearly independent set and the a_i's are not all zero, we conclude $\ker(T) \neq \{\mathbf{0}\}$ which contradicts that T is injective.
Conversely suppose that $\{T(v_1), \ldots, T(v_n)\}$ is a linearly independent subset of W, but that T is not injective. Then $\ker(T) \neq \{\mathbf{0}\}$, and since $\{v_1, \ldots, v_n\}$ is a basis for V, there exist scalars $a_1, \ldots, a_n$ not all zero so that $a_1v_1 + \cdots + a_nv_n \in \ker(T)$. But this in turn says that

$$\mathbf{0} = T(a_1v_1 + \cdots + a_nv_n) = a_1T(v_1) + \cdots + a_nT(v_n),$$

(again by Proposition 1.1.5) showing that $\{T(v_1), \ldots, T(v_n)\}$ is linearly dependent, a contradiction. ∎

Proof of (2). First suppose that $\{T(v_1), \ldots, T(v_n)\}$ is a spanning set for W. Since $T(V)$, the image of T, is a subspace of W, and $\{T(v_1), \ldots, T(v_n)\} \subset T(V)$

$$W = \operatorname{Span}\{T(v_1), \ldots, T(v_n)\} \subseteq T(V),$$

so T is surjective.
Conversely if T is surjective, then $T(V) = W$. But with a very slight generalization of Proposition 1.1.5, we see that

$$W = T(V) = T(\operatorname{Span}\{v_1, \ldots, v_n\}) = \operatorname{Span}\{T(v_1), \ldots, T(v_n)\},$$

showing that $(v_1), \ldots, T(v_n)\}$ is a spanning set for W. ∎

Example 1.2.4 Some easy-to-check isomorphisms.

- For an integer $n \geq 1$, the vector spaces $V = F^{n+1}$ and $W = P_n(F)$ are isomorphic. One bijective linear map which demonstrates this is $T : V \to W$ given by $T(a_0, \ldots, a_n) = a_0 + a_1 x + \cdots + a_n x^n$ where we have written the element $(a_0, \ldots, a_n) \in F^{n+1}$ as a row vector for typographical simplicity.

- A more explicit example is that F^6 is isomorphic to $M_{2\times 3}(F)$ via $T(a_1, \ldots, a_6) = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \end{bmatrix}$.

□

1.2.2 Notions connected to isomorphism

There are many important concepts related to isomorphism. Taking a top-down approach, one of the most important theorems in the classification of vector spaces applies to finite-dimensional vector spaces. The classification theorem is

Theorem 1.2.5 Classification theorem for finite-dimensional vector spaces. *Two finite-dimensional vector spaces V and W defined over the same field F are isomorphic if and only if* $\dim V = \dim W$.

The proof of this theorem (often stated succinctly as "map a basis to a basis") captures a great deal about the dynamics of linear algebra including how to define a map known to be linear and how to determine whether it is injective or surjective. Try to write the proof on your own.

Proof. First let's suppose that $\dim V = \dim W$. That means that any bases for the two spaces have the same cardinality. Let $\{v_1, \ldots, v_n\}$ be a basis for V, and $\{w_1, \ldots, w_n\}$ be a basis for W. By Theorem 1.1.6, there is a unique linear map which takes $T(v_i) = w_i$, for $i = 1, \ldots, n$. By Proposition 1.2.3, it follows that T is both injective and surjective, hence an isomorphism.
Conversely, suppose that $T : V \to W$ is an isomorphism and the $\{v_1, \ldots, v_n\}$ is a basis for V. Once again by Proposition 1.2.3, it follows that $\{T(v_1), \ldots, T(v_n)\}$ is a basis for W, and since the cardinality of any basis determines the dimension of the space, we have $\dim V = \dim W$. ■

1.3 Rank and Nullity

1.3.1 Some fundamental subspaces

Let $T : V \to W$ be a linear map between vector spaces over a field F. We have defined the kernel of T, $\ker(T) = \mathrm{Null}(T)$, (also called the **nullspace**) and noted that it is a subspace of the domain V. The image of T, $\mathrm{Im}(T)$, is a subspace of the codomain W.

1.3.2 The rank-nullity theorem

Given a linear map $T : V \to W$, with V finite dimensional, there is a fundamental theorem relating the dimension of V to the dimensions of $\ker(T)$ and $\operatorname{Im}(T)$.

Theorem 1.3.1 The Rank-Nullity Theorem (aka the dimension theorem). *Let $T : V \to W$ be a linear map, with V a finite-dimensional vector space. Then*

$$\dim V = \operatorname{rank}(T) + \operatorname{nullity}(T) = \dim \operatorname{Im}(T) + \dim \ker(T).$$

Proof. Let $n = \dim V$, and recall that if $\{v_1, \ldots, v_n\}$ is any basis for V, then $\operatorname{Im}(T) = \operatorname{Span}(\{T(v_1), \ldots, T(v_n)\})$.
First consider the case that T is injective. This means that $\ker(T) = \{0\}$, so that $\operatorname{nullity}(T) = 0$. By Proposition 1.2.3, the set $\{T(v_1), \ldots, T(v_n)\}$ is linearly independent, and since this set spans $\operatorname{Im}(T)$, it is a basis for $\operatorname{Im}(T)$, so its cardinality equals the dimension of the image, i.e., $\operatorname{rank}(T)$. Thus $\operatorname{rank}(T) = n$, and we see that

$$n = \dim V = n + 0 = \operatorname{rank}(T) + \operatorname{nullity}(T).$$

Now consider the case where $\ker(T) \neq \{0\}$. Let $\{u_1, \ldots, u_k\}$ be a basis for $\ker(T)$, hence $\operatorname{nullity}(T) = k$. Since $\{u_1, \ldots, u_k\}$ is a linearly independent set, by `[provisional cross-reference: prop-extend-independent-set-to-basis]`, it can be extended to a basis for V:

$$\{u_1, \ldots, u_k, u_{k+1}, \ldots, u_n\}$$

To establish the theorem, we need only show that $\operatorname{rank}(T) = n - k$. Since $\{u_1, \ldots, u_k, u_{k+1}, \ldots, u_n\}$ is a basis for V, $\operatorname{Im}(T) = \operatorname{Span}(\{T(u_1), \ldots, T(u_n)\})$, but we recall that $u_1, \ldots, u_k \in \ker(T)$, so that $\operatorname{Im}(T) = \operatorname{Span}(\{T(u_{k+1}), \ldots, T(u_n)\})$. Thus we know $\operatorname{rank} T \leq n - k$. To obtain an equality, we need only show that the set $\{T(u_{k+1}), \ldots, T(u_n)\}$ is linearly independent.
Suppose to the contrary, that the set is linearly independent. Then there exists scalars $a_i \in F$, not all zero, so that

$$\sum_{i=k+1}^{n} a_i T(u_i) = 0.$$

By linearity, this says $T(\sum_{i=k+1}^{r} a_i u_i) = 0$, which means $\sum_{i=k+1}^{r} a_i u_i \in \ker(T)$. But this in turn says that $\sum_{i=k+1}^{r} a_i u_i \in \operatorname{Span}(\{u_1, \ldots, u_k\})$ implying the full set $\{u_1, \ldots, u_n\}$ is linearly dependent, contradicting that it is a basis for V. This completes the proof. ■

Let's do a simple example.

Example 1.3.2 Consider the linear map $T : P_3(\mathbb{R}) \to P_3(\mathbb{R})$ given by $T(f) = f'' + f'$, where f' and f'' are the first and second derivatives of f.

The domain has dimension 4 with standard basis $\{1, x, x^2, x^3\}$, so

$$\operatorname{Im}(T) = \operatorname{Span}\{T(1), T(x), T(x^2), T(x^3)\}$$

One easily checks that $\operatorname{Im}(T) = \operatorname{Span}\{0, 1, 2 + 2x, 6x + 3x^2\} = \operatorname{Span}\{1, x, x^2\}$. At the very least we know that $\operatorname{rank}(T) \leq 3$, and since $T(1) = 0$, we must have $\operatorname{nullity}(T) \geq 1$. Now since $\{1, x, x^2\}$ is a linearly independent set, we know that $\operatorname{rank}(T) = 3$ which means that $\operatorname{nullity}(T) = 1$ by Theorem 1.3.1. It follows that $\{1\}$ is a basis for the nullspace. □

1.3.3 Computing rank and nullity

Let $A \in M_{m\times n}(F)$ be a matrix. Then $T(x) = Ax$ defines a linear map $T : F^n \to F^m$. Indeed in Subsection 1.4.2, we shall see how to translate the action of an arbitrary linear map between finite-dimensional vectors spaces into an action of a matrix on column vectors.

Let's recall how to extract the image and kernel of the linear map $x \mapsto Ax$. We know that the image of any linear map is obtained by taking the span of $T(e_1), \ldots, T(e_n)$ where $\{e_1, \ldots, e_n\}$ is any basis for F^n, the domain. Indeed if we choose the e_i to be the standard basis vectors (with a 1 in the ith coordinate and zeroes elsewhere), then $T(e_j)$ is simply the jth column of the matrix A. Thus $\operatorname{Im}(T)$ is the **column space of A.** However to determine the rank of A, we would need to know which columns form a basis. We'll get to that in a moment.

The nullspace of T, is the set of solutions to the homogeneous linear system $Ax = 0$. You may recall that a standard method to deduce the solutions is to put the matrix A in **reduced echelon form**. That means that all rows of zeros are at the bottom, the leading nonzero entry of each row is a one, and in every column containing a leading 1, all other entries are zero. These leading ones play several roles.

Proposition 1.3.3

- *Given the variables $x_1, \ldots, x_n$ in the system $Ax = 0$, a 1 in the jth column of the reduced row echelon form of A, called a **pivot**, means that the variable x_j is a **constrained variable** while the remaining variables are **free** variables. Thus if there are r pivots, there are $n - r$ free variables, and $n - r = \operatorname{nullity}(T)$; it follows that $r = \operatorname{rank}(T)$.*

- *The pivot columns of A (the columns of A in which there is a pivot in the reduced row echelon form of A) can be taken as a basis of the column space of A.*

- *The **row rank** of A (number of linearly independent rows) equals the **column rank** of A (number of linearly independent columns).*

1.3.4 Elementary Row and Column operations

The following are a series of facts about elementary row and column operations on an $m \times n$ matrix A.

- The matrix A is put in reduced row echelon form by a sequence of elementary row operations.
- Each elementary row operation can be achieved by left multiplication of A ($A \mapsto EA$) by an $m \times m$ elementary matrix.
- Each elementary column operation can be achieved by right multiplication of A ($A \mapsto AE$) by an $n \times n$ elementary matrix.
- Every elementary matrix is invertible and its inverse in again an elementary matrix of the same type.
- The rank of an $m \times n$ matrix is unchanged by elementary row or column operations, that is $\text{rank}(EA) = \text{rank}(A)$ and $\text{rank}(AE) = \text{rank}(A)$ for appropriately sized elementary matrices E.

Every invertible matrix is a product of elementary matrices, and this leads to the

Algorithm 1.3.4 *To determine whether an $n \times n$ matrix A is invertible and if so find its inverse, reduce to row-echelon form the "augmented" $n \times 2n$ matrix*

$$[A|I_n] \mapsto [R|A'].$$

The matrix A is invertible if and only if $R = I_n$, and in that case A' is the inverse A^{-1}.

Exercises

1. Let A be an $m \times n$ matrix and E an elementary matrix of the appropriate size.

- Are the row spaces of A and EA the same?
- Are the column spaces of A and AE the same?
- If R is the reduced row-echelon form of A, are the nonzero rows of R a basis for the row space of A?
- If R is the reduced row-echelon form of A, is the column space of R the same as the column space of A?

Answer. yes; yes; yes (why?); no; If $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, then $R = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, and

$$\text{Span}(\begin{bmatrix} 1 \\ 1 \end{bmatrix}) \neq \text{Span}(\begin{bmatrix} 1 \\ 0 \end{bmatrix})$$

2. Given an $m \times n$ matrix A, show that there exist (appropriately sized) elementary matrices U, V so that UAV has the form

$$UAV = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

where I_r is an $r \times r$ identity matrix with $r = \operatorname{rank}(A)$, and the other entries are all zeros.

Note that when we work with modules over a PID instead of vector spaces over a field, this construct leads to a diagonal matrix called the **Smith normal form** of the matrix A.

1.4 Coordinates and Matrices

1.4.1 Coordinate Vectors

Let V be a finite-dimensional vector space over a field F with basis $\mathcal{B} = \{v_1, \dots, v_n\}$. Since $\mathcal{B}$ is a spanning set for V, every vector $v \in V$ can be expressed as a linear combination of the vectors in $\mathcal{B}$: $v = a_1 v_1 + \cdots + a_n v_n$ with $a_i \in F$.

And, since $\mathcal{B}$ is a linearly independent set, the coefficients a_i are uniquely determined. We record those uniquely determined coefficients as

Definition 1.4.1 The **coordinate vector** of $v = a_1 v_1 + \cdots + a_n v_n$ with respect to the basis $\mathcal{B} = \{v_1, \dots, v_n\}$ is denoted as the column vector:

$$[v]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{1.4.1}$$

◊

1.4.2 Matrix of a linear map

Let V and W be two finite-dimensional vector spaces defined over a field F. Suppose that $\dim V = n$ and $\dim W = m$, and we choose bases $\mathcal{B} = \{v_1, \dots, v_n\}$ for V, and $\mathcal{C} = \{w_1, \dots, w_m\}$ for W. By Theorem 1.1.6, any linear map $T : V \to W$ is completely determined by the set of vectors $\{T(v_1), \dots, T(v_n)\}$, and since $\mathcal{C}$ is a basis for W, for each index j, there are uniquely determined scalars $a_{ij} \in F$ with

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

We record that data as a matrix A with $A_{ij} = a_{ij}$. We define the **matrix of**

T **with respect to the bases $\mathcal{B}$ and $\mathcal{C}$**, as

$$[T]_{\mathcal{B}}^{\mathcal{C}} = A = [a_{ij}] \tag{1.4.2}$$

Example 1.4.2 The companion matrix of a polynomial. Let $f = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ be a polynomial with coefficients in a field F. Let V be a finite-dimensional vector space over the field F with basis $\mathcal{B} = \{v_1, \ldots, v_n\}$. Define a linear map $T : V \to V$ (called an **endomorphism** or **linear operator** since the domain and codomain are the same vector space) by:

$$\begin{aligned} T(v_1) &= v_2 \\ T(v_2) &= v_3 \\ &\vdots \\ T(v_{n-1}) &= v_n \\ T(v_n) &= -a_0v_1 - a_1v_2 - \cdots - a_{n-1}v_{n-1}. \end{aligned}$$

The matrix of T with respect to the basis $\mathcal{B}$ is called the **companion matrix of** f, and is given by

$$[T]_{\mathcal{B}} := [T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 0 & & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & 0 & 1 & -a_{n-1} \end{bmatrix}$$

One can show that both the **minimal polynomial** and characteristic polynomial of this companion matrix is the polynomial f. The companion matrix is an essential component in the **rational canonical form** of an arbitrary square matrix A where the polynomials f that occur are the **invariant factors** associated to A. □

Observation 1.4.3 When constructing the matrix of a linear map, it is very useful to recognize the connection with coordinate vectors. For example in constructing the matrix $[T]_{\mathcal{B}}^{\mathcal{C}}$ in (1.4.2), the jth column of the matrix is the coordinate vector $[T(v_j)]_{\mathcal{C}}$. Thus a mnemonic device for remembering how to construct the matrix of a linear map is that

$$[T]_{\mathcal{B}}^{\mathcal{C}} = A = [a_{ij}] = \begin{bmatrix} | & | & \cdots & | \\ [T(v_1)]_{\mathcal{C}} & [T(v_2)]_{\mathcal{C}} & \cdots & [T(v_n)]_{\mathcal{C}} \\ | & | & \cdots & | \end{bmatrix}. \tag{1.4.3}$$

1.4.3 Matrix associated to a composition

Suppose that U, V, and W are vector spaces over a field F, and $S : U \to V$ and $T : V \to W$ are linear maps. The the composition $T \circ S$ (usually denoted TS) is a linear map, $T \circ S : U \to W$.

Now suppose that all three vector spaces are finite-dimensional, say $\dim U = n$, $\dim V = p$, and $\dim W = m$, with bases $\mathcal{B}_U, \mathcal{B}_V, \mathcal{B}_W$. If we consider the matrices of the corresponding linear maps, we see that the matrix sizes are

$$\begin{aligned} [S]_{\mathcal{B}_U}^{\mathcal{B}_V} &\text{ is } p \times n \\ [T]_{\mathcal{B}_V}^{\mathcal{B}_W} &\text{ is } m \times p \\ [TS]_{\mathcal{B}_U}^{\mathcal{B}_W} &\text{ is } m \times n \end{aligned}$$

The fundamental result connecting these is

Theorem 1.4.4 Matrix of a composition.

$$[TS]_{\mathcal{B}_U}^{\mathcal{B}_W} = [T]_{\mathcal{B}_V}^{\mathcal{B}_W} [S]_{\mathcal{B}_U}^{\mathcal{B}_V} \tag{1.4.4}$$

This result will be of critical importance when we discuss **change of basis.**

As more or less a special case of the above theorem, we have the corresponding result with coordinate vectors: that the coordinate vector of $T(v)$ is the product of the matrix of T with the coordinate vector of v. More precisely,

Corollary 1.4.5 *With the notation as above, for $v \in V$*

$$[T(v)]_{\mathcal{B}_W} = [T]_{\mathcal{B}_V}^{\mathcal{B}_W} [v]_{\mathcal{B}_V}.$$

Example 1.4.6 Let $V = P_4(\mathbb{R})$ and $W = P_3(\mathbb{R})$ be the vector spaces of polynomials with coefficients in $\mathbb{R}$ having degree less than or equal to 4 and 3 respectively. Let $D : V \to W$ be the (linear) derivative map, $D(f) = f'$, where f' is the usual derivative for polynomials. Let's take standard bases for V and W, namely $\mathcal{B}_V = \{1, x, x^2, x^3, x^4\}$ and $\mathcal{B}_W = \{1, x, x^2, x^3\}$. One computes:

$$[D]_{\mathcal{B}_V}^{\mathcal{B}_W} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Let $f = 2 + 3x + 5x^3$. We know of course that $D(f) = 3 + 15x^2$, but we want to see this with coordinate vectors. We know that

$$[f]_{\mathcal{B}_V} = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 5 \\ 0 \end{bmatrix} \text{ and } [D(f)]_{\mathcal{B}_W} = \begin{bmatrix} 3 \\ 0 \\ 15 \\ 0 \end{bmatrix}$$

and verify that

$$[D(f)]_{\mathcal{B}_W} = \begin{bmatrix} 3 \\ 0 \\ 15 \\ 0 \end{bmatrix} = [D]_{\mathcal{B}_V}^{\mathcal{B}_W} [f]_{\mathcal{B}_V} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 0 \\ 5 \\ 0 \end{bmatrix}.$$

□

1.4.4 Change of basis

A **change of basis** or **change of coordinates** is an enormously useful concept. It plays a pivotal role in diagonalization, triangularization, and more generally in putting a matrix into a **canonical form**. It's practical uses are easy to envision. We may think of the usual orthonormal basis of $\mathbb{R}^3$ along the coordinate axes as the **standard basis** for $\mathbb{R}^3$, but when one want to create computer graphics which projects the image of an object onto a plane, the natural frame includes a direction parallel to the line of sight of the observer, so it defines a **natural basis** for this application.

First, let's understand what we are doing intuitively. Suppose our vector space $V = \mathbb{R}^3$, and we have two bases for it with elements written as row vectors, $\mathcal{B}_1 = \{e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)\}$ and $\mathcal{B}_2 = \{v_1 = (1,1,1), v_2 = (0,1,1), v_3 = (0,0,1)\}$.

Checkpoint 1.4.7 Is $\mathcal{B}_2$ really a basis? Let's recall a useful fact that allows us to quickly verify that $\mathcal{B}_2$ is actually a basis for $\mathbb{R}^3$. While in principle we must check the set is both linearly independent and spans $\mathbb{R}^3$, since we know the dimension of $\mathbb{R}^3$, and the set has 3 elements, it follows that either condition implies the other.

Hint. To show $\mathcal{B}_2$ spans, it is enough to show that $\operatorname{Span}(\mathcal{B}_2)$ contains a spanning set for $\mathbb{R}^3$

Normally when we think of a vector in $\mathbb{R}^3$, we think of it as a coordinate vector with respect to the standard basis, so that a vector we write as $v = (a,b,c)$ is really the coordinate vector with respect to the standard basis:

$$v = [v]_{\mathcal{B}_1} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

The problem is when we want to find $[v]_{\mathcal{B}_2}$. For some vectors this is easy. For example,

$$[v]_{\mathcal{B}_1} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ is equivalent to } [v]_{\mathcal{B}_2} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

or

$$[v]_{\mathcal{B}_1} = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix} \text{ is equivalent to } [v]_{\mathcal{B}_2} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},$$

but what is going on in general?

Recall from Corollary 1.4.5, that for a linear transformation $T : V \to W$, and $v \in V$ that

$$[T(v)]_{\mathcal{B}_W} = [T]_{\mathcal{B}_V}^{\mathcal{B}_W} [v]_{\mathcal{B}_V}.$$

In our current situation $V = W$ and T is the identity transformation, $T(v) = v$, which we shall denote by I, so that

$$[v]_{\mathcal{B}_2} = [I]_{\mathcal{B}_1}^{\mathcal{B}_2} [v]_{\mathcal{B}_1}.$$

The matrix $[I]_{\mathcal{B}_1}^{\mathcal{B}_2}$ is called the **change of basis** or **change of coordinates** matrix (converting $\mathcal{B}_1$ coordinates to $\mathcal{B}_2$ coordinates), and these change of basis matrices come in pairs

$$[I]_{\mathcal{B}_1}^{\mathcal{B}_2} \text{ and } [I]_{\mathcal{B}_2}^{\mathcal{B}_1}.$$

Now in our case, both matrices are easy to compute:

$$[I]_{\mathcal{B}_1}^{\mathcal{B}_2} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \text{ and } [I]_{\mathcal{B}_2}^{\mathcal{B}_1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix},$$

and it should come as no surprise that the columns of the second are just the elements of the $\mathcal{B}_2$-basis in standard coordinates. But the nice part is that the first matrix is related to the second affording a means to compute it when computations by hand are not so simple.

Using Equation (1.4.4) on the matrix of a composition

$$[TS]_{\mathcal{B}_U}^{\mathcal{B}_W} = [T]_{\mathcal{B}_V}^{\mathcal{B}_W} [S]_{\mathcal{B}_U}^{\mathcal{B}_V},$$

with $V = U = W$, and $T = S = I$, we arrive at

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [I]_{\mathcal{B}_1}^{\mathcal{B}_1} = [I]_{\mathcal{B}_1}^{\mathcal{B}_2} [I]_{\mathcal{B}_2}^{\mathcal{B}_1},$$

that is $[I]_{\mathcal{B}_1}^{\mathcal{B}_2}$ and $[I]_{\mathcal{B}_2}^{\mathcal{B}_1}$ are inverse matrices, and this is always the case.

Theorem 1.4.8 *Given two bases $\mathcal{B}_1$ and $\mathcal{B}_2$ for a finite-dimensional vector space V, the change of basis matrices $[I]_{\mathcal{B}_1}^{\mathcal{B}_2}$ and $[I]_{\mathcal{B}_2}^{\mathcal{B}_1}$ are inverse matrices.*

Finally we apply this to the matrix of a linear map $T : V \to V$ on a finite-dimensional vector space V with bases $\mathcal{B}_1$ and $\mathcal{B}_2$:

Theorem 1.4.9

$$[T]_{\mathcal{B}_2} = [I]_{\mathcal{B}_1}^{\mathcal{B}_2} [T]_{\mathcal{B}_1} [I]_{\mathcal{B}_2}^{\mathcal{B}_1}.$$

Example 1.4.10 We often express the matrix of a linear map in terms of the standard basis, but many times such a matrix is complicated and does not easily reveal what the linear map is actually doing. For example, using our bases $\mathcal{B}_1$ and $\mathcal{B}_2$ for $\mathbb{R}^3$ given above, suppose we have a linear map $T : \mathbb{R}^3 \to \mathbb{R}^3$ whose matrix with respect to the standard basis $\mathcal{B}_1$ is

$$[T]_{\mathcal{B}_1} = \begin{bmatrix} 4 & 0 & 0 \\ -1 & 5 & 0 \\ -1 & -1 & 6 \end{bmatrix}.$$

It is easy enough to compute the value of T on a given vector (recall from equation (1.4.3), the columns of the above matrix are simply $T(v_1), T(v_2), T(v_3)$ written with respect to the standard basis ($\mathcal{B}_1$) for $\mathbb{R}^3$).

However, using Theorem 1.4.9, we compute

$$[T]_{\mathcal{B}_2} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 6 \end{bmatrix},$$

which makes much clearer how the map T is acting on $\mathbb{R}^3$ (strecthing by a factor of 4, 5, 6 in the directions of w_1, w_2, w_3. □

1.5 Eigenvalues, eigenvectors, diagonalization

1.5.1 The big picture

Given a linear operator $T : V \to V$ on a finite-dimensional vector space V, T is said to be **diagonalizable** if there exists a basis $\mathcal{E} = \{v_1, \ldots, v_n\}$ of V so that the matrix of T with respect to $\mathcal{E}$ is diagonal:

$$[T]_{\mathcal{E}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}$$

where the λ_i are scalars in F, not necessarily distinct. A trivial example is the identity linear operator which is diagonalizable with respect to any basis and its matrix is the $n \times n$ identity matrix.

Note that the diagonal form of the matrix above encodes the information, $T(v_i) = \lambda_i v_i$ for $i = 1, \ldots, n$.

In general, given a linear map $T : V \to V$ on a vector space V over a field F, one can ask whether for a given scalar $\lambda \in F$, there exist nonzero vectors $v \in V$, so that $T(v) = \lambda v$. If they exist, λ is called an **eigenvalue** of T, and $v \neq 0$ an **eigenvector** for T corresponding to the eigenvalue λ. Thus T is **diagonalizable** if and only if there is a basis for V consisting of eigenvectors for T.

While at first glance this may appear an odd notion, consider the case of $\lambda = 0$. Asking for a nonzero vector v so that $T(v) = 0v = 0$ is simply asking whether T has a nontrivial kernel.

Let's look at several examples. Let $U = \mathbb{R}[x]$ be the vector space of all polynomials with coefficients in $\mathbb{R}$, and let $V = C^\infty(\mathbb{R})$ be the vector space of all functions which are infinitely differentiable. Note that U is a subspace of V.

Example 1.5.1 **$T : \mathbb{R}[x] \to \mathbb{R}[x]$ given by $T(f) = f'$.** Let $T : \mathbb{R}[x] \to \mathbb{R}[x]$ be the linear map which takes a polynomial to its first derivative, $T(f) = f'$. Does T have any eigenvectors or eigenvalues?

We must ask how is it possible that

$$T(f) = f' = \lambda f$$

for a nonzero polynomial f?

If $\lambda \neq 0$, there can be no nonzero f since the degrees of f' and λf differ by one. So the only possibility left is $\lambda = 0$. Do we know any nonzero polynomials f so that $T(f) = f' = 0 \cdot f = 0$? Calculus tells us that the only solution to the problem are the constant polynomials. Well maybe not so interesting, but still instructive. $\square$

Example 1.5.2 $T : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ **given by** $T(f) = f'$**.** Next consider $T : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ to be the same derivative map, but now on the vector space $V = C^\infty(\mathbb{R})$. We consider the same problem of finding scalars λ and nonzero functions f so that

$$f' = \lambda f.$$

Once again, calculus solves this problem completely as the functions f are simply the solutions to the first order homogeneous linear differential equation $y' - \lambda y = 0$, the solutions to which are all of the form $f(x) = Ce^{\lambda x}$. Note this includes $\lambda = 0$ from the previous case. $\square$

Example 1.5.3 $S : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ **given by** $S(f) = f''$**.** Finally consider the map $S : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ given by $S(f) = f''$, the second derivative map, so now we seek functions for which $S(f) = f'' = \lambda f$, or in calculus terms solutions to the second order homogeneous differential equation

$$y'' - \lambda y = 0.$$

This is an interesting example since the answer depends on the sign of λ. For $\lambda = 0$, the fundamental theorem of calculus tells us that solutions are all linear polynomials $f(x) = ax + b$.

For $\lambda < 0$, we can write $\lambda = -\omega^2$. We see that $\sin(\omega x)$ and $\cos(\omega x)$ are eigenvectors for S with eigenvalue $\lambda = -\omega^2$. Indeed every eigenvector with eigenvalue $\lambda = -\omega^2 < 0$ is a linear combination of these two.

For $\lambda > 0$, we write $\lambda = \omega^2$, we see that $e^{\pm\omega x}$ are solutions and as above every eigenvector with eigenvalue $\lambda = \omega^2 > 0$ is a linear combination of these two. $\square$

With a few examples under our belt, we return to the problem of finding a systematic way to determine eigenvalues and eigenvectors. The condition $T(v) = \lambda v$ is the same as the condition that $(T - \lambda I)v = 0$, where I is the identity linear operator ($I(v) = v$) on V. So let's put

$$E_\lambda = \{v \in V \mid T(v) = \lambda v\}.$$

Then as we just said, $E_\lambda = \ker(T - \lambda I)$, so we know that E_λ is a subspace of V, called **the** λ**−eigenspace** of T.

Since E_λ is a subspace of V, 0 is always an element, but $T(0) = \lambda 0 = 0$ for any λ which is not terribly discriminating, and our goal is to find a basis of the space consisting of eigenvectors, so the zero vector must be excluded.

On a finite-dimensional vector space, finding the eigenvalues and a basis for the corresponding eigenspace is rather algorithmic, at least in principle. Let A

be the matrix of T with respect to any basis $\mathcal{B}$ (it does not matter which). Since $T(v) = \lambda v$ if and only if

$$A[v]_{\mathcal{B}} = [T]_{\mathcal{B}}[v]_{\mathcal{B}} = [T(v)]_{\mathcal{B}} = [\lambda v]_{\mathcal{B}} = \lambda[v]_{\mathcal{B}},$$

we can simply describe how to find eigenvalues of the matrix A.

So now we are looking for scalars λ for which there are nonzero vectors $v \in F^n$ with $Av = \lambda v$. As before, it is more useful to phrase this as seeking values of λ for which $(A - \lambda I_n)$ has a nontrivial kernel. But now remember that $(A - \lambda I_n) : F^n \to F^n$ is a linear operator on F^n, so it has a nontrivial kernel if and only if it is not invertible, and invertibility can be detected with the determinant. Thus $E_\lambda \neq 0$ if and only if $\det(A - \lambda I) = 0$.

Remark 1.5.4 Since for any $n \times n$ matrix B, $\det(-B) = (-1)^n \det B$, we have $\det(A - \lambda I_n) = 0$ if and only if $\det(\lambda I_n - A) = 0$. One of these expression is more convenient for the theory, while the other one is more convenient for computation.

Since we want to find all values of λ with $\det(\lambda I_n - A) = 0$, we set the problem up with a variable and define the function

$$\chi_A(x) := \det(xI - A).$$

One shows that χ_A is a monic polynomial of degree n, called the **characteristic polynomial** of A. The roots of this polynomial are the **eigenvalues** of A, so the first part of the algorithm is to find the roots of the characteristic polynomial. In particular, an $n \times n$ matrix can have at most n eigenvalues in F, counted with multiplicity.

Now for each eigenvalue λ, there is a corresponding eigenspace, E_λ which is the kernel of $\lambda I_n - A$, or equivalently of $A - \lambda I_n$. Finding the kernel is simply finding the solutions for the system of homogeneous linear equations $(A - \lambda I_n)X = 0$, which one can easily do via row reduction.

1.5.2 Taking stock of where we are

- Given a matrix $A \in M_n(F)$, we consider the characteristic polynomial $\chi_A(x) = \det(xI - A)$ which is a monic polynomial of degree n in $F[x]$. When $F = \mathbb{C}$ (or any algebraically closed field), χ_A is guaranteed to have all of its roots in F, but not so otherwise. For example, if $F = \mathbb{R}$ and

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, B = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}, \text{ and } C = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

then $\chi_A(x) = x^2+1$ and $\chi_B(x) = (x-4)(x^2+1)$, so neither A nor B has all its eigenvalues in $F = \mathbb{R}$. On the other hand, $\chi_C(x) = (x-4)(x-1)(x+1)$ does have all its eigenvalues in F.

- So in the general case, a matrix $A \in M_n(F)$ will have a characteristic polynomial χ_A exhibiting a factorization of the form:

$$\chi_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r} q(x),$$

where either $q(x)$ is the constant 1 or is a polynomial of degree ≥ 2 with no roots in F. It will follow that if $q(x) \neq 1$, then A cannot be diagonalized, though something can still be said.

- Let's assume that

$$\chi_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r},$$

with $\lambda_1, \ldots, \lambda_r$ the distinct eigenvalues of A in F. The exponents m_i are called the **algebraic multiplicities** of the corresponding eigenvalues.

By comparing degrees, we see that

$$n = m_1 + \cdots + m_r.$$

Moreover since the λ_k are roots of the characteristic polynomial, we know that $\det(A - \lambda_k I) = 0$, which guarantees that $E_{\lambda_k} \neq \{0\}$. Indeed, it is not hard to show that

$$1 \leq \dim E_{\lambda_k} \leq m_k, \text{ for } k = 1, \ldots, r. \tag{1.5.1}$$

Another important result is the

Proposition 1.5.5 *Suppose that the matrix A has distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, and that the eigenspace E_{λ_k} has basis $\mathcal{B}_k$, $k = 1, \ldots, r$. Then $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_r$ is a linearly independent set.*

Now recall that a linear operator $T : V \to V$ (resp. square matrix $A \in M_n(F)$) is diagonalizable if and only if there is a basis of V (resp. F^n) consisting of eigenvectors for T. From the proposition above, the largest linearly independent set of eigenvectors which can be constructed has size

$$\begin{aligned} |\mathcal{B}| &= \dim E_{\lambda_1} + \cdots + \dim(E_{\lambda_r}) \\ &\leq m_1 + \cdots + m_r = n = \dim V. \end{aligned}$$

We summarize our results as

Theorem 1.5.6 Diagonalizability criterion. *A matrix $A \in M_n(F)$ is diagonalizable if and only if*

- *The characteristic polynomial χ_A factors into linear factors over F:*

$$\chi_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r}$$

with the λ_i distinct, and

- $\dim E_{\lambda_i} = m_i$, *for $i = 1, \ldots, r$.*

Corollary 1.5.7 A sufficient condition for diagonalizability. *Suppose the matrix $A \in M_n(F)$ has characteristic polynomial which factors into distinct linear factors over F:*

$$\chi_A(x) = (x - \lambda_1) \cdots (x - \lambda_n)$$

with the λ_i distinct. Then A is diagonalizable.

Proof. We know that there are n eigenspaces each with dimension at least one which gives at least n linearly independent eigenvectors. As F^n is n-dimensional, these form a basis for the space, so A is diagonalizable. ■

1.5.3 An alternate characterization of diagonalizable

We want to make sense of an alternate definition that an $n \times n$ matrix $A \in M_n(F)$is **diagonalizable** if there is an invertible matrix $P \in M_n(F)$, so that $D = P^{-1}AP$ is a diagonal matrix. Recall that in this setting we say that the matrix A is similar to a diagonal matrix.

Suppose that the matrix A is given to us as the matrix of a linear transformation $T : V \to V$ with respect to a basis $\mathcal{B}$ for V, $A = [T]_{\mathcal{B}}$. Now T is diagonalizable if and only if there is a basis $\mathcal{E}$ of V consisting of eigenvectors for T. We know that $[T]_{\mathcal{E}}$ is diagonal. But we recall from Theorem 1.4.9 that

$$[T]_{\mathcal{E}} = [I]_{\mathcal{B}}^{\mathcal{E}}[T]_{\mathcal{B}}[I]_{\mathcal{E}}^{\mathcal{B}} = P^{-1}AP,$$

where $P = [I]_{\mathcal{E}}^{\mathcal{B}}$ is the invertible matrix. Also note that when $\mathcal{B}$ is a standard basis, the columns of $P = [I]_{\mathcal{E}}^{\mathcal{B}}$ are simply the coordinate vectors of the eigenvector basis $\mathcal{E}$. This is quite a mouthful, so we should look at some examples.

Example 1.5.8 A simple example to start. Let $A = \begin{bmatrix} 5 & 6 & 0 \\ 0 & 5 & 8 \\ 0 & 0 & 9 \end{bmatrix}$. Then $\chi_A(x) = (x-5)^2(x-9)$, so we have two eigenvalues 5 and 9. We need to compute the corresponding eigenspaces.

For each eigenvalue λ, we compute $\ker(A - \lambda I_3)$, that is find all solutions to $(A - \lambda I_3)x = \mathbf{0}$.

$$A - 9I = \begin{bmatrix} -4 & 6 & 0 \\ 0 & -4 & 8 \\ 0 & 0 & 0 \end{bmatrix} \overset{RREF}{\mapsto} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix},$$

so $E_9(A) = \ker(A - 9I) = \operatorname{Span}\left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right\}$. Similarly,

$$A - 5I = \begin{bmatrix} 0 & 6 & 0 \\ 0 & 0 & 8 \\ 0 & 0 & 4 \end{bmatrix} \overset{RREF}{\mapsto} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

so $E_5(A) = \ker(A - 5I) = \operatorname{Span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}\right\}$.

But $\left\{\begin{bmatrix}3\\2\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}\right\}$ is not a basis for $\mathbb{R}^3$, so A is not diagonalizable. $\square$

Remark 1.5.9 It is important to note in the example above that if we simply wanted to know whether A is diagonalizable or not, we did not have to do all of this work. Diagonalizability is possible if and only if the algebraic multiplicity of each eigenvalue equals the dimension of the corresponding eigenspace. An eigenvalue with algebraic multiplicity one (a simple root of χ_A) will always have a one-dimensional eigenspace, so the issue for us was discovering that $\dim E_5(A) = 1$ while the algebraic multiplicity of $\lambda = 5$ is 2.

Example 1.5.10 A more involved example. Let $A = \begin{bmatrix}3&0&2&0\\1&3&1&0\\0&1&1&0\\0&0&0&4\end{bmatrix}$.

Think of A as $A = [T]_{\mathcal{B}}$, the matrix of the linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^4$ with respect to the standard basis $\mathcal{B}$ of $\mathbb{R}^4$. Then A has characteristic polynomial $\chi_A(x) = x^4 - 11x^3 + 42x^2 - 64x + 32 = (x-1)(x-2)(x-4)^2$.

We know that the eigenspaces E_1 and E_2 will each have dimension one, so are no obstruction to diagonalizability, but since we want to do a bit more with this example, we compute bases for the eigenspaces. If we let $\mathcal{E}_\lambda$ denote a basis for the eigenspace $E_\lambda = \ker(A - \lambda I)$, then as in the previous example via row reduction, we find $\mathcal{E}_1 = \left\{v_1 = \begin{bmatrix}1\\0\\-1\\0\end{bmatrix}\right\}$ and $\mathcal{E}_2 = \left\{v_2 = \begin{bmatrix}2\\-1\\-1\\0\end{bmatrix}\right\}$.

By Equation (1.5.1), we know that $1 \le \dim E_4 \le 2$. If the dimension is 1, then A is not diagonalizable. As it turns out the dimension is 2, and $\mathcal{E}_4 = \{v_3, v_4\} = \left\{\begin{bmatrix}2\\3\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right\}$ is a basis for E_4.

Let $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_4 = \{v_1, v_2, v_3, v_4\}$ be the basis of eigenvectors. Then

$$D = [T]_{\mathcal{E}} = \begin{bmatrix}1&0&0&0\\0&2&0&0\\0&0&4&0\\0&0&0&4\end{bmatrix} = P^{-1}AP,$$

where

$$P = [I]_{\mathcal{E}}^{\mathcal{B}} = \begin{bmatrix} 1 & 2 & 2 & 0 \\ 0 & -1 & 3 & 0 \\ -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note how the columns of P are the (coordinate vectors of) the eigenvector basis. $\square$

1.6 Minimal and characteristic polynomials

We review a few important facts about minimal and characteristic polynomials.

1.6.1 Annihilating polynomials

Let $A \in M_n(F)$ be a square matrix. One can ask if there is a nonzero polynomial $f(x) = a_m x^m + \cdots + a_0 \in F[x]$ for which $f(A) = a_m A^m + \cdots + a_1 A + a_0 I_n = \mathbf{0}$, the zero matrix. If we think of trying to find a polynomial, this may seem a challenging task.

However, if we consider that $M_n(F)$ is a vector space of dimension n^2, then Theorem 1.1.2 tells us that the set

$$\{I_n, A, A^2, \ldots, A^{n^2}\}$$

must be a linearly dependent set, and that means there are scalars $a_0, a_1, \ldots, a_{n^2} \in F$, not all zero, for which $a_{n^2} A^{n^2} + \cdots + a_1 A + a_0 I_n = \mathbf{0}$, so that $f(x) = a_{n^2} x^{n^2} + \cdots + a_0$ is one nonzero polynomial which annihilates A.

1.6.2 The minimal polynomial

Given a matrix $A \in M_n(F)$, we have seen there is a nonzero polynomial which annihilates it, so we consider the set

$$J = \{f \in F[x] \mid f(A) = \mathbf{0}\}.$$

In the language of abstract algebra, J is an ideal in the polynomial ring $F[x]$, and since F is a field, $F[x]$ is a PID (principal ideal domain), the ideal J is principally generated: $J = \langle \mu_A \rangle$, where μ_A is the monic generator of this ideal. In less technical terms, μ_A is the monic polynomial of least degree which annihilates A, and every element of J is a (polynomial) multiple of μ_A. The polynomial μ_A is called the **minimal polynomial** of the matrix A.

A more constructive version of finding the minimal polynomial comes from the observation that if $f, g \in J$, that if $f(A) = g(A) = \mathbf{0}$, then $h(A) = 0$, where h is the greatest common divisor **gcd**, of f and g. In particular, if $f(A) = \mathbf{0}$, then μ_A must divide f, so if we can factor f, there are only finitely many possibilities for μ_A.

Example 1.6.1 $\boldsymbol{A^8 = I_n}$**.** Let's suppose that $A \in M_n(\mathbb{Q})$ and $A^8 = I_n$. This means that $f(x) = x^8 - 1$ is a polynomial which annihilates A, so μ_A must divide it. Over $\mathbb{Q}$, we have the following factorization into irreducibles:

$$x^8 - 1 = \Phi_8\Phi_4\Phi_2\Phi_1 = (x^4 + 1)(x^2 + 1)(x + 1)(x - 1),$$

where (for those with abstract algebra background) the Φ_d are the dth cyclotomic polynomials defined recursively as an (irreducible) factorization over $\mathbb{Q}$ by

$$x^n - 1 = \prod_{d|n} \Phi_d.$$

Thus $\Phi_1 = (x-1)$, $x^2-1 = \Phi_1\Phi_2$, so $\Phi_2 = x+1$, and x^8-1 has the factorization given above. □

1.6.3 The characteristic polynomial

Given a matrix $A \in M_n(F)$, we have seen that there is a polynomial of degree at most n^2 which annihilates A, and given one such nonzero polynomial there is one of minimal degree. But the key to finding a minimal polynomial is obtaining at least one. The idea of trying to find a linear dependence relation among $I_n, A, A^2, \ldots, A^{n^2}$ is far from appealing, but fortunately there is a polynomial we have used before which annhilates A.

Theorem 1.6.2 Cayley-Hamilton. *Let $A \in M_n(F)$, and $\chi_A(x) = \det(xI_n - A)$ be its characteristic polynomial. Then $\chi_A(A) = \mathbf{0}$, that is χ_A is a monic polynomial of degree n which annihilates A.*

In particular, the minimal polynomial, μ_A, divides the characteristic polynomial, χ_A.

Example 1.6.3 Are there any elements of order 8 in $\boldsymbol{GL_3(\mathbb{Q})}$? The question asks whether there is an invertible 3×3 matrix A so that 8 is the smallest positive integer k with $A^k = I_3$.

Since $A^8 = I_3$, we know $\det A \neq 0$, so such a matrix will necessarily be invertible, hence an element of $GL_3(\mathbb{Q})$. In the example above, we saw that any matrix which satisfies $A^8 = I_3$ must have minimal polynomial μ_A which divides $x^8 - 1 = (x^4 + 1)(x^2 + 1)(x + 1)(x - 1)$. But the Cayley-Hamilton theorem tells us that μ_A must also divide the characteristic polynomial χ_A which must have degree 3, and the only way to create a polynomial of degree 3 with the factors listed above is to have $\chi_A \mid x^4-1$, which forces $A^4 = I_3$, so there are no elements of order 8 in $GL_3(\mathbb{Q})$. □

On the other hand,

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

has $\mu_A = \chi_A = x^4 + 1$, so A is an element of order 8 in $GL_4(\mathbb{Q})$. The matrix A is the **companion matrix** to the polynomial $x^4 + 1$. See Example 1.4.2 for more detail.

1.7 Some Sage examples

Here are some common uses of Sage with linear algebra applications.

1.7.1 Row reduction, echelon form, kernel, column space

A random 4×5 rational matrix with rank 3. Note that while every matrix has a reduced row echelon form, this algorithm will generate a matrix of a desired size and rank, over a desired ring, whose reduced row-echelon form has only integral values.

```
%display latex
latex.matrix_delimiters("[", "]")
A=random_matrix(QQ,4,5,algorithm='echelonizable', rank=3,
   upper_bound=10); A
```

The example below was chosen so that the pivots were not all in the first three columns.

```
%display latex
latex.matrix_delimiters("[", "]")
A = matrix(QQ,4,5,[[0 ,0 , 0 , 1 , 1],
[0 , 0 , 1 , 2 , 5],
[-1 , 5 , 0 , 6 , 8 ],
[0 , 0 , 1 , -1 , 2]]); A
```

It's reduced row echelon form which allows us to confirm the rank and make it easier to find the kernel.

```
A.echelon_form()
```

The kernel or nullspace of the matrix, $\{x \in \mathbb{Q}^5 \mid Ax = 0\}$, written as row vectors.

```
A.right_kernel()
```

The column space of the matrix: a basis of column space written as row vectors; note they are the pivot columns.

```
[A.column(i) for i in A.pivots()]
```

The following is simply a basis for the column space of A written as row vectors. Note that this is the reduced row-echelon form of the matrix formed from the rows above: A matrix and its RREF have the same row space.

```
A.column_space()
```

1.7.2 Eigenvalues, eigenvectors, and diagonalization

Generate a diagonalizable 8×8 integer matrix.

```
%display latex
latex.matrix_delimiters("[", "]")
B=random_matrix(ZZ,8,8,algorithm='diagonalizable')
B
```

Compute the characteristic polynomial and factor it. The characteristic polynomial will necessarily factor into linear factors. To make things more interesting, run the Sage script until you get a characteristic polynomial with some algebraic multiplicities greater than one.

```
B.characteristic_polynomial().factor()
```

Compute the eigenvalues and bases for the corresponding eigenspaces. The output is a list giving each eigenvalue and a basis for the corresponding eigenspace. Watch for these to show up as the columns of the change of basis matrix.

```
B.eigenspaces_right()
```

Another way of getting the same data

```
B.eigenvectors_right()
```

The diagonalized matrix $D = P^{-1}BP$ where P is the change of basis matrix whose columns are the eigenvectors spanning the eigenspaces.

```
B.eigenmatrix_right()
```

1.7.3 Rational and Jordan canonical forms

Example slightly modified from the Sage Reference Manual[1].

```
%display latex
latex.matrix_delimiters("[", "]")
C=matrix(QQ,8,[[0,-8,4,-6,-2,5,-3,11], \
```

[1] doc.sagemath.org/pdf/en/reference/matrices/matrices.pdf

```
[-2,-4,2,-4,-2,4,-2,6], [5, 14, -7, 12, 3,-8,6,-27], \
[-3,8,7,-5,0,2,6,17], [0,5,0,2,4, -4, 1, 2], \
[-3, -7, 5, -6, -1, 5, -4, 14], \
[6, 18, -10, 14, 4, -10, 10, -28], \
[-2, -6, 4, -5, -1, 3, -3, 13]]);C
```

We see the factored characteristic polynomial is divisible by a quadratic which is irreducible over $\mathbb{Q}$, so the matrix will have a rational canonical form, but not a Jordan form over $\mathbb{Q}$.

```
C.characteristic_polynomial().factor()
```

Here is the minimal polynomial, the largest of the invariant factors.

```
m=C.minimal_polynomial()
m,m.factor()
```

Here is a list of the invariant factors, given as a lists of coefficients of the polynomials they represent.

```
C.rational_form(format='invariants')
```

Here we turn those lists into polynomials. The rational canonical form is a block diagonal matrix with each block being the **companion matrix**.

```
invariants=C.rational_form(format='invariants')
R=PolynomialRing(QQ,'x')
[R(p).factor() for p in invariants]
```

The matrix C is not diagaonalizable over any field since the minimal polynomial has a multiple root.

```
C.rational_form(format='right')
```

Since the minimal(characteristic) polynomial has an irreducible quadratic factor, we need to extend the field $\mathbb{Q}$ to a quadratic extension which contains a root in order to produce a Jordan form.

```
K.<a>=NumberField(x^2+6*x-20);K
```

Now C has a Jordan canonical form over the field K.

```
C.jordan_form(K)
```

```
%display latex
latex.matrix_delimiters("[", "]")
D=matrix(QQ,8,[[0,-8,4,-6,-2,5,-3,11], \
[-2,-4,2,-4,-2,4,-2,6], [5, 14, -7, 12, 3,-8,6,-27], \
```

```
[-3,-8,7,-5,0,2,-6,17], [0,5,0,2,4, -4, 1, 2], \
[-3, -7, 5, -6, -1, 5, -4, 14], \
[6, 18, -10, 14, 4, -10, 10, -28], \
[-2, -6, 4, -5, -1, 3, -3, 13]]);D
```

Example taken from the Sage Reference Manual[2], has all invariant factors a power of $(x-2)$.

```
D.characteristic_polynomial().factor()
```

```
m=D.minimal_polynomial()
m,m.factor()
```

```
invariants=D.rational_form(format='invariants')
R=PolynomialRing(QQ,'x')
[R(p).factor() for p in invariants]
```

```
D.rational_form(format='right')
```

```
D.jordan_form()
```

1.8 Exercises (with solutions)

Exercises

1. Let H be the subset of $\mathbb{R}^4$ defined by

$$H = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} : x_1 + x_2 + x_3 + x_4 = 0 \right\}.$$

Either show that H is a subspace of $\mathbb{R}^4$, or demonstrate how it fails to have a necessary property.

Solution. The easiest way to show that H is a subspace is to note that it is the kernel of a linear map. Let A be the 1×4 matrix $A = [1\ 1\ 1\ 1]$. Then

$$H = \{x \in \mathbb{R}^4 \mid Ax = 0\},$$

is the nullspace of A, which is always a subspace.

Alternatively of course you could check that 0 is in the set and that it is closed under addition and scalar multiplication.

[2] doc.sagemath.org/pdf/en/reference/matrices/matrices.pdf

2. Suppose that $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear map satisfying

$$T\left(\begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 6 \\ -3 \\ 6 \end{bmatrix}, T\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \text{ and } T\left(\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 6 \\ 2 \end{bmatrix}.$$

(a) If the standard basis for $\mathbb{R}^3$ is $\mathcal{E} = \{e_1, e_2, e_3\}$, determine

$$T(e_1), T(e_2), \text{ and } T(e_3).$$

Solution. Using linearity, we are given $T(3e_1) = 3T(e_1) = \begin{bmatrix} 6 \\ -3 \\ 6 \end{bmatrix}$, so $T(e_1) = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}$.

We are given $T(e_1 + e_2) = T(e_1) + T(e_2) = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$, so

$$T(e_2) = T(e_1 + e_2) - T(e_1) = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

Finally, $T(2e_3) = \begin{bmatrix} 4 \\ 6 \\ 2 \end{bmatrix}$, so $T(e_3) = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$.

(b) Find $T\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)$.

Solution. We compute

$$T\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) = T(e_1) + T(e_2) + T(e_3) = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}.$$

3. Consider the upper triangular matrix

$$A = \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix},$$

with $x, y, z \in \mathbb{R}$.

(a) Give as many reasons as you can that shows the matrix A is invertible.

Solution. We see that A is already in echelon (not RREF) form, which tells us there is a pivot in each column. Since there are only three variables the system $Ax = 0$ has only the trivial solution, to the linear map $x \mapsto Ax$ is injective. Three pivots also means the column space is spanned by three independent vectors, so is all of $\mathbb{R}^3$. So the linear map is bijective, hence invertible.

One could also say that since the RREF of A is the identity matrix, it is invertible.

If you know about determinants, you could say the determinant equals 1, hence is nonzero, which means A is invertible.

(b) Find the inverse of the matrix A.

Solution. We row-reduce

$$\left[\begin{array}{ccc|ccc} 1 & x & z & 1 & 0 & 0 \\ 0 & 1 & y & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \mapsto \left[\begin{array}{ccc|ccc} 1 & x & 0 & 1 & 0 & -z \\ 0 & 1 & 0 & 0 & 1 & -y \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$
$$\mapsto \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -x & -z+xy \\ 0 & 1 & 0 & 0 & 1 & -y \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right].$$

So

$$A^{-1} = \begin{bmatrix} 1 & -x & -z+xy \\ 0 & 1 & -y \\ 0 & 0 & 1 \end{bmatrix}.$$

4. Consider the linear transformation $T : \mathbb{R}^5 \to \mathbb{R}^4$ given by $T(x) = Ax$ where A and its reduced row-echelon form R are given by:

$$A = \begin{bmatrix} 1 & -1 & 2 & 6 & -3 \\ 2 & -1 & 0 & 7 & 10 \\ -2 & 3 & -7 & -15 & 17 \\ 2 & -2 & 2 & 8 & 5 \end{bmatrix} \text{ and } R = \begin{bmatrix} 1 & 0 & 0 & 5 & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(a) Determine $\ker T$, the kernel of T.

Solution. The kernel of T is the nullspace of A, which we know is the same as the nullspace of R which we can read off:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -5x_4 \\ -3x_4 \\ -2x_4 \\ x_4 \\ 0 \end{bmatrix} = x_4 \begin{bmatrix} -5 \\ -3 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

(b) Determine $\operatorname{Im} T$, the image of T.

Solution. Depending upon what you already know, you could observe that the RREF R has a pivot in each row which means the columns of A span all of $\mathbb{R}^4$.

Or you may know that looking at R tells us there are four pivot columns in A, meaning the column space is spanned by 4 linearly independent vectors, hence the image is all of $\mathbb{R}^4$.

Or, if you have already learned the rank-nullity theorem, then from the previous part we would know the nullity is one, and so rank-nullity says the rank is $5-1=4$, so the image is a dimension 4 subspace of $\mathbb{R}^4$, which is all of $\mathbb{R}^4$.

5. Let K be the set of solutions in $\mathbb{R}^5$ to the homogeneous linear system

$$\begin{aligned} x_1+x_2+x_3+x_4 \qquad &= 0 \\ x_5 &= 0. \end{aligned}$$

(a) Find a basis $\mathcal{B}_0$ for K.

Solution. The coefficient matrix for the system is

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

which is already in reduced row-echelon form. We see there are two pivots, hence 3 free variables, meaning $\dim K = 3$. By inspection (or working out the details of finding all solutions), one finds a basis can be taken to

$$\mathcal{B}_0 = \left\{ v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, v_3 = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

(b) Extend the basis $\mathcal{B}_0$ from the previous part to a basis $\mathcal{B}$ for all of $\mathbb{R}^5$.

Solution. To extend a linearly independent set, one must add something not in the original span (see Theorem 1.1.4). There are many correct answers possible, but the vectors

$$v_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \text{ and } v_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

are clearly not in K since v_4 does not satisfy the first defining equation, and v_5 does not satisfy the second. So thinking algorithmically, $\mathcal{B}_0 \cup$

$\{v_4\}$ is linearly independent, and v_5 is certainly not in the span of those four vectors since their last coordinates are all zero. Thus we may take (as one possible solution)

$$\mathcal{B} = \mathcal{B}_0 \cup \{v_4, v_5\}.$$

(c) Define a linear transformation $T : \mathbb{R}^5 \to \mathbb{R}^5$ with kernel K and image equal to the set of all vectors with $x_3 = x_4 = x_5 = 0$.

Solution. By Theorem 1.1.6, a linear map is uniquely defined by its action on a basis. It should be clear that the desired image is defined by the standard basis vectors e_1 and e_2. So with the given basis $\mathcal{B} = \{v_1, \ldots, v_5\}$, we must have

$$T(v_i) = 0, \text{ for } i = 1, 2, 3,$$

and $T(v_4), T(v_5)$ linearly independent vectors in the image, say

$$T(v_4) = e_1 \text{ and } T(v_5) = e_2.$$

6. Let $M_{2\times 2}$ be the vector space of 2×2 matrices with real entries, and fix a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{2\times 2}$. Consider the linear transformation $T : M_{2\times 2} \to M_{2\times 2}$ defined by $T(X) = AX$, which (left) multiplies an arbitrary 2×2 matrix X by the fixed matrix A. Let $\mathcal{E} = \left\{ \mathbf{e}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathbf{e}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ be a basis for $M_{2\times 2}$.

(a) Find the matrix of T with respect to the basis $\mathcal{E}$, that is $[T]_{\mathcal{E}}$.

Solution.

$$T(\mathbf{e}_1) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a & 0 \\ c & 0 \end{bmatrix} = a\mathbf{e}_1 + c\mathbf{e}_3$$
$$T(\mathbf{e}_2) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & a \\ 0 & c \end{bmatrix} = a\mathbf{e}_2 + c\mathbf{e}_4$$
$$T(\mathbf{e}_3) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} b & 0 \\ d & 0 \end{bmatrix} = b\mathbf{e}_1 + d\mathbf{e}_3$$
$$T(\mathbf{e}_4) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & b \\ 0 & d \end{bmatrix} = b\mathbf{e}_2 + d\mathbf{e}_4$$

We now simply record the data as coordinate vectors:

$$[T]_{\mathcal{E}} = \begin{bmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{bmatrix}$$

(b) Now let $\mathcal{B}$ be the basis, $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2, \mathbf{e}_4\}$, that is, the same elements as $\mathcal{E}$, but with the second and third elements interchanged. Write down the appropriate change of basis matrix, $[I]^{\mathcal{E}}_{\mathcal{B}}$, and use it to compute the matrix of T with respect to the basis $\mathcal{B}$, that is $[T]_{\mathcal{B}}$.

Solution. The change of basis matrices $[I]^{\mathcal{E}}_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} =$ $[I]^{\mathcal{B}}_{\mathcal{E}}$, so

$$\begin{aligned} [T]_{\mathcal{B}} &= [I]^{\mathcal{B}}_{\mathcal{E}}\,[T]_{\mathcal{E}}\,[T]^{\mathcal{E}}_{\mathcal{B}} \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{bmatrix}. \end{aligned}$$

Of course it was possible to write down $[T]_{\mathcal{B}}$ simply from the information in part (a).

7. Write down an explicit linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ that has as its image the plane $x - 4y + 5z = 0$. What is the kernel of T?

Hint. Any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ has the form $T(x) = Ax$ where A is the matrix for T with respect to the standard bases. How is the image of T related to the matrix A?

Solution. We know that T can be given by $T(x) = Ax$ where A is the 3×2 matrix whose columns are $T(e_1)$ and $T(e_2)$. They must span the given plane, so for example, $A = \begin{bmatrix} 4 & -5 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ will do.

By rank-nullity, the kernel must be trivial.

8. Let $A \in M_n(\mathbb{R})$ which is invertible. Show that the columns of A form a basis for $\mathbb{R}^n$.

Solution. Since A is invertible, we know that we can find its inverse by row reducing the augmented matrix

$$[A|I_n] \mapsto [I_n|A^{-1}].$$

In particular, this says that the RREF form of A is I_n.

One way to finish is that the information above says that $Ax = 0$ has only the trivial solution, which means that the n columns of A are linearly independent. Since there are $n = \dim \mathbb{R}^n$ of them, by Theorem 1.1.3, they must be a basis.

Another approach is that the linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = Ax$ is an isomorphism with the inverse map being given $x \mapsto A^{-1}x$. In particular, T is surjective and its image is the column space of A. That means that the n columns of A span all of $\mathbb{R}^n$, and hence must be a basis again by Theorem 1.1.3.

9. Consider the vector space $M_2(\mathbb{R})$ of all 2×2 matrices with real entries. Let's consider a number of subspaces and their bases. Let $\mathcal{E} = \{E_{11}, E_{12}, E_{21}, E_{22}\} = \{\left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]\}$ be the standard basis for $M_2(\mathbb{R})$.

(a) Define a map $T : M_2(\mathbb{R}) \to \mathbb{R}$ by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a + d.$$

The quantity $a + d$ (the sum of the diagonal entries) is called the **trace** of the matrix. You may assume that T is a linear map. Find a basis for its kernel, K.

Solution. It is easy to see that T is a surjective map, so by the rank-nullity theorem, $\dim K = 3$. Extracting from the standard basis, we see that $E_{12}, E_{21} \in K$ so are part of a basis for K. We just need to add one more matrix which is not in the span of the two chosen basis vectors.

Certainly, the matrix must have the form $\begin{bmatrix} a & b \\ c & -a \end{bmatrix}$, and we need $a \neq 0$, otherwise our matrix is in the span of the other two vectors. But once we realize that, we may as well assume that $b = c = 0$, so that $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ is a nice choice, and since it is not in the span of the other two, adding it still gives us an independent set.

(b) Now let's consider the subspace S consisting of all symmetric matrices, those for which $A^T = A$. It should be clear this is a proper subspace, but what is its dimension. Actually finding a basis helps answer that question.

Hint. If you don't like the "brute force" force of the tack of the solution, you could take the high road and consider the space of **skew-symmetric matrices,** those for which $A^T = -A$. It is pretty easy to determine its dimension and then you can use the fact that every matrix can be written as the sum of symmetric and skew-symmetric matrix to tell you the dimension of S.

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

Solution. Once again, it is clear that some elements of the standard basis are in S, like E_{11}, E_{22}. Since it is a proper subspace, its dimension is either 2 or 3, and a few moments thought convinces you that

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = E_{12} + E_{21}$$

is symmetric, not in the span of the other two, so forms an independent set in S. So $\dim S = 3$, this must be a basis for S.

(c) Now $K \cap S$ is also a subspace of $M_2(\mathbb{R})$. Can we find its dimension.

Solution. Once again, it is useful to know the dimension of the space. Certainly it is at most 3, but then not every symmetric matrix has zero trace, so it is at most two. Staring at the bases for each of S and K separately, we see that both

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

are in the intersection and are clearly linearly independent, so they must be a basis.

(d) Extend the basis you found for $K \cap S$ to bases for S and for K.

Solution. Since $\dim(K \cap S) = 2$, we need only find one matrix not in their span to give a basis for either K or S. For K, we could choose E_{12}, and for S we could choose E_{11}. Knowing the dimension is clearly a powerful tool since it tells you when you are done.

10. The matrix $B = \begin{bmatrix} 1 & 4 & -7 \\ -3 & -11 & 19 \\ -1 & -9 & 18 \end{bmatrix}$ is invertible with inverse $B^{-1} = \begin{bmatrix} -27 & -9 & -1 \\ 35 & 11 & 2 \\ 16 & 5 & 1 \end{bmatrix}$. Since the columns of B are linearly independent, they form a basis for $\mathbb{R}^3$:

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 \\ -11 \\ -9 \end{bmatrix}, \begin{bmatrix} -7 \\ 19 \\ 18 \end{bmatrix} \right\}.$$

Let $\mathcal{E}$ be the standard basis for $\mathbb{R}^3$.

(a) Suppose that a vector $v \in \mathbb{R}^3$ has coordinate vector $[v]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Find$[v]_{\mathcal{E}}$.

Solution. The matrix B is the change of basis matrix $[I]_{\mathcal{B}}^{\mathcal{E}}$ so

$$[v]_{\mathcal{E}} = [I]_{\mathcal{B}}^{\mathcal{E}}[v]_{\mathcal{B}} = \begin{bmatrix} 1 & 4 & -7 \\ -3 & -11 & 19 \\ -1 & -9 & 18 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -12 \\ 32 \\ 35 \end{bmatrix}$$

(b) Suppose that $T : \mathbb{R}^3 \to \mathbb{R}^3$ is the linear map given by $T(x) = Ax$ where

$$A = [T]_{\mathcal{E}} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Write down an appropriate product of matrices which equal $[T]_{\mathcal{B}}$.

Solution. By Theorem 1.4.9

$$[T]_{\mathcal{B}} = [I]_{\mathcal{E}}^{\mathcal{B}}[T]_{\mathcal{E}}[I]_{\mathcal{B}}^{\mathcal{E}} = B^{-1}AB.$$

11. Let W be the subspace of $M_2(\mathbb{R})$ spanned by the set S, where

$$S = \left\{ \begin{bmatrix} 0 & -1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 9 \end{bmatrix}, \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} \right\}.$$

(a) Use the standard basis $\mathcal{B} = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ for $M_2(\mathbb{R})$ to express each element of S as a coordinate vector with respect to the basis $\mathcal{B}$.

Solution. We write the coordinate vectors as columns of the matrix:

$$\begin{bmatrix} 0 & 1 & 2 & 1 \\ -1 & 2 & 1 & -2 \\ -1 & 2 & 1 & -2 \\ 1 & 3 & 9 & 4 \end{bmatrix}.$$

(b) Determine a basis for W.

Hint. By staring at the matrix, it is immediate that that rank is at most 3. What are the pivots?

Solution. We start a row reduction:

$$A \mapsto \begin{bmatrix} 0 & 1 & 2 & 1 \\ -1 & 2 & 1 & -2 \\ 1 & 3 & 9 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 3 & 9 & 4 \\ 0 & 1 & 2 & 1 \\ -1 & 2 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mapsto \begin{bmatrix} 1 & 3 & 9 & 4 \\ 0 & 1 & 2 & 1 \\ 0 & 5 & 10 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 3 & 9 & 4 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus the pivot columns are the first, second, and fourth, so we may take the first, second and fourth elements of S as a basis for W.

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

(a) Compute the rank and nullity of A.

Solution. Too easy! It is obvious that the rank is 1 since all columns are multiples of the first. Rank-nullity tells us that the nullity is $3 - 1 = 2$.

(b) Compute $A \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, and use your answer to help conclude (without computing the characteristic polynomial) that A is diagonalizable.

Solution. $A \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \\ 6 \end{bmatrix} = 6 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, which means that 6 is a eigenvalue for A, and $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is an eigenvector.

The nullity is 2, which means that 0 is an eigenvalue and that the eigenspace corresponding to 0 (the nullspace of A) has dimension 2, so that there exists a basis of $\mathbb{R}^3$ consisting of eigenvectors. Recall that by Proposition 1.5.5 the eigenvectors from different eigenspaces are linearly independent.

(c) Determine the characteristic polynomial of A from what you have observed.

Solution. $\chi_A(x) = x^2(x-6)$. There are two eigenvalues, 0 and 6, and since the matrix is diagonalizable the algebraic multiplicities to which they occur equal their geometric multiplicities (i.e., the dimension of the corresponding eigenspaces), see Theorem 1.5.6.

(d) Determine a matrix P so that

$$\begin{bmatrix} 6 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = P^{-1}AP.$$

Solution. We already know that $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is an eigenvector for the eigenvalue 6, and since 6 occurs as the first entry in the diagonal matrix, that should be the first column of P.

To find a basis of eigenvectors for the eigenvalue 0, we need to find the nullspace of A. It is immediate to see that the reduced row-echelon form of A is

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

which tells us the solutions are

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_2 - 3x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}.$$

We may choose either of those vectors (or some linear combinations of them) to fill out the last columns of P. So one choice for P is

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

13. Let $\mathcal{E}_1 = \{E_{11}, E_{12}, E_{21}, E_{22}\} = \{\left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]\}$ be the standard basis for $M_2(\mathbb{R})$, and $\mathcal{E}_2 = \{1, x, x^2, x^3\}$ the standard basis for $\mathcal{P}_3(\mathbb{R})$. Let $T : M_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ be defined by

$$T(\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]) = 2a + (b-d)x - (a+c)x^2 + (a+b-c-d)x^3.$$

(a) Find the matrix of T with respect to the two bases: $[T]_{\mathcal{E}_1}^{\mathcal{E}_2}$.

Solution. The columns of the matrix $[T]_{\mathcal{E}_1}^{\mathcal{E}_2}$ are the coordinate vectors $[T(E_{ij})]_{\mathcal{E}_2}$, so

$$[T]_{\mathcal{E}_1}^{\mathcal{E}_2} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & -1 & 0 \\ 1 & 1 & -1 & -1 \end{bmatrix}.$$

(b) Determine the rank and nullity of T.

Solution. It is almost immediate that the first three columns of the matrix are pivot columns (think RREF), so the rank is at least three. Then we notice that the last column is a multiple of the second, which means the rank is at most three. Thus rank is 3 and nullity is 1.

(c) Find a basis of the image of T.

Solution. The first three columns of $[T]_{\mathcal{E}_1}^{\mathcal{E}_2}$ are a basis for the column space of the matrix, but we recall that they are coordinate vectors and the codomain is $P_3(\mathbb{R})$, so a basis for the image is:

$$\{2 - x^2 + x^3, x + x^3, -x^2 - x^3\}.$$

(d) Find a basis of the kernel of T.

Solution. Since

$$T(\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]) = 2a + (b-d)x - (a+c)x^2 + (a+b-c-d)x^3,$$

we must characterize all matrices which yield the zero polynomial. We quickly deduce we must have

$$a = c = 0, \text{ and } b = d,$$

so one can choose $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ as a basis for the kernel.

14. Let V be a vector space with basis $\mathcal{B} = \{v_1, \dots, v_4\}$. Define a linear transformation by

$$T(v_1) = v_2, \quad T(v_2) = v_3, \quad T(v_3) = v_4, \quad T(v_4) = av_1 + bv_2 + cv_3 + dv_4.$$

(a) What is the matrix of T with respect to the basis $\mathcal{B}$?

Solution. $[T]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 & a \\ 1 & 0 & 0 & b \\ 0 & 1 & 0 & c \\ 0 & 0 & 1 & d \end{bmatrix}.$

(b) Determine necessary and sufficient conditions on a, b, c, d so that T is invertible.

Hint. What is the determinant of T, or what happens when you row reduce the matrix?

Solution. The determinant of the matrix is $-a$, so T is invertible if and only if $a \neq 0$. The values of b, c, d do not matter.

(c) What is the rank of T and how does the answer depend upon the values of a, b, c, d?

Solution. With one elementary row operation, we reduce the original matrix to $\begin{bmatrix} 1 & 0 & 0 & b \\ 0 & 1 & 0 & c \\ 0 & 0 & 1 & d \\ 0 & 0 & 0 & a \end{bmatrix}$ which is in echelon form. If $a = 0$, the rank is 3, otherwise it is 4.

15. Define a map $T : M_{m\times n}(\mathbb{R}) \to \mathbb{R}^m$ as follows: For $A = [a_{ij}] \in M_{m\times n}(\mathbb{R})$, define $T(A) = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ where $b_k = \sum_{j=1}^{n} a_{kj}$, that is, b_k is the sum of all the elements in the k-th row of A. Assume that T is linear.

(a) Find the rank and nullity of T.

Hint. If you find this too abstract, try an example first, say with $m = 2$ and $n = 3$. And finding the rank is the easier first step.

Solution. Using the standard basis $\{E_{ij}\}$ for $M_{m\times n}(\mathbb{R})$, we see that $T(E_{k1}) = e_k$ where $\{e_1, \ldots, e_m\}$ is the standard basis for $\mathbb{R}^m$. Since a spanning set for $\mathbb{R}^m$ is in the image of T, the map must be surjective, which means the rank is m. By rank-nullity, the nullity is $nm - m$.

(b) For $m = 2$, and $n = 3$ find a basis for the nullspace of T.

Hint. For an element to be in the nullspace, the sum of the entries in each of its rows needs to be zero. Can you make a basis with one row in each matrix all zero?

Solution. Consider the set

$$\left\{\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}\right\}$$

Notice that the 1 which occurs in each matrix occurs in a different location in each matrix. It is now easy to show that any linear combination of these matrices which equals the zero matrix must have all coefficients equal to zero, so the set is linearly independent. Since it has the correct size, it must be a basis for the nullspace.

16. This exercise is about how to deal with determining independent and spanning sets in vector spaces other than F^n. Let $V = P_3(\mathbb{R})$, the vector space of polynomials of degree at most 3 with real coefficients. Suppose that some process has handed you the set of polynomials

$$S = \{p_1 = 1+2x+3x^2+3x^3, p_2 = 5+6x+7x^2+8x^3, p_3 = 9+10x+11x^2+12x^3, p_4 = 13+14x+15x^2-$$

We want to know whether S is a basis for V, or barring that extract a maximal linearly independent subset.

(a) How can we translate this problem about polynomials into one about vectors in $\mathbb{R}^n$?

Solution. We know that Theorem 1.2.5 tells us that $P_3(\mathbb{R})$ is isomorphic to $\mathbb{R}^4$, and all we need to do is map a basis to a basis, but we would like a little more information at our disposal.

Let $\mathcal{B} = \{1, x, x^2, x^3\}$ be the standard basis for $V = P_3(\mathbb{R})$. Then the map

$$T(v) = [v]_{\mathcal{B}}$$

which takes a vector v to its coordinate vector is such an isomorphism. What is important is that linear dependence relations among the vectors in S are automatically reflected in linear dependence relations among the coordinate vectors.

(b) Determine a maximal linearly independent subset of S.

Solution. If we record the coordinate vectors for the polynomials in S as columns of a matrix, we produce a matrix A and its RREF R:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix} \mapsto R = \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So we see that the first two columns are pivot columns which means $S_0 = \{p_1, p_2\}$ is a maximal linearly independent set.

We also recall that from the RREF, we can read off the linear dependencies with the other two vecotrs:

$$p_3 = -p_1 + 2p_2 \text{ and } p_4 = -2p_1 + 3p_2.$$

(c) Extend the linearly independent set from the previous part to a basis for $P_3(\mathbb{R})$.

Solution. Since we are free to add whatever vectors we want to the given set, we can add column vectors to the ones for p_1 and p_2 to see if we can extend the basis. We know that $\{p_1, p_2, 1, x, x^2, x^3\}$ is a linearly dependent spanning set. We convert to coordinates and row reduce to find the pivots. So we build a matrix B and its RREF:

$$\begin{bmatrix} 1 & 5 & 1 & 0 & 0 & 0 \\ 2 & 6 & 0 & 1 & 0 & 0 \\ 3 & 7 & 0 & 0 & 1 & 0 \\ 4 & 8 & 0 & 0 & 0 & 1 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 0 & 0 & -2 & \frac{7}{4} \\ 0 & 1 & 0 & 0 & 1 & -\frac{3}{4} \\ 0 & 0 & 1 & 0 & -3 & 2 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

We see the first 4 columns are pivots, so we may take $\{p_1, p_2, 1, x\}$ as one such basis.

17. Let $A \in M_5(\mathbb{R})$ be the block matrix (with off diagonal blocks all zero) given by:

$$A = \begin{bmatrix} -1 & 0 & & & \\ \alpha & 2 & & & \\ & & 3 & 0 & 0 \\ & & \beta & 3 & 0 \\ & & 0 & \gamma & 3 \end{bmatrix}.$$

Determine all values of α, β, γ for which A is diagonalizable.

Solution. Since the matrix is lower triangular, it is easy to compute the characteristic polynomial:

$$\chi_A = (x + 1)(x - 2)(x - 3)^3.$$

The eigenspaces for $\lambda = -1, 2$ each have dimension 1 (the required minimum) and equal to the algebraic multiplicity, so the only question is what happens with the eigenvalue $\lambda = 3$. Consider the matrix $A - 3I = \begin{bmatrix} -4 & 0 & & & \\ \alpha & -1 & & & \\ & & 0 & 0 & 0 \\ & & \beta & 0 & 0 \\ & & 0 & \gamma & 0 \end{bmatrix}$. For the nullspace of $A - 3I$ to have dimension 3, the rank must be 2. Clearly the first two rows are linearly independent (independent of α), while if either β or γ is nonzero, this will increase the rank beyond two. So the answer is α can be anything, but β and γ must both be zero.

18. Let $A = \begin{bmatrix} 3 & 0 & 0 \\ 6 & -1 & 6 \\ 1 & 0 & 2 \end{bmatrix} \in M_3(\mathbb{R})$.

(a) Find the characteristic polynomial of A.

Solution. $\chi_A = \det(xI - A) = \det\left(\begin{bmatrix} x-3 & 0 & 0 \\ -6 & x+1 & -6 \\ -1 & 0 & x-2 \end{bmatrix}\right)$. Expanding along the first row shows that $\chi_A = (x-3)(x-2)(x+1)$.

(b) Show that A is invertible.

Solution. Many answers are possible: $\det A = -6 \neq 0$, or 0 is not an eigenvalue, or one could row reduce the matrix to the identity. All show A is invertible.

(c) Justify that the columns of A form a basis for $\mathbb{R}^3$.

Solution. Since A is invertible, the rank of A is 3, which is the dimension of the column space. So the column space spans all of $\mathbb{R}^3$, which means the columns must be linearly independent either by Theorem 1.1.3 or directly since the nullspace is trivial. Thus the columns form a basis.

(d) Let $\mathcal{B} = \{v_1, v_2, v_3\}$ be the columns of A, and let $\mathcal{E}$ be the standard basis for $\mathbb{R}^3$. Suppose that $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear map for which $A = [T]_{\mathcal{E}}$. Determine $[T]_{\mathcal{B}}$.

Solution. We know that $[T]_{\mathcal{B}} = Q^{-1}[T]_{\mathcal{E}}Q$, where $Q = [I]_{\mathcal{B}}^{\mathcal{E}}$ is a change of basis matrix. But we see that $Q = [I]_{\mathcal{B}}^{\mathcal{E}} = A$ by definition and since $[T]_{\mathcal{E}} = A$ as well, we check that $[T]_{\mathcal{B}} = Q^{-1}[T]_{\mathcal{E}}Q = A^{-1}AA = A$.

Chapter 2

Vector space constructions

This chapter contains material the reader may or may not have seen. The most critical is the first section on sums and direct sums. The later two sections are more important to advanced readers.

One goal of linear algebra is to understand the properties of a linear map $T : V \to W$ between two vector spaces. There are many ways in which to do this depending on the end goal.

One approach is try to separate the original problem into smaller subproblems. For example, it is often the case that restricted to a subspace $U \subseteq V$, the behavior of T is well understood. An easy instance of this is that we know exactly what T does when restricted to $U = \ker(T)$. If we could somehow write V as a sum of spaces, $\ker(T)+U'$, the job of understanding the action of T would be reduced to understanding T on the smaller subspace U'. We shall make this precise below.

In a very different direction than direct sums, let's try to understand the image of a linear map $T : V \to W$. Of course if T is injective, then there is an isomorphic copy of V sitting inside W. But if T is not injective, how do we describe the image of the map in terms of V?

Suppose w is in the image of T, say with $T(v_0) = w$. We characterized the inverve image, $T^{-1}(w)$ in Equation (1.2.1) as

$$T^{-1}(w) = \{v_0 + k \mid k \in \ker(T)\} = v_0 + \ker(T).$$

So all those vectors get mapped to the same point $w \in W$. This suggests setting up an equivalence relation among the vectors in V, so that each equivalence class is in bijective correspondence with the points in the image of T. We shall make this clearer in the sections below.

2.1 Sums and Direct Sums

Let's return to the example above in which $T : V \to W$ is a linear map, and suppose $\{v_1, \ldots, v_k\}$ is a basis for $U = \ker(T)$, and we extend that basis to a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ for V. If we put $U' = \mathrm{Span}(\{v_{k+1}, \ldots, v_n\})$, then every element $v \in V$ can be written as $v = u + u'$ for unique vectors $u \in U$ and $u' \in U'$.

Taking this one step further, if $v = u + u'$ as above we know that

$$T(v) = T(u + u') = T(u) + T(u') = 0 + T(u') = T(u'),$$

so that understanding the action of T on V has been reduced to understanding the action on the subspace U'. So effectively we have reduced the size of our problem.

The situation we described above is actually rather special, so let's begin with a slightly more general notion.

Let U, W be subspaces of a vector space V. Denote by

$$U + W := \{u + w \mid u \in U,\ w \in W\}.$$

That is, $U + W$ is the set of vectors $v \in V$ which can be written as $v = u + w$ for some $u \in U$ and some $w \in W$. That seems very similar to what happened in the example above, except in that example, the vectors u, w were uniquely determined.

It is easy to check that $U + W$ is a subspace of V, (indeed the *smallest* subspace of V containing U and W), but before going too far, we should make a few simple observations. First, it is immediate to check that $U + W = W + U$ since addition in a vector space is commutative. What if we have more than two subspaces?

If we had three subspaces U_i, $i = 1, 2, 3$, we could easily check (since we know how to add pairs of subspaces) that

$$(U_1 + U_2) + U_3 = U_1 + (U_2 + U_3),$$

so we can unambiguously define

$$U_1 + U_2 + U_3 := (U_1 + U_2) + U_3,$$

and inductively we define

$$U_1 + \cdots + U_n := (U_1 + \cdots + U_{n-1}) + U_n.$$

But as with any new concept, some examples help us better understand it.

Example 2.1.1 A standard decomposition of F^n. Let $\{e_1, \ldots, e_n\}$ be the standard basis for F^n, and put $U_i = \mathrm{Span}\{e_i\}$, the line through the origin in the direction of e_i. So when $n = 3$, these subspaces are just the x, y, and z axes. Then we see that $F^n = U_1 + U_2 + \cdots + U_n$. We also see that every element of F^n is the sum of uniquely determined elements from the U_i. As row vectors,

$$(a_1, \ldots, a_n) = (a_1, 0, \ldots, 0) + (0, a_2, 0, \ldots, 0) + \cdots + (0, \ldots, 0, a_n).$$

□

Example 2.1.2 Decomposing $V = F^3$. Let $\{e_1, e_2, e_3\}$ be the standard basis for V, and let $U = \text{Span}\{e_1, e_2\}$ and let $W = \text{Span}\{e_3, e_1 + e_2 + e_3\}$. It is straightforward to show that

$$U + W = \text{Span}\{e_1, e_2, e_3, e_1 + e_2 + e_3\} = \text{Span}\{e_1, e_2, e_3\} = V,$$

so every element of V can be written as the sum of vectors from U and W, but in this case not necessarily uniquely.

As a trivial example, let $v = e_3$. Then v can be written as $v = u + w$ with $u = 0$ and $w = e_3$, or with $u = -e_1 - e_2$ and $w = e_1 + e_2 + e_3$. □

The source of this non-uniqueness is actually easy to discover. Suppose that $V = U + W$, and for some $v \in V$,

$$v = u_1 + w_1 = u_2 + w_2.$$

Then of course $r = u_1 - u_2 = w_2 - w_1$. For uniqueness, we would need $u_1 = u_2$ and $w_1 = w_2$. Said another way, we would need $r = u_1 - u_2 = w_2 - w_1 = 0$. But $u_1 - u_2 \in U$ and $w_2 - w_1 \in W$, so the only way to force uniqueness is if $U \cap W = \{0\}$.

We summarize this as

Proposition 2.1.3 *Let U, W be subspaces of a vector space V, and suppose that $V = U + W$. Then every element of V is representable as a sum of uniquely determined elements of U and W if and only if $U \cap W = \{0\}$.*

In the case that $V = U + W$, and $U \cap W = \{0\}$, we write

$$V = U \oplus W$$

and call V the **direct sum** of the subspaces U and W.

Checkpoint 2.1.4 Suppose that U_i, $i = 1, 2, 3$ are subspaces of a vector space V, and that $V = U_1 + U_2 + U_3$. We want necessary and sufficient conditions so that every element of V can be represented as a unique sum of elements from the U_i. What about when $V = U_1 + \cdots + U_n$ for $n \geq 3$?

Hint. To gain some insight, first find an example in $\mathbb{R}^3$ where $U_i \cap U_j = \{0\}$ whenever $i \neq j$, but not every element of $\mathbb{R}^3$ has a unique representation as a sum.

2.2 Quotient Spaces

We need a very different construction than what we used for direct sums; we now want to build an entirely new space from a given vector space V and an arbitrry subspace U. Let's motivate this construction with a familiar example. Given a linear map $T : V \to W$ with $T(v_0) = w$, we characterized the inverve image, $T^{-1}(w)$ in Equation (1.2.1) as

$$T^{-1}(w) = \{v_0 + k \mid k \in \ker(T)\} = v_0 + \ker(T).$$

Let's put an equivalence relation on V by saying that $v_1 \sim v_2$ if and only if $v_1 - v_2 \in \ker(T)$. It is easy to check that this can be rephrased as $v_1 \sim v_2$ if and only if $T(v_1) = T(v_2)$. We let $V/\ker(T)$ denote the set of equivalence classes. So if we let $[v]$ denote the equivalence class containing v, then

$$V/\ker(T) = \{[v] \mid v \in V\} \text{ where } [v] = \{v + w \mid w \in \ker(T)\}.$$

As a shorthand, we write

$$v + \ker(T) := \{v + w \mid w \in \ker(T)\}.$$

Now there was nothing particularly special about using $\ker(T)$ for the construction. So if V is a vector space, and U is any subspace, we define an equivalence relation on V by $v_1 \sim_U v_2$ iff $v_1 - v_2 \in U$. As in the previous example the equivalence classes have the form

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

Finally, we denote by

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

Definition 2.2.1 Let V be a vector space, and U any subspace. The set V/U is called a **quotient space**, and V/U is read V **mod** U or simply the **quotient of** V **by** U. The elements $v + U$ are called **cosets** of U in V. $\Diamond$

As with any equivalence relation, the equivalence classes partition the original set, so that V is the disjoint union of the cosets:

$$V = \bigsqcup_{v \in V} (v + U).$$

Example 2.2.2 Let $V = \mathbb{R}^2$ and $U = \mathrm{Span}\{(a, b)\}$ where $(a, b) \neq (0, 0)$, thus the subspace U is simply a line through the origin. Now let $v = (c, d)$. Then the coset $v + U$ is simply the line through (c, d) in the direction of (a, b). We also see that the union of these lines is all of $V = \mathbb{R}^2$. $\square$

Now that we understand V/U as a set, we want to introduce an algebraic structure on it inherited naturally from V. We define addition of cosets in a natural manner:

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

It follows that the additive identity is $0 + U (= U)$, the additive inverse of $v + U$ is $-v + U$, and scalar multiplication is defined by $\lambda(v + U) = \lambda v + U$. One checks the operations are well-defined and makes V/U a vector space over the same scalar field as V.

One checks that $V/\{0\} \cong V$, and $V/V \cong \{0\}$, but how to think about V/U in general is the subject of the next section.

2.3 Linear Maps out of quotients

Suppose we have vector spaces V, W and a subspace $U \subseteq V$. How should one define a linear map $T : V/U \to W$? In general one should not (as least directly)! We want to show that every well-defined linear map $T : V/U \to W$ arises in a natural way from a linear map $T_0 : V \to W$.

Why such a fuss? How hard is it to define a map on cosets? While it isn't that hard, each such definition requires an extra step. To define $T(v + U)$ one must show that the definition is well-defined, meaning if $v + U = v' + U$, then $T(v + U) = T(v' + U)$. The method we shall propose will do this once and the result will apply to all maps.

To begin, we first note that there is a natural linear map (called a **projection**) $\pi : V \to V/U$ defined by

$$\pi(v) = v + U.$$

It is easy to check that this is a surjective linear map with $\ker(\pi) = U$.

As an immediate corollary of properties of the projection map, we deduce:

Corollary 2.3.1 *Let V be a finite-dimensional vector space over a field F, and let U be a subspace of V. Then*

$$\dim(V/U) = \dim V - \dim U.$$

Proof. Consider the projection map $\pi : V \to V/U$ (given by $\pi(v) = v + U$). We have already stated that π is a surjective linear map with $\ker \pi = U$. So by the rank-nullity theorem we have that

$$\dim V = \dim(V/U) + \dim U,$$

from which the result follows. ■

So now we suppose that we have a linear map $T : V \to W$, and a subspace $U \subseteq V$. We want to know when we can induce a linear map $T_* : V/U \to W$ which makes the diagram below **commute.** What that means is that starting with a vector $v \in V$, following either path to W produces the same result. In terms of the functions, this means that

$$T(v) = (T_* \circ \pi)(v) = T_*(v + U).$$

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\pi"] & W \\ V/U \arrow[ur, dashed, "T_*"'] & \end{tikzcd}$$

Figure 2.3.2 Factoring a map through a quotient

It is evident that if such a linear map exists, it can have only one definition:

$$T_*(v+U) = T(v),$$

and it is here we confront and deal with the issue of T_* being well-defined. If $v+U = v'+U$, we need that $T(v) = T(v')$.

By definition, the condition $v+U = v'+U$ is equivalent to $v - v' \in U$, say $v = v' + u$ for some $u \in U$. The requirement that $T(v) = T(v')$ demands that

$$T(v) = T(v'+u) = T(v') + T(u) = T(v'),$$

so we must have $T(u) = 0$. Thus a necessary and sufficient condition that the map be well-defined is that $U \subseteq \ker(T)$.

Theorem 2.3.3 Fundamental theorem on linear maps. *Let $T : V \to W$ be a linear map and U a subspace of V with $U \subseteq \ker(T)$. Then there is a unique linear map $T_* : V/U \to W$ (defined by $T_*(v+U) = T(v)$) with $\mathrm{Im}(T_*) = \mathrm{Im}(T)$, and with $\ker(T_*) = \ker(T)/U$.*

Corollary 2.3.4 *If $T : V \to W$ is a linear map, then the induced map $T_* : V/\ker T \to W$ is injective.*

What this says is that if the original map T is not injective (many things in V mapping to the same element in W), the coset $v + \ker T$ collects together all the elements of V which map to the element $T(v)$ in W. In this way we obtain an isomorphic copy of $V/\ker T$ inside of W.

Corollary 2.3.5 First Isomorphism theorem. *Let $T : V \to W$ be a linear map. Then $V/\ker(T) \cong \mathrm{Im}(T)$ via the map $T_*(v+\ker(T)) = T(v)$. In particular, if T is a surjective map then $V/\ker(T) \cong W$.*

In the interest of full disclosure, the first isomorphism theorem (and its corollaries) are not that robust in linear algebra since vector spaces are classified up to isomophism by their dimension. They become much more important in group and ring theory, but let's try to give a sense of what they accomplish.

Example 2.3.6 Let's take a simple example from multivariable calculus. Let a, b, c be real numbers, not all zero. Consider the linear map $T : \mathbb{R}^3 \to \mathbb{R}$ given by $T(x, y, z) = ax + by + cz$. This is certainly linear as it is of the form $\mathbf{x} \mapsto A\mathbf{x}$ where A is the 1×3 matrix $[a\ b\ c]$.

Let $P = \ker T$, that is P is the plane through the origin $ax + by + cz = 0$. How do we characterize the quotient space $\mathbb{R}^3/P$?

It is trivial to check that the map T is surjective, so the first isomorphism theorem says that $\mathbb{R}^3/P \cong \mathbb{R}$.

Alternatively, we know that $\dim P = 2$, so $\dim(\mathbb{R}^3/P) = 3 - 2 = 1$, and any two spaces of the same (finite) dimension are isomorphic, though the orignial map T is more intrinsic to the problem. □

Chapter 3

Inner Product Spaces

This chapter contains the material that every linear algebra course wants to cover, but which often gets short shrift as time runs short and students strain to keep all the new concepts straight. So a point is made to take time with this material.

It is in this chapter that we find some of the most important applications of linear algebra as well as some of the deepest results, many of which have vast generalizations in the realm of functional analysis.

Starting from basic definitions and properties, we move to the fundamental notion of orthogonality and orthogonal projection. While grounded with geometric intuition, this notion has profound applications to high-dimensional spaces where our geometric intuition fails. Applications include least squares solutions to inconsistent linear systems as well as spectral decompositions for real symmetric and unitary/normal complex matrices. We discuss results over the complex numbers, and note where differences arise with the results over the reals. We state without proof the spectral theorems and leverage them to develop the singular value decomposition of a matrix. We give an application an application to image compression and explore some of the underlying duality.

3.1 Inner Product Spaces

While a great deal of linear algebra applies to all vector spaces, by restricting attention to those with some notion of distance and orthogonality, we can go much further.

3.1.1 Definitions and examples

Our discussion of inner product spaces will generally ***restrict to the setting of a vector space over a field F being either the real or complex numbers.***

Recall the axioms of an inner product. They are often paraphrased with higher level concepts. For example, the first two axioms combined says that the inner product is **linear in the first variable** (with the second variable held constant). What that means is that if we fix a vector $w \in V$ and define $T : V \to V$ by $T(v) = \langle v, w \rangle$, then T is a linear operator on V.

Remark 3.1.1 We note that the third axiom tells us that the inner product is **conjugate linear** in the second variable (or that the function of two variables,$\langle \cdot, \cdot \rangle$, is **sesquilinear).** Using the first three axioms, if we fix $v \in V$, and define $S : V \to V$ by $S(w) := \langle v, w \rangle$, we observe

$$\begin{aligned} S(u+w) = \langle v, u+w \rangle &= \overline{\langle u+w, v \rangle} = \overline{\langle u, v \rangle + \langle w, v \rangle} \\ &= \overline{\langle u, v \rangle} + \overline{\langle w, v \rangle} = \langle v, u \rangle + \langle v, w \rangle = S(u) + S(v), \end{aligned}$$

and

$$S(\lambda u) = \langle v, \lambda u \rangle = \overline{\langle \lambda u, v \rangle} = \overline{\lambda \langle u, v \rangle} = \overline{\lambda} \langle v, u \rangle = \overline{\lambda} S(u),$$

hence the term **conjugate linear.**

Remark 3.1.2 We also note that if we are dealing with a real inner product space (i.e., $F = \mathbb{R}$), then the inner product is linear in both variables leading mathematicians to call it **bilinear,** that is linear in each variable while holding the other fixed.

Remark 3.1.3 An inner product on a vector space V will give us a notion of when two vectors are **orthogonal.** The positivity condition on an inner product ($\langle v, v \rangle > 0$ unless $v = 0$) gives us a notion of length. We define the **norm** of a vector $v \in V$ by

$$\|v\| := \sqrt{\langle v, v \rangle}.$$

First we assemble a collection of inner products, and their norms.

Example 3.1.4 $V = F^n$**.** Let $v = (a_, \ldots, a_n), w = (b_1, \ldots, b_n) \in F^n$ (written as row vectors). Define

$$\langle v, w \rangle := \sum_{i=1}^{n} a_i \overline{b_i}.$$

This inner product is called the **standard inner product** on F^n. When $F = \mathbb{R}$, this is the usual **dot product**.

If $v = (a_, \ldots, a_n)$, we see that

$$\|v\| = \langle v, v \rangle = \sqrt{\sum_{i=1}^{n} a_i \overline{a}_i} = \sqrt{\sum_{i=1}^{n} |a_i|^2}$$

$\square$

Example 3.1.5 $V = M_{m\times n}(\mathbb{C})$**.** Let $A, B \in V = M_{m\times n}(\mathbb{C})$. Define the **Frobenius inner product** of A and B by

$$\langle A, B\rangle := \operatorname{tr}(AB^*) = \operatorname{tr}(B^*A),$$

where B^* is the conjugate transpose of B, and tr is the trace of the matrix.

Here the norm is $\|A\| = \sqrt{\operatorname{tr}(A^*A)}$. □

Example 3.1.6 $V = C([0,1])$**.** Let $V = C([0,1])$ be the set of real-valued continuous functions defined on the interval $[0,1]$. For $f, g \in C([0,1])$, define their inner product on V by:

$$\langle f, g\rangle := \int_0^1 f(t)g(t)\,dt.$$

If instead f and g are complex-valued, then the inner product becomes:

$$\langle f, g\rangle := \int_0^1 f(t)\overline{g(t)}\,dt.$$

Here the norm is $\|f\| := \sqrt{\int_0^1 f(t)\overline{f(t)}} = \sqrt{\int_0^1 |f(t)|^2}$, where $|\cdot|$ is the usual absolute value on the complex numbers. □

If $(V, \langle\cdot,\cdot\rangle)$ is an inner product space, we say that

- $u, v \in V$ are **orthogonal** if $\langle u, v\rangle = 0$.
- Two subsets $S, T \subseteq V$ are **orthogonal** if $\langle u, v\rangle = 0$ for every $u \in S$ and $v \in T$.
- $v \in V$ is a **unit vector** if $\|v\| = 1$.

3.1.2 Basic Properties

We list some basic properties of inner products and their norms which can be found in any of the standard references.

Let V be an inner product space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$.

Theorem 3.1.7 *For all $u, v, w \in V$ and $\lambda \in F$*

- *If $\langle v, u\rangle = \langle v, w\rangle$ for all $v \in V$, then $u = w$.*
- $\|\lambda v\| = |\lambda|\,\|v\|$.
- *$\|v\| \geq 0$ for all v, and $\|v\| = 0$ if and only if $v = 0$.*
- *(Cauchy-Schwarz Inequality): $|\langle u, v\rangle| \leq \|u\|\,\|v\|$.*
- *(Triangle inequality): $\|u + v\| \leq \|u\| + \|v\|$.*
- *(Pythagorean theorem) If $\langle u, v\rangle = 0$, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.*

Remark 3.1.8 The angle between vectors. For nonzero vectors $u, v \in \mathbb{R}^n$, the Cauchy-Schwarz inequality says that

$$\frac{|\langle u, v\rangle|}{\|u\|\,\|v\|} \leq 1, \text{ equivalently } -1 \leq \frac{\langle u, v\rangle}{\|u\|\,\|v\|} \leq 1.$$

Thus it makes sense to define a unique angle $\theta \in [0, \pi]$ with

$$\cos\theta := \frac{\langle u, v\rangle}{\|u\|\,\|v\|}$$

which we can call the angle between the vectors u, v. In some statistical interpretations of the vectors, the value of $\cos\theta$ is called a **correlation coefficient.**

3.2 Orthogonality and applications

Throughout all vector spaces are inner product spaces over the field $F = \mathbb{R}$ or $\mathbb{C}$ with inner product $\langle\cdot,\cdot\rangle$. Generally the vector spaces are finite-dimensional unless noted.

3.2.1 Orthogonal and Orthonormal Bases

Recall that a set S of vectors is **orthogonal** if every pair of distinct vectors in S is orthogonal, and the set is **orthonormal** if S is an orthogonal set of unit vectors.

Example 3.2.1 The standard basis in F^n. Let $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ be the standard basis in F^n (e_i has a one in the ith coordinate and zeros elsewhere). It is immediate to check that this is an orthonormal basis for F^n. □

We first make a very simple observation about an orthogonal set of nonzero vectors; they are linearly independent.

Proposition 3.2.2 *Let $S = \{v_i\}_{i\in I}$ be an orthogonal set of nonzero vectors. Then S is a linearly independent set.*

Here S can be an infinite set which is why we index its elements by a set I, but since the notion of linear (in)dependence only involves a finite number of vectors at a time, our proposition holds true in this broader setting.

Proof. Suppose that S is a linearly dependent set. Then there exist vectors $v_{i_1}, \ldots, v_{i_k} \in S$ and scalars a_{i_j} not all zero so that

$$v = a_{i_1}v_{i_1} + \cdots + a_{i_k}v_{i_k} = 0.$$

Indeed, there is no loss to assume all the coefficients are nonzero, so let's say $a_{i_1} \neq 0$. We know that since $v = 0$, $\langle v, v_{i_1}\rangle = 0$, but we now compute it differently and see

$$0 = \langle v, v_{i_1} \rangle = \sum_{j=1}^{k} a_{i_j} \langle v_{i_j}, v_{i_1} \rangle = a_{i_1} \langle v_{i_1}, v_{i_1} \rangle = a_{i_1} \|v_{i_1}\|^2.$$

But $v_{i_1} \neq 0$, so its length is nonzero, forcing $a_{i_1} = 0$, a contradiction. ■

Orthonormal bases offer distinct advantages in terms of representing coordinate vectors or the matrix of a linear map. For example if $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for a vector space V, we know that every $v \in V$ has a unique representation as $v = a_1 v_1 + \cdots + a_n v_n$ the coefficients of which provide the coordinate vector $[v]_{\mathcal{B}}$. But determining the coordinates is often a task which requires some work. With an orthonormal basis, this process is completely mechanical.

Theorem 3.2.3 *Let V, W be finite-dimensional inner product spaces with orthonormal bases $\mathcal{B}_V = \{e_1, \dots, e_n\}$ and $\mathcal{B}_W = \{f_1, \dots, f_m\}$.*

1. *Every vector $v \in V$ has a unique representation as $v = a_1 e_1 + \cdots + a_n e_n$ where $a_j = \langle v, e_j \rangle$.*

2. *If $T : V \to W$ is a linear map and $A = [T]_{\mathcal{B}_V}^{\mathcal{B}_W}$, then $A_{ij} = \langle T(e_j), f_i \rangle$.*

Proof of (1). Write $v = a_1 e_1 + \cdots + a_n e_n$. Then using the linearity of the inner product in the first variable and $\langle e_i, e_j \rangle = \delta_{ij}$, the Kronecker delta, we have

$$\langle v, e_j \rangle = \sum_{i=1}^{n} a_i \langle e_i, e_j \rangle = a_j.$$

■

Proof of (2). In Subsection 1.4.2, we saw that the matrix of T is given by $A = [T]_{\mathcal{B}_V}^{\mathcal{B}_W}$ where

$$T(e_j) = \sum_{k=1}^{m} A_{kj} f_k.$$

We now compute

$$\langle T(e_j), f_i \rangle = \langle \sum_{k=1}^{m} A_{kj} f_k, f_i \rangle = \sum_{k=1}^{m} A_{kj} \langle f_k, f_i \rangle = A_{ij}.$$

■

It is clear that orthonormal bases have distinct advantages and there is a standard algorithm to produce one from an arbitrary basis, but to understand why the algorithm should work, we need to review projections.

From applications of vector calculus, one recalls the **orthogonal projection** of a vector v onto the line spanned by a vector u. The projection is a vector parallel to u, so is of the form λu for some scalar λ. Referring to the figure below, if θ is the angle between the vectors u and v, then the length of $\operatorname{proj}_u v$ is $\|v\| \cos\theta$ (technically its absolute value). But $\cos\theta = \langle u, v \rangle / (\|u\| \|v\|)$, and the direction of u is given by the unit vector, $\dfrac{u}{\|u\|}$, parallel to u, so putting things together we see that

$$\operatorname{proj}_u v = (\|v\| \cos\theta)\frac{u}{\|u\|} = \|v\|\frac{\langle u, v\rangle}{\|u\|\|v\|}\frac{u}{\|u\|} = \frac{\langle u, v\rangle}{\|u\|^2}u,$$

so the scalar λ referred to above is $\dfrac{\langle u, v\rangle}{\|u\|^2}$. We also note that the vector $w := v - \operatorname{proj}_u v$ is orthogonal to u.

Now the key to an algorithm which takes an arbitrary basis to an orthogonal one is the above construction. Note that in the figure below, the vectors u and v are not parallel, so form a linearly independent set. The vectors u and w are orthogonal (hence linearly independent) and have the same span as the original vectors. Thus we have turned an arbitrary basis of two elements into an orthogonal one. The **Gram-Schmidt** process below extends this idea inductively.

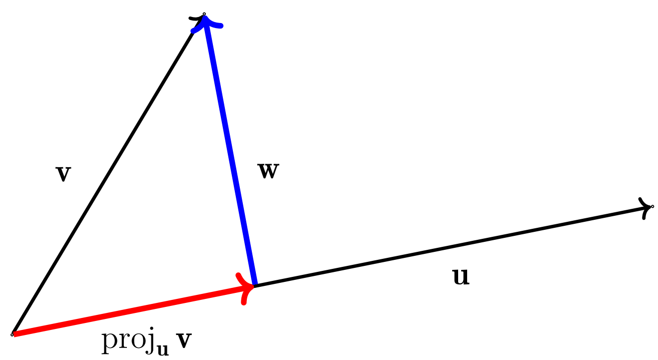

Figure 3.2.4 Orthogonal projection of vector v onto u

Algorithm 3.2.5 Gram-Schmidt process. *Let V be an inner product space, and W a subspace with basis $\mathcal{B} = \{v_1, \ldots, v_m\}$. To produce an **orthogonal** basis $\mathcal{E} = \{e_1, \ldots, e_m\}$ for W, proceed inductively.*

- *Let $e_1 = v_1$.*
- *Let $e_k = v_k - \displaystyle\sum_{j=1}^{k-1} \frac{\langle v_k, e_j\rangle}{\|e_j\|^2} e_j$, for $2 \le k \le m$.*

*To produce an **orthonormal** basis, normalize each vector replacing e_j with $e_j/\|e_j\|$.*

We note that the first two steps of the Gram-Schmidt process are exactly what we did above with the orthogonal projection.

3.2.2 Orthogonal complements and projections

Let V be an inner product space and W a subspace. Define

$$W^\perp = \{v \in V \mid \langle v, W\rangle = 0\}.$$

The set $W^\perp$ is called the **orthogonal complement of W in V**. The notation $\langle v, W\rangle = 0$ means that $\langle v, w\rangle = 0$ for all $w \in W$, so every vector in $W^\perp$ is orthogonal to every vector of W.

Example 3.2.6 The orthogonal complement of a plane. For example, if $V = \mathbb{R}^3$, and W is a line through the origin, then $W^\perp$, the orthogonal complement of W, is a plane through the origin for which the line defines the normal vector. □

Checkpoint 3.2.7 Is the orthogonal complement a subspace? If W is a subspace of a vector space V, is $W^\perp$ necessarily a subspace of V?

Hint. How do we check? Is $0 \in W^\perp$ (why?). If $u_1, u_2 \in W^\perp$ what about u_1+u_2 and λu_1? (why?)}

If may occur to you that the task of finding a vector in $W^\perp$ could be daunting since you have to check it is orthogonal to every vector in W. Or do you?

Checkpoint 3.2.8 How do we check if a vector is in the orthogonal complement? Let S be a set of vectors in a vector space V, and $W = \operatorname{Span}(S)$. Show that a vector $v \in W^\perp$ if and only if $\langle v, s\rangle = 0$ for every $s \in S$. This means there is only a finite amount of work for any subspace with a finite basis.

Moreover, we know that $W^\perp$ is a subspace of V, but what you have shown is that $S^\perp = W^\perp$ is also.

Hint. Everything in $\operatorname{Span}(S)$ is a linear combination of the elements of S, and we know how to expand $\langle v, \sum_{k=1}^m \lambda_i s_i\rangle$.

We shall see below that if V is an inner product space and W a finite-dimensional subspace, then every vector in V can be written uniquely as $v = w + w^\perp$, i.e., for unique $w \in W$ and $w^\perp \in W^\perp$. In different notation, that will say that $V = W \oplus W^\perp$, that V is the direct sum of W and $W^\perp$.

For now let us verify only the simple part of showing it is a direct sum, showing that $W \cap W^\perp = \{0\}$.

Proposition 3.2.9 *If V is an inner product space and W any subspace, then $W \cap W^\perp = \{0\}$.*

Proof. Let $w \in W \cap W^\perp$. If $w \neq 0$, then by the properties of an inner product $\langle w, w\rangle \neq 0$. But since $w \in W^\perp$, the vector w is orthogonal to every vector in W, in particular to w, a contradiction. ■

3.2.3 What good is an orthogonal complement anyway?

Let's say that after a great deal of work we have obtained an $m \times n$ matrix A and column vector b, and desperately want to solve the linear system $Ax = b$.

We know that the system is solvable if and only if b is in $C(A)$, the column space of A. But what if b is not is the column space? We want to solve this problem, right? Should we just throw up our hands?

This dilemma is not dissimilar from trying to find a rational number equal to $\sqrt{2}$. It cannot be done. But there are rational numbers arbitrarily close to $\sqrt{2}$. Perhaps an approximation to a solution would be good enough.

So now let's make the problem geometric. Suppose we have a plane P in $\mathbb{R}^3$ and a point x not on the plane. How would we find the point on P closest to the point x? Intuitively, we might "drop a perpendicular" from the point to the plane and the point x_0 where it intersects would be the desired closest point.

This is correct and gives us the intuition to develop the notion of an **orthogonal projection.** To apply it to our inconsistent linear system, we want to find a column vector $\hat{b}$ (in the column space of A) closest to b. We then check (see Corollary 3.2.15) that the solution $\hat{x}$ to $Ax = \hat{b}$ satisfies the property that

$$\|A\hat{x} - b\| \leq \|Ax - b\| \text{ for any } x \in \mathbb{R}^n.$$

Since the original system $Ax = b$ is not solvable, we know that $\|Ax - b\| > 0$ for every x, and that difference is an error term given by the distance between Ax and b. The value $\hat{x}$ minimizes the error, and is called the **least squares solution** to $Ax = b$ (since there is no exact solution). We shall explore this in more detail a bit later.

3.2.4 Orthogonal Projections

Now we want to take our intuitive example of "dropping a perpendicular" and develop it into a formal tool for inner product spaces.

Let V be an inner product space and W be a finite-dimensional subspace. Since W has a basis, we can use the Gram-Schmidt process to produce and orthogonal basis $\{w_1, \ldots, w_r\}$ for W.

Theorem 3.2.10 *Let $\{w_1, \ldots, w_r\}$ be an orthogonal basis for a subspace W of an inner product space V. Each vector $v \in V$ can be represented uniquely as $v = w^\perp + w$ where $w \in W$, and $w^\perp \in W^\perp$, that is $w^\perp$ is orthogonal to W. Moreover,*

$$w = \frac{\langle v, w_1\rangle}{\langle w_1, w_1\rangle} w_1 + \cdots + \frac{\langle v, w_r\rangle}{\langle w_r, w_r\rangle} w_r. \tag{3.2.1}$$

Proof. Certainly w as defined is an element of W, and to see that $w^\perp = v - w$ is orthogonal to W, it is sufficient by Checkpoint 3.2.8 to verify that $\langle w^\perp, w_i\rangle = 0$ for each $i = 1, \ldots, r$.
Using the definition of $w^\perp$ and bilinearity of the inner product we have

$$\langle w^\perp, w_i\rangle = \langle v - w, w_i\rangle = \langle v, w_i\rangle - \langle w, w_i\rangle,$$

and since the$\{w_j\}$ form an orthogonal basis, the expression for w in (3.2.1) gives

$$\langle w, w_i\rangle = \langle \frac{\langle v, w_i\rangle}{\langle w_i, w_i\rangle} w_i, w_i\rangle = \langle \frac{\langle v, w_i\rangle}{\langle w_i, w_i\rangle} \langle w_i, w_i\rangle = \langle v, w_i\rangle.$$

It is now immediate from the first displayed equation that

$$\langle w^\perp, w_i\rangle = \langle v - w, w_i\rangle = \langle v, w_i\rangle - \langle w, w_i\rangle = \langle v, w_i\rangle - \langle v, w_i\rangle = 0,$$

as desired.
Finally to see that $w^\perp$ and w are uniquely determined by these conditions, suppose that as above $v = w^\perp + w$, and also $v = w_1^\perp + w_1$ with $w_1 \in W$ and $w_1^\perp \in W^\perp$.
Setting the two expressions equal to each other and solving gives that

$$w - w_1 = w_1^\perp - w^\perp.$$

But the left hand side is an element of W while the right hand side is an element of $W^\perp$, so by Proposition 3.2.9, both expressions equal zero, which gives the uniqueness. ■

Corollary 3.2.11 *Let V be an inner product space and W be a finite-dimensional subspace. Then*

$$V = W \oplus W^\perp.$$

*In this case the direct sum is an **orthogonal sum**, so the expression is often written as*

$$V = W \boxplus W^\perp.$$

Another useful property of the orthogonal complement is

Corollary 3.2.12 *Let V be an inner product space and W a finite-dimensional subspace. Then*

$$(W^\perp)^\perp = W.$$

Proof. Recall that

$$W^\perp = \{v \in V \mid \langle v, W\rangle = 0\},$$

so

$$(W^\perp)^\perp = \{v \in V \mid \langle v, W^\perp\rangle = 0.$$

In particular, every $w \in W$ is orthogonal to all of $W^\perp$, so that $W \subseteq (W^\perp)^\perp$. The other containment takes a bit more care.
Let $v \in (W^\perp)^\perp$. Since W is finite-dimensional, Theorem 3.2.10 says that v can be written uniquely as

$$v = w^\perp + w$$

where $w \in W$ and $w^\perp \in W^\perp$. The goal is to show that $w^\perp = 0$.
Consider $w^\perp = v - w$. Since $v \in (W^\perp)^\perp$, and $w \in W \subseteq (W^\perp)^\perp$, we conclude $w^\perp \in (W^\perp)^\perp$, so $\langle w^\perp, W^\perp\rangle = 0$. But $w^\perp \in W^\perp$ by the theorem, so $\langle w^\perp, w^\perp\rangle = 0$ implying that $w^\perp = 0$ by the axioms for an inner product. Thus $v = w \in W$, meaning $(W^\perp)^\perp \subseteq W$, giving us the desired equality. ■

Definition 3.2.13 If V is an inner product space and W a finite-dimensional subspace with orthogonal basis $\{w_1, \dots, w_r\}$, then the **orthogonal projection of a vector v onto the subspace** W is given by the expression in Theorem 3.2.10:

$$\operatorname{proj}_W v := \frac{\langle v, w_1\rangle}{\langle w_1, w_1\rangle} w_1 + \cdots + \frac{\langle v, w_r\rangle}{\langle w_r, w_r\rangle} w_r.$$

◊

Corollary 3.2.14 *Let V be an inner product space and W be a finite-dimensional subspace. If $w \in W$, then*

$$\operatorname{proj}_W w = w.$$

Proof. Combining Theorem 3.2.10 with the definition of projection, we know that w can be written uniquely as $w = w^\perp + \operatorname{proj}_W w$, where $w^\perp \in W^\perp$. But $w = 0 + w$, so $w^\perp = 0$ and $w = \operatorname{proj}_W w$. ∎

To complete our formalization of the idea of dropping a perpendicular, we now show that the projection $\operatorname{proj}_W v$ of a vector v is the unique vector in W closest to v.

Corollary 3.2.15 *Let V be an inner product space and W be a finite-dimensional subspace. If $v \in V$, then*

$$\|v - \operatorname{proj}_W v\| < \|v - w\|$$

for all $w \in W$, with $w \neq \operatorname{proj}_W v$.

Proof. By Corollary 3.2.14, we may assume that $v \notin W$, so consider any $w \in W$ with $w \neq \operatorname{proj}_W v$. We certainly know that

$$v - w = v - \operatorname{proj}_W v + \operatorname{proj}_W v - w,$$

and we know that $\operatorname{proj}_W v - w \in W$ while by Theorem 3.2.10 we know that $v - \operatorname{proj}_W v \in W^\perp$. Thus the vectors $v - w$, $v - \operatorname{proj}_W v$ and $\operatorname{proj}_W v - w$ form a right triangle whose lengths satisfy the Pythagorean identity:

$$\|v - w\|^2 = \|v - \operatorname{proj}_W v\|^2 + \|\operatorname{proj}_W v - w\|^2.$$

It follows that if $w \neq \operatorname{proj}_W v$, that $\|\operatorname{proj}_W v - w\| > 0$, so that $\|v - w\| > \|v - \operatorname{proj}_W v\|$. ∎

3.2.5 A first look at the four fundamental subspaces

While in the previous section, we have seen how orthogonal projections and complements are related, there is another prominent place in which orthogonal complements arise naturally.

Let $A \in M_{m\times n}(\mathbb{C})$. Associated to A we have a linear transformation $L_A : \mathbb{C}^n \to \mathbb{C}^m$ given by left multiplication by A. To obviate the need to introduce L_A, we often write $\ker A$ for $\ker L_A$, and $\operatorname{range} A$ for $\operatorname{range} L_A$ which we know is the column space, $C(A)$, of A.

Additionally, we also have a linear transformation $L_{A^*} : \mathbb{C}^m \to \mathbb{C}^n$ given by left multiplication by A^*. We have the following very useful property relating A and A^*:

Proposition 3.2.16 *Let $A \in M_{m\times n}(\mathbb{C})$. For $x \in \mathbb{C}^n$ and $y \in \mathbb{C}^m$, we have*

$$\langle Ax, y\rangle_m = \langle x, A^*y\rangle_n,$$

where we have subscripted the inner product symbols to remind the reader of the ambient inner product space, $\mathbb{C}^m$ or $\mathbb{C}^n$.

Proof. Recall the inner product $\langle v, w\rangle$ in $\mathbb{C}^\ell$ is w^*v the matrix product of a $1 \times \ell$ row vector with an $\ell \times 1$ column vector. Thus

$$\langle Ax, y\rangle_m = y^*Ax = (A^*y)^*x = \langle x, A^*y\rangle_n.$$

■

Many authors, e.g., [2] and [3], define the **four fundamental subspaces**. For complex matrices, these are most easily described by the kernel and range of A and A^*. For real matrices, the same identities can be rewritten in terms of the row and column spaces of A and A^T. The significance of these four subspaces will be evident when we discuss the **singular value decomposition of a matrix** in Section 3.6, but for now we reveal their basic relations.

Theorem 3.2.17 *Let $A \in M_{m\times n}(\mathbb{C})$. Then*

$$\ker(A^*) = \operatorname{range}(A)^\perp \text{ and } \operatorname{range}(A^*) = C(A^*) = \ker(A)^\perp.$$

Proof. Let $w \in \ker A^*$. Then $A^*w = 0$, hence $\langle A^*w, v\rangle = 0$ for all $v \in \mathbb{C}^n$. By taking complex conjugates in Proposition 3.2.16,

$$0 = \langle A^*w, v\rangle = \langle w, Av\rangle,$$

so w is orthogonal to everything in $\operatorname{range}(A) = C(A)$. This gives the inclusion $\ker(A^*) \subseteq \operatorname{range}(A)^\perp$.
Conversely, if $w \in \operatorname{range}(A)^\perp$, then for all $v \in \mathbb{C}^n$,

$$0 = \langle w, Av\rangle = \langle A^*w, v\rangle.$$

In particular, taking $v = A^*w$, we have $\langle A^*w, A^*w\rangle = 0$ which means that $A^*w = 0$, showing that $\operatorname{range}(A)^\perp \subseteq \ker(A^*)$, giving us the first equality.
Since the first equality is valid for any matrix A, we replace A by A^*, and use that $A^{**} = A$ to conclude that

$$\ker(A) = \operatorname{range}(A^*)^\perp.$$

Using Corollary 3.2.12 yields

$$\ker(A)^\perp = \operatorname{range}(A^*).$$

■

For real matrices, these become

Corollary 3.2.18 *Let $A \in M_{m\times n}(\mathbb{R})$. Then*

$$C(A)^\perp = \ker(A^T) \text{ and } R(A)^\perp = \ker A.$$

Proof. The first statement is immediate from the previous theorem since $\mathrm{range}(A) = C(A)$. For the second, we had deduced above that $\ker(A) = \mathrm{range}(A^*)^\perp$. Now if A is a real matrix,

$$\mathrm{range}(A^*) = \mathrm{range}(A^T) = C(A^T) = R(A)$$

which finishes the proof.

■

3.3 Orthogonal Projections and Least Squares Approximations

We begin with the notion of orthogonal projection introduced in the previous section. We find different ways to compute it other than from the definiton, and give an application to least squares approximations.

3.3.1 Orthonormal bases and orthogonal/unitary matrices.

Consider the inner product space $V = F^n$ where $F = \mathbb{R}$ or $\mathbb{C}$, and denote by $\overline{z}$ the complex conjugate of z.

If $v = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $w = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ are two vectors in F^n, we defined their inner product by:

$$\langle v, w \rangle := \sum_{i=1}^{n} a_i \overline{b_i}.$$

It is very convenient to recognize values of the inner product via matrix multiplication. In particular, regarding the column vectors v, w as $n \times 1$ matrices

$$\langle v, w \rangle := \sum_{i=1}^{n} a_i \overline{b_i} = w^* v$$

is the 1×1 matrix product $w^* v$ where w^* is the $1 \times n$ conjugate-transpose matrix to w.

For vectors v, w as above, we have seen the meaning of $w^* v$. It is more than idle curiosity to inquire about the meaning of vw^*. We can certainly compute it, but first we note that while $w^* v = \langle v, w \rangle$ is a scalar (a 1×1 matrix), in the reverse order, vw^* is an $n \times n$ matrix, specifically:

$$vw^* = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} \overline{b_1} & \cdots & \overline{b_n} \end{bmatrix} = \begin{bmatrix} a_1\overline{b_1} & \cdots & a_1\overline{b_n} \\ a_2\overline{b_1} & \cdots & a_2\overline{b_n} \\ \vdots & \vdots & \vdots \\ a_n\overline{b_1} & \cdots & a_n\overline{b_n} \end{bmatrix}.$$

It is probably a bit more useful to see how this product arises naturally in what we have done so far.

Let apply the definition of an orthogonal projection to the inner product space $V = \mathbb{C}^n$; what happens for $V = \mathbb{R}^n$ will be clear.

Let W be an $r-$dimensional subspace of V with ***orthonormal*** basis $\{w_1, \dots, w_r\}$. Then Definition 3.2.13 tells us that the orthogonal projection of a vector v into the subspace W is given by:

$$\operatorname{proj}_W v := \langle v, w_1\rangle w_1 + \cdots + \langle v, w_r\rangle w_r.$$

Now while for a vector space V over a field F, we have defined multiplication of a scalar λ times a vector v as λv, you might ask if we would get into trouble if we defined $v\lambda := \lambda v$. Since multiplication in a field is commutative, this turns out to be just fine, but in more general structures (modules over rings) there can be significant issues. So with that as preamble, let's consider a summand $\langle v, w_j\rangle w_j$ in the expression for an orthogonal projection. First we use that scalar multiplication can be thought of on the right or the left and then we use the specific nature of the inner product on $\mathbb{C}^n$, so that

$$\langle v, w_j\rangle w_j = w_j\langle v, w_j\rangle = w_j w_j^* v$$

Thus as a corollary we obtain a matrix-specific characterization of an orthogonal projection to a subspace of $\mathbb{C}^n$.

Corollary 3.3.1 *Let W be a subspace of $\mathbb{C}^n$ with orthonormal basis $\{w_1, \dots, w_r\}$. Then for any vector $v \in \mathbb{C}^n$,*

$$\operatorname{proj}_W v := \langle v, w_1\rangle w_1 + \cdots + \langle v, w_r\rangle w_r = \sum_{k=1}^{r} w_k w_k^* v = \left(\sum_{k=1}^{r} w_k w_k^*\right) v,$$

where we note that the last expression is the matrix multiplication of an $n \times n$ matrix times the $n \times 1$ vector v.

Our next goal is to give a more intrinsic characterization of the matrix $\sum_{k=1}^{r} w_k w_k^*$. Let A be the $n \times r$ matrix whose columns are the orthonormal basis $\{w_1, \dots, w_r\}$ of the subspace W. What should the matrix A^*A look like?

Using our familiar (row-column) method of multiplying two matrices together, the ij entry of the product is

$$w_i^* w_j = \langle w_j, w_i\rangle = \delta_{ij} \text{ (Kronecker delta)},$$

so that $A^*A = I_r$, the $r \times r$ identity matrix.

In the other order we claim that

$$AA^* = \sum_{k=1}^{r} w_k w_k^*$$

from Corollary 3.3.1, that is, AA^* is the matrix of the orthogonal projection (with respect to the standard basis) of a vector to the subspace W.

This claim is most easily justified using the "column-row" expansion of a matrix product as given in [2]. If A is an $n \times r$ matrix (as it is for us), and B is an $r \times m$ matrix, then

$$AB = \mathrm{col}_1(A)\mathrm{row}_1(B) + \cdots + \mathrm{col}_r(A)\mathrm{row}_r(B).$$

Proof. The proof is simply a computation, but it is easy to make an error, so we do it out explicitly. Note that each summand is the product of an $n \times 1$ matrix times an $1 \times m$ matrix.

$$\mathrm{col}_1(A)\mathrm{row}_1(B) + \cdots + \mathrm{col}_r(A)\mathrm{row}_r(B) =$$
$$\begin{bmatrix} a_{11}b_{11} & \cdots & a_{11}b_{1m} \\ a_{21}b_{11} & \cdots & a_{21}b_{1m} \\ \vdots & \vdots & \vdots \\ a_{n1}b_{11} & \cdots & a_{n1}b_{1m} \end{bmatrix} + \cdots + \begin{bmatrix} a_{1r}b_{r1} & \cdots & a_{1r1}b_{rm} \\ a_{2r}b_{r1} & \cdots & a_{2r}b_{rm} \\ \vdots & \vdots & \vdots \\ a_{nr}b_{r1} & \cdots & a_{nr}b_{rm} \end{bmatrix}.$$

Now from the row-column rule we know that the ij entry of AB is $(AB)_{ij} = \sum_{k=1}^{r} a_{ik}b_{kj}$, which is exactly the sum of the ij entries from each of the r matrices above. ■

Now we apply this to the product of the matrices AA^*. The column-row rule immediately gives that

$$AA^* = w_1w_1^* + \cdots + w_rw_r^*$$

as claimed. We summarize this as

Corollary 3.3.2 *Let W be a subspace of $\mathbb{C}^n$ with orthonormal basis $\{w_1, \ldots, w_r\}$, and let A be the $n \times r$ matrix whose columns are those orthonormal basis vectors. Then for any vector $v \in \mathbb{C}^n$,*

$$\mathrm{proj}_W\, v := AA^*v \text{ and } A^*A = I_r.$$

While this is a very pretty expression for the orthogonal projection onto a subspace W, it is predicated on having an orthonormal basis for the subspace. Of course Gram-Schmidt can be employed, but it is an interesting exercise to produce a matrix representation of the projection starting from an arbitrary basis for the subspace. We reproduce Proposition 4.18 of [3] including a proof which includes several interesting ideas.

Proposition 3.3.3 *Let W be a subspace of $\mathbb{C}^n$ (or $\mathbb{R}^n$) with arbitrary basis $\{v_1, \ldots, v_r\}$. Let A be the $n \times r$ matrix with columns $v_1, \ldots, v_r$. Then the matrix of the orthogonal projection, proj_W, with respect to the standard basis is*

$$A(A^*A)^{-1}A^*.$$

Before giving the proof, let's make a few observations. First is that we must prove that the matrix A^*A is invertible. Second, what does this more complicated expression look like when the given basis is actually orthonormal? But that one is easy. We observed above that under those assumptions, A^*A was just the $r \times r$ identity matrix, so our complicated expression in the proposition reduces to AA^* as we proved in the earlier case. So there is some measure of confidence.

Proof. Given a vector v, we know its orthogonal projection, $\operatorname{proj}_W v$ is an element of W so a linear combination of the basis for W, say

$$\operatorname{proj}_W v = \lambda_1 v_1 + \cdots + \lambda_r v_r.$$

On the other hand this linear combination can be represented as the matrix product

$$\lambda_1 v_1 + \cdots + \lambda_r v_r = A\boldsymbol{\lambda}$$

where

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_r \end{bmatrix}.$$

Thus we begin with the identity

$$\operatorname{proj}_W v = A\boldsymbol{\lambda}.$$

By Theorem 3.2.10, we know that $v - \operatorname{proj}_W v = v - A\boldsymbol{\lambda} \in W^\perp$ so that for all $j = 1, \ldots, r$

$$\langle v - A\boldsymbol{\lambda}, v_j \rangle = v_j^*(v - A\boldsymbol{\lambda}) = 0.$$

Writing the system of r equations as a single matrix equation, we have

$$A^*(v - A\boldsymbol{\lambda}) = 0 \text{ or equivalently } A^*v = A^*A\boldsymbol{\lambda}.$$

Assuming for the moment that A^*A is invertible, we multiply both sides of $A^*v = A^*A\boldsymbol{\lambda}$ by $A(A^*A)^{-1}$ to obtain

$$A(A^*A)^{-1}A^*v = A(A^*A)^{-1}(A^*A)\boldsymbol{\lambda} = A\boldsymbol{\lambda} = \operatorname{proj}_W v,$$

as desired.

Finally, we must check that the $r \times r$ matrix A^*A is invertible. By the rank-nullity theorem it suffices to know that A^*A has trivial nullspace. So suppose that $A^*Av = 0$. Since $\langle 0, v \rangle = 0$, we can write

$$0 = \langle A^*Av, v \rangle = v^*(A^*Av) = (Av)^*(Av) = \|Av\|^2.$$

Thus $A^*Av = 0$ implies $Av = 0$, but A is an $n \times r$ matrix which defines a linear map from $\mathbb{C}^r \to \mathbb{C}^n$. Since A has r linearly independent columns, it has rank r. By the rank-nullity theorem, it follows that the nullity of A is zero, so $Av = 0$ implies $v = 0$. Thus A^*A has trivial nullspace and so is invertible. ■

Let's work through an example showing an orthogonal projection using the three different characterizations given above. We fix the vector space $V = \mathbb{R}^3$, and let $w_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ and $w_2 = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$, $W = \operatorname{Span}\{w_1, w_2\}$, and $y = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. We note that w_1 and w_2 are orthogonal, but not orthonormal and claim $y \notin W$.

Example 3.3.4 From the definition. Using Definition 3.2.13, we see that

$$\operatorname{proj}_W v = \frac{\langle y, w_1\rangle}{\langle w_1, w_1\rangle} w_1 + \frac{\langle y, w_2\rangle}{\langle w_r, w_r\rangle} w_2 = \frac{-2}{6}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + \frac{2}{30}\begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/5 \\ 4/5 \end{bmatrix}.$$

We also check that

$$y^\perp = y - \operatorname{proj}_W y = \begin{bmatrix} 0 \\ 2/5 \\ 1/5 \end{bmatrix} \in W^\perp.$$

□

Example 3.3.5 Using a matrix with orthonormal columns. Normalizing the vectors w_1 and w_2, we obtain a matrix with orthonormal columns spanning W:

$$A = \begin{bmatrix} 1/\sqrt{6} & 5/\sqrt{30} \\ 1/\sqrt{6} & -1/\sqrt{30} \\ -2/\sqrt{6} & 2/\sqrt{30} \end{bmatrix}$$

That A has orthonormal columns implies that $A^*A (= A^T A) = I_2$ (the two-by-two identity matrix), but that the matrix of proj_W with respect to the standard basis for $\mathbb{R}^3$ is

$$[\operatorname{proj}_W] = AA^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/5 & -2/5 \\ 0 & -2/5 & 4/5 \end{bmatrix}$$

and we check that

$$\operatorname{proj}_W y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/5 & -2/5 \\ 0 & -2/5 & 4/5 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/5 \\ 4/5 \end{bmatrix}.$$

□

Example 3.3.6 Using the given vectors in matrix form. Now we use Proposition 3.3.3 with the original vectors as the columns of the matrix

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

So the matrix of the projection is

$$[\operatorname{proj}_W] = A(A^*A)^{-1}A^*.$$

We note that

$$A^*A = \begin{bmatrix} 6 & 0 \\ 0 & 30 \end{bmatrix} \text{ so } (A^*A)^{-1} = \begin{bmatrix} 1/6 & 0 \\ 0 & 1/30 \end{bmatrix}$$

and

$$[\operatorname{proj}_W] = A(A^*A)^{-1}A^* =$$
$$\begin{bmatrix} 1 & 5 \\ 1 & -1 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} 1/6 & 0 \\ 0 & 1/30 \end{bmatrix} \begin{bmatrix} 1 & 1 & -2 \\ 5 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/5 & -2/5 \\ 0 & -2/5 & 4/5 \end{bmatrix}$$

as in the previous computation. □

Remark 3.3.7 Which is the better method? At first blush (maybe second too), it sure looks like the first example gives a method with the least amount of work. So why should we even consider the second or third methods?

The answer depends upon the intended application. If there is a single computation to make, the first method is mostly likely the most efficient, but if you must compute the orthogonal projection of many vectors into the same subspace, then the matrix method is far superior since you only compute the matrix once.

Examples of multiple projections include writing a computer graphics program which renders a three dimensional image on a flat screen (aka a plane).

Remark 3.3.8 One final comment of note. Since

$$V = W \boxplus W^\perp,$$

we know that the identity operator I_V can be written as

$$I_V = \operatorname{proj}_W + \operatorname{proj}_{W^\perp}.$$

This means that

$$\operatorname{proj}_W v = v - \operatorname{proj}_{W^\perp} v,$$

so if the dimension of $W^\perp$ is smaller than that of W, it may make more sense to compute $\operatorname{proj}_{W^\perp}$ and subtract it from the identity to obtain the desired projection.

Example 3.3.9 Point closest to a plane. Let's do another example illustrating some of the concepts above. Let $V = \mathbb{R}^3$ and W be the subpace described by $3x - y - 5z = 0$. Let's find the point on the plane closest to the point $v = (1, 1, 1)$.

We know that the plane W is spanned by any two linearly independent vectors in W, say

$$v_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \text{ and } v_2 = \begin{bmatrix} 0 \\ -5 \\ 1 \end{bmatrix}.$$

We form the matrix whose columns are v_1 and v_2, and use Proposition 3.3.3 to compute the matrix of the projection (with respect to the standard basis) as

$$[\mathrm{proj}_W] = \begin{bmatrix} \frac{26}{35} & \frac{3}{35} & \frac{3}{7} \\ \frac{3}{35} & \frac{34}{35} & -\frac{1}{7} \\ \frac{3}{7} & -\frac{1}{7} & \frac{2}{7} \end{bmatrix}$$

Thus

$$\mathrm{proj}_W v = \begin{bmatrix} \frac{26}{35} & \frac{3}{35} & \frac{3}{7} \\ \frac{3}{35} & \frac{34}{35} & -\frac{1}{7} \\ \frac{3}{7} & -\frac{1}{7} & \frac{2}{7} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{35} \begin{bmatrix} 44 \\ 32 \\ 20 \end{bmatrix}.$$

On the other hand, we could arrive at the answer via $\mathrm{proj}_{W^\perp}$. Since $W^\perp$ is spanned by $v_3 = \begin{bmatrix} 3 \\ -1 \\ -5 \end{bmatrix}$

$$\mathrm{proj}_{W^\perp} v = \frac{\langle v, v_3 \rangle}{\langle v_3, v_3 \rangle} v_3 = \frac{1}{35} \begin{bmatrix} -9 \\ 3 \\ 15 \end{bmatrix},$$

so

$$\mathrm{proj}_W v = v - \mathrm{proj}_{W^\perp} v = \frac{1}{35} \begin{bmatrix} 44 \\ 32 \\ 20 \end{bmatrix}.$$

□

3.3.2 Sage Compuations

In this section, we use Sage to make some of the computations in the above examples. In those examples, we started with an orthogonal basis spanning the subspace W in $V = \mathbb{R}^3$, given by $w_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ and $w_2 = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$.

Of course, more typically we have an arbitrary basis and need to use Gram-Schmidt to produce an orthogonal one. Also recall that Gram-Schmidt simply accepts the first of the given vectors as the first in the orthogonal basis. So let's start with the basis $w_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ and $w_2' = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = w_1 + w_2$ (so that the basis is not already orthogonal).

So we build a matrix A whose ***row*** vectors are w_1 and w_2'. The Gram-Schmidt algorithm in Sage returns two matrices: G is a the matrix whose rows are an orthogonal basis, and M is the matrix which tells the linear combinations of the given rows used to produce the orthogonal rows. As you will see, we return to our original orthogonal basis.

```
%display latex
latex.matrix_delimiters("[", "]")
A=matrix(QQbar, [[1,1,-2], [6,0,0]])
G,M=A.gram_schmidt()
(A,G)
```

```
([ 1  1 -2]  [ 1  1 -2]
[ 6  0  0], [ 5 -1  2])

\left(\left[\begin{array}{rrr}
1 \amp 1 \amp -2 \\
6 \amp 0 \amp 0
\end{array}\right], \left[\begin{array}{rrr}
1 \amp 1 \amp -2 \\
5 \amp -1 \amp 2
\end{array}\right]\right)
```

Next we compute the orthogonal projection of the vector $v = [0, 0, 1]$ onto the plane W using the definition of the orthogonal projection. Notice that the rows of the matrix A are now the orthogonal basis for W.

```
v=vector(QQbar,[0,0,1])
A=matrix(QQbar,  [G[0],G[1]])
OP = vector(QQbar,[0,0,0])
for i in range(G.nrows()):
    scalar =
       v.inner_product(G[i])/(G[i].inner_product(G[i]))
    OP = OP + scalar*G[i]
OP
```

```
(0, -2/5, 4/5)
```

Finally here we make A into a matrix with orthogonal columns to coincide with Proposition 3.3.3. We then compute the matrix of the proj_W with respect to the standard basis.

```
A= A.transpose()
A*  (A.conjugate_transpose()*A).inverse()  *
   A.conjugate_transpose()
```

```
[   1    0    0]
[   0  1/5 -2/5]
```

```
[   0 -2/5  4/5]

\left[\begin{array}{rrr}
1 \amp 0 \amp 0 \\
0 \amp \frac{1}{5} \amp -\frac{2}{5} \\
0 \amp -\frac{2}{5} \amp \frac{4}{5}
\end{array}\right]
```

3.3.3 More on orthogonal projections

We return to a motivating example: how to deal with inconsistent linear system $Ax = b$. Since the system is inconsistent, we know that $\|Ax - b\| > 0$ for every x in the domain. Want we want to do is find a vector $\hat{x}$ which minimizes the error, that is for which

$$\|A\hat{x} - b\| \leq \|Ax - b\|$$

for every x in the domain.

So we let W be the column space of the $m \times n$ matrix A and let $\hat{b} = \operatorname{proj}_W b$. Since $\hat{b}$ is in the column space, the system $Ax = \hat{b}$ is consistent. With $\hat{x}$ any solution to $Ax = \hat{b}$ Corollary 3.2.15 says that

$$\|A\hat{x} - b\| \leq \|Ax - b\|$$

for every x in the domain.

To compute this solution, there are multiple paths. Of course, we could compute the orthogonal projection, $\hat{b}$ and solve the consistent system $Ax = \hat{b}$, but what if we could solve it without finding the orthogonal projection? That would be a significant time-saver.

Let's start from the premise that we have found the orthogonal projection, $\hat{b}$ of b into $W = C(A)$, and a solution $\hat{x}$ to $Ax = \hat{b}$. Now by Theorem 3.2.10, $b = \hat{b} + b^\perp$ where

$$b^\perp = b - \hat{b} = b - A\hat{x} \in W^\perp.$$

Since W, the column space of A, is the image (range) of the linear map $x \mapsto Ax$, we deduce that

$$\langle Ax, A\hat{x} - b\rangle_m = 0.$$

By Proposition 3.2.16, we deduce

$$\langle Ax, A\hat{x} - b\rangle_m = \langle x, A^*(A\hat{x} - b)\rangle_n = 0,$$

for every $x \in \mathbb{C}^n$. By the positivity property of any inner product, that means that $A^*(A\hat{x} - b) = 0$. Thus to find $\hat{x}$, we need only find a solution to the new linear system

$$A^*A\hat{x} = A^*b.$$

We summarize this as

Corollary 3.3.10 *Let* $A \in M_{m\times n}(\mathbb{C})$, *and* $b \in \mathbb{C}^m$. *Then there is an* $\hat{x} \in \mathbb{C}^n$ *so that*

$$\|A\hat{x} - b\| \le \|Ax - b\|$$

for all $x \in \mathbb{C}^n$. *Moreover, the solution(s),* $\hat{x}$ *are acquired by solving the consistent linear system* $A^*A\hat{x} = A^*b$.

Checkpoint 3.3.11 Does it matter which solution $\hat{x}$ we pick? In a theoretical sense the answer is no, but in a computational sense, the answer is probably. Of course if the system has a unique solution, the issue is resolved, but if it has more than one solution, there are infinitely many since any two differ by something in the nullspace of A. How should one choose?

3.3.4 Least Squares Examples

A common problem is determining a curve which best describes a set of observed data points. The curve may be a polynomial, exponential, logarithmic, or something else. Below we investigate how to produce a polynomial which represents a **least squares** approximation to a set of data points. We begin with the simplest example, linear regression.

Consider the figure below in which two observed data points are plotted at (x_i, y_i) and (x_j, y_j). The goal is to find an equation of a line of the $y = mx + b$ which "best describes" the given data, but what does that mean? Since in general, not all data points will lie on any chosen line, each choice of line will produce some **error** in approximation. Our first job is to decide on a method to measure the error. Looking at this generally, suppose we have observed data $\{(x_i, y_i) \mid i = 1 \dots n\}$ and we are trying the find the best function $y = f(x)$ which fits the data.

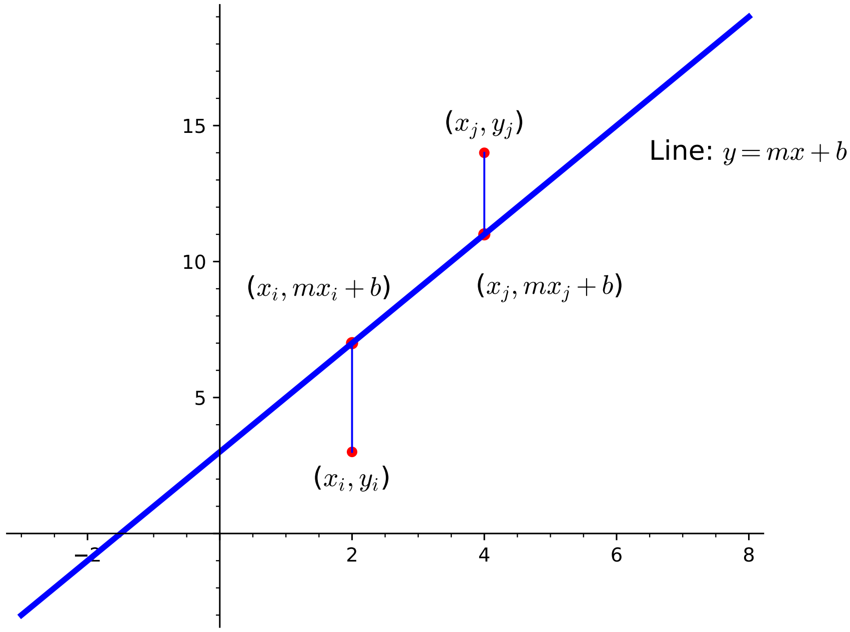

Figure 3.3.12 The concept of a least squares approximation

We could set the error to be

$$E = \sum_{i=1}^{n} (y_i - f(x_i)),$$

but we immediately see this is a poor choice for when $y_i > f(x_i)$ the error is counted as positive while when $y_i < f(x_i)$, the error is counted as negative, so it would be possible for a really poor approximation to produce a small error by having positive errors balanced by negative ones. Of course one solution would be simply to take absolute values, but they are often a bit challenging to work with, so for this and reasons connected to the inner product on $\mathbb{R}^n$, we choose a sum of squares of the errors:

$$E = \sum_{i=1}^{n} (y_i - f(x_i))^2,$$

so for our linear model the error is

$$E = \sum_{i=1}^{n} (y_i - mx_i - b)^2.$$

So where is the linear algebra? It might occur to you in staring that the expression for the error that if we had two vectors

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \text{ and } Z = \begin{bmatrix} mx_1 + b \\ \vdots \\ mx_n + b \end{bmatrix}$$

that our error is
$$E = \|Y - Z\|^2 = \langle Y - Z, Y - Z\rangle.$$

It is clear where the vector Y comes from, but let's see if we can get a matrix involved to describe Z. Let
$$A = \begin{bmatrix} x_1 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix}.$$

Then
$$Z = \begin{bmatrix} mx_1 + b \\ \vdots \\ mx_n + b \end{bmatrix} = A \begin{bmatrix} m \\ b \end{bmatrix}.$$

What would it mean if all the data points were to lie on the line? Of course it would mean the error is zero, but to move us in the direction of work we have already done, it would mean that
$$A \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

in other words the linear system
$$AX = Y$$

is solvable with solution $X = \begin{bmatrix} m \\ b \end{bmatrix}$.

When the data points do not all lie on the line, the original system is inconsistent, but Corollary 3.3.10 tells us how to find the best solution $\hat{X} = \begin{bmatrix} m \\ b \end{bmatrix}$ for which
$$\|A\hat{X} - Y\| \le \|AX - Y\|$$

for all $X \in \mathbb{R}^2$. Recalling that our error $E = \|A\hat{X} - Y\|^2$, this will solve our problem.

A simple example.

Suppose we have collected the following data points (x, y):
$$\{(2, 1), (5, 2), (7, 3), (8, 3)\}.$$

We construct the matrix
$$A = \begin{bmatrix} x_1 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix},$$

and

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

Using Corollary 3.3.10, we solve the consistent linear system

$$A^*A\hat{X} = A^*Y :$$

$$A^*A = \begin{bmatrix} 2 & 5 & 7 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix} = \begin{bmatrix} 142 & 22 \\ 22 & 4 \end{bmatrix}$$

and

$$A^*Y = \begin{bmatrix} 2 & 5 & 7 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 57 \\ 9 \end{bmatrix}$$

We note that A^*A is invertible, so that

$$\begin{bmatrix} 142 & 22 \\ 22 & 4 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 57 \\ 9 \end{bmatrix}$$

has a unique solution:

$$\hat{X} = \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 5/14 \\ 2/7 \end{bmatrix},$$

that is the desired line is $y = 5/14x + 2/7$. We plot the data points and the line of regression below. Note that the first point lies on the line.

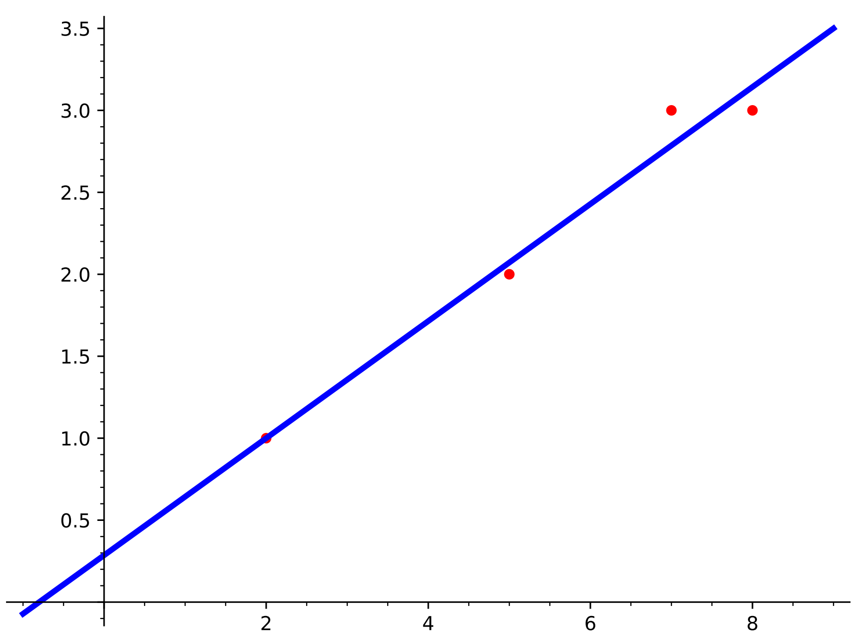

Figure 3.3.13 A simple linear regression

We now consider higher degree polynomial approximations. For background, we know that two points determine a line so we need to use linear regression as soon as we have more than two points. **Lagrange interpolation** tells us that given n points in the plane, no two on a vertical line, there is a unique polynomial of degree $n-1$ which passes through them. We consider the quadratic case. So as soon as there are more than 3 points, we are no longer guaranteed a unique quadratic curve passing through them, so we desire a least squares approximation.

Now we are looking for coefficients b_0, b_1, b_2 so that $y = b_2x^2 + b_1x + b_0$ best approximates the data. As before assume that we have observed data

$$(x_i, y_i) = (60, 3.1), (61, 3.6), (62, 3.8), (63, 4), (65, 4.1), i = 1 \dots 5.$$

In our quadratic model we have five equations of the form:

$$y_i = b_2x_i^2 + b_1x_i + b_0 + \varepsilon_i$$

where ε_i is the difference between the observed value and the value predicted by the quadratic. As before we have a matrix equation of the form

$$Y = AX + \varepsilon(X)$$

where

$$A = \begin{bmatrix} 60^2 & 60 & 1 \\ 61^2 & 61 & 1 \\ 62^2 & 62 & 1 \\ 63^2 & 63 & 1 \\ 65^2 & 65 & 1 \end{bmatrix}, \quad X = \begin{bmatrix} b_2 \\ b_1 \\ b_0 \end{bmatrix}, \text{ and } Y = \begin{bmatrix} 3.1 \\ 3.6 \\ 3.8 \\ 4 \\ 4.1 \end{bmatrix}.$$

Again, we seek an $\hat{X}$ so that

$$\|Y = A\hat{X}\| \leq \|y - AX\| = \|\varepsilon(X)\| \quad (E(X) = \|\varepsilon(X)\|^2).$$

So we want to solve the consistent system

$$A^*A\hat{X} = A^*Y.$$

We have

$$A^*A = \begin{bmatrix} 75185763 & 1205981 & 19359 \\ 1205981 & 19359 & 311 \\ 19359 & 311 & 5 \end{bmatrix}, A^*Y = \begin{bmatrix} \frac{723613}{10} \\ \\ \frac{11597}{10} \\ \\ \frac{93}{5} \end{bmatrix}, \text{ and } \hat{X} = \begin{bmatrix} -\frac{141}{2716} \\ \\ \frac{90733}{13580} \\ \\ -\frac{715864}{3395} \end{bmatrix}.$$

So the quadratic is

$$y = -\frac{141}{2716}x^2 + \frac{90733}{13580}x - \frac{715864}{3395}.$$

The points together with the approximating quadratic are displayed below.

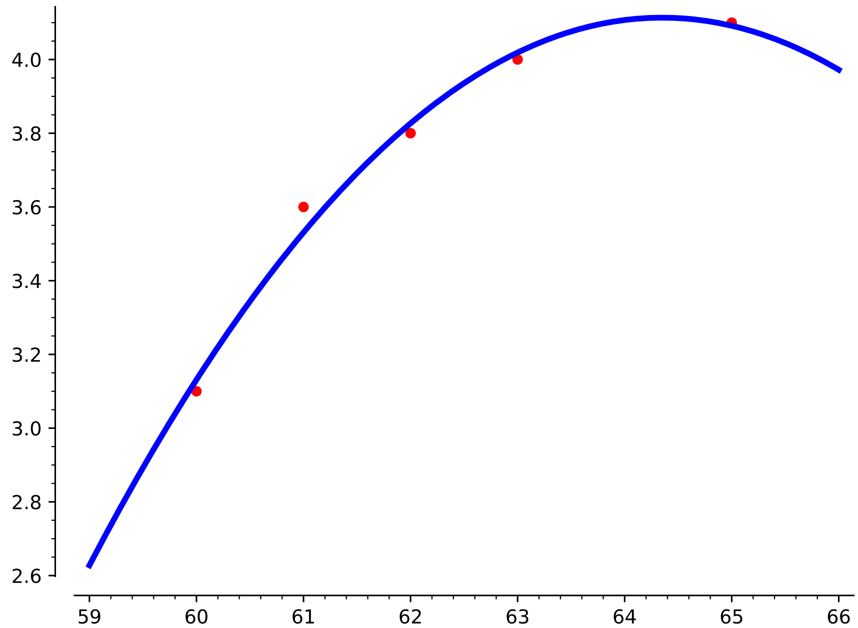

Figure 3.3.14 A quadratic least squares approximation

3.4 Diagonalization of matrices in Inner Product Spaces

We examine properties of a matrix in a inner product space which guarantee it is diagonalizable. We also lay the ground work for singular value decomposition of an arbitrary matrix.

In particular, we shall show that a real symmetric matrix and a complex unitary matrix can always be diagonalized.

While such a result is remarkable in and of itself since these properties must somehow guarantee that for such matrices each eigenspace has geometric multiplicity equal to its algebraic multiplicity, it leads us to discover an important result about the representation of any real or complex $m \times n$ matrix A. The key is that for any such matrix, both A^*A and AA^* are Hermitian matrices. What is even more interesting is that diagonalization of A^*A still tells us very important information about the original matrix A.

3.4.1 Some relations between A and A^*

Let's begin with some simple properties concerning the rank of a matrix.

Proposition 3.4.1 *Let A be an $m \times n$ matrix with entries in any field F.*

1. *Let P (resp. Q) be any invertible $m \times m$ (resp. $n \times n$) matrix with entries*

in F. Then

$$\operatorname{rank}(PAQ) = \operatorname{rank} A.$$

Equivalently, one can say that elementary row or column operations on a matrix do not change its rank.

2. $\operatorname{rank} A = \operatorname{rank} A^T$ *(i.e., row rank is equal to column rank).*

3. *If A has complex entries then* $\operatorname{rank} A = \operatorname{rank} A^*$.

Proof of (1). See Theorem 3.4 of [1]. ■

Proof of (2). Recall the the number of pivots is equal to the row and column rank, so consider the reduced row-echelon form of the matrix, noting that elementary row operations do not change the row space nor the dimension of the column space. ■

Proof of (3). The difference between A^* and A^T is simply that the entries of A^T have been replaced by their complex conjugates, so if there were a linear dependence among the rows of (say) A^*, conjugating that relation would produce a linear dependence among the rows of A^T. ■

3.4.2 A closer look at matrices A^*A and AA^*.

In Corollary 3.3.2, we have seen both of these products of matrices when the columns of A are orthonormal; one product producing an identity matrix, the other the matrix of the orthogonal projection into the column space of A. But what can we say in general (when the columns are not orthonormal vectors)?

Proposition 3.4.2 *Let A be any $m \times n$ matrix with real or complex entries. Then*

$$\operatorname{rank} A = \operatorname{rank}(A^*A) = \operatorname{rank}(AA^*).$$

Proof. We first show that $\operatorname{rank} A = \operatorname{rank}(A^*A)$. Since A is $m \times n$ and A^*A is $n \times n$, both matrices represent linear transformations from a domain of dimension n. As such, the rank-nullity theorem says that

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

We show that the two nullspaces (kernels) are equal, hence have the same dimension, and the statement about ranks will follow.
Since any linear map takes 0 to 0, it is clear that $\ker A \subseteq \ker A^*A$. Conversely, suppose that $x \in \ker A^*A$. Then $A^*Ax = 0$, hence $\langle x, A^*Ax\rangle = 0$, so by Proposition 3.2.16,

$$0 = \langle x, A^*Ax\rangle = \langle Ax, Ax\rangle$$

which implies $Ax = 0$ by the positivity of the inner product. Thus $\ker A^*A \subseteq \ker A$, giving us the desired equality.
To show that $\operatorname{rank} A = \operatorname{rank} AA^*$, we show equivalently (see Proposition 3.4.1) that $\operatorname{rank} A^* = \operatorname{rank} AA^*$. We showed above that for any matrix B, $\operatorname{rank} B =$

rank B^*B, so letting $B = A^*$, we conclude

$$\operatorname{rank} A^* = \operatorname{rank}((A^*)^*A^*) = \operatorname{rank}(AA^*).$$

■

Let us note another common property of AA^* and A^*A.

Proposition 3.4.3 *Let A be any $m \times n$ matrix with real or complex entries. Then the nonzero eigenvalues of A^*A and AA^* are the same. Note that zero may be an eigenvalue of one product, but not the other.*

Proof. This result is fairly general. Suppose that A, B are two matrices for which both products AB and BA are defined, and suppose that λ is a nonzero eigenvalue for AB. This implies there exists a nonzero vector v for which $ABv = \lambda v$. Multiplying both sides by B and noting multiplication by B is a linear map, we conclude that

$$(BA)Bv = \lambda(Bv),$$

which shows that λ is an eigenvalue of BA so long as $Bv \neq 0$ (eigenvectors need to be nonzero). But if $Bv = 0$, then $ABv = \lambda v = 0$ which implies $\lambda = 0$, contrary to assumption.
For the eigenvalue $\lambda = 0$, the situation can be (and often is) different. Let $A = \begin{bmatrix} 1 & 1 \end{bmatrix}$, and consider $B = A^T$. Then

$$AB = \begin{bmatrix} 2 \end{bmatrix} \text{ while } BA = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

The matrix AB is clearly non-singular, while the rank of BA is one, hence having a non-trivial nullspace. ■

Before proceeding, we need to make a few more definitions and raise one cautionary note. For the caution, observe that in general results we state for complex matrices A hold analogously for real matrices, replacing A^* by A^T. The Spectral Theorem for complex matrices and the Spectral Theorem for real matrices have distinctly different hypotheses, and we want to spend a bit of time explaining why.

While all the terms we list below are defined in the section on definitions, it is useful for comparison to list them explicitly here. Let $A \in M_n(\mathbb{C})$ and $B \in M_n(\mathbb{R})$

- A is **normal** if $AA^* = A^*A$.
- A is **unitary** if $AA^* = A^*A = I_n$
- A is **Hermitian** if $A = A^*$.
- B is **normal** if $BB^T = B^TB$.
- B is **orthogonal** if $BB^T = B^TB = I_n$

- B is **symmetric** if $B = B^T$.

Note that both Hermitian and unitary matrices are normal, though for example a Hermitian matrix is unitary only if $A^2 = I_n$. Analogous observations are true for real matrices. The point here is that the complex Spectral Theorem holds for the broad class of normal matrices, but the real Spectral Theorem holds only for the narrower class of real symmetric matrices. We still need to understand why.

We first consider some properties of real orthogonal matrices and complex unitary matrices.

Proposition 3.4.4 *Let $P \in M_n(\mathbb{R})$ (resp. $U \in M_n(\mathbb{C})$). The following statements are equivalent:*

1. *P is an orthogonal matrix (resp. U is a unitary matrix).*
2. *$P^T P = I_n$ (resp. $U^*U = I_n$).*
3. *$PP^T = I_n$ (resp. $UU^* = I_n$).*
4. *$P^{-1} = P^T$ (resp. $U^{-1} = U^*$)*
5. *$\langle Pv, Pw\rangle = \langle v, w\rangle$ for all $v, w \in \mathbb{R}^n$ (resp. $\langle Uv, Uw\rangle = \langle v, w\rangle$ for all $v, w \in \mathbb{C}^n$.)*

Proof. As a sample consider the case where $A^*A = I_n$. This says that A has a left inverse, but since A is a square matrix, it also has a right one and they are equal.
For the last statement, recall from Proposition 3.2.16 that for any matrix $A \in M_n(C)$,

$$\langle Av, w\rangle = \langle v, A^*w\rangle$$

for all $v, w \in \mathbb{C}^n$. It follows that for an orthogonal/unitary matrix

$$\langle Av, Aw\rangle = \langle v, A^*Aw\rangle = \langle v, w\rangle.$$

■

Below we state some simple versions of the spectral theorems.

Theorem 3.4.5 The Spectral Theorem for normal matrices. *If $A \in M_n(\mathbb{C})$ is a normal matrix, then there is a unitary matrix U and complex scalars $\lambda_1, \ldots, \lambda_n$ so that*

$$\mathrm{diag}(\lambda_1, \ldots, \lambda_n) = U^*AU.$$

In particular, any complex normal matrix can be unitarily diagonalized. The columns of U are eigenvectors for A and form an orthonormal basis for $\mathbb{C}^n$.

Theorem 3.4.6 The Spectral Theorem for real symmetric matrices. *If $A \in M_n(\mathbb{R})$ is a symmetric matrix, then there exists an orthogonal matrix P and* *real scalars* $\lambda_1, \dots, \lambda_n$ *so that*

$$\operatorname{diag}(\lambda_1, \dots, \lambda_n) = P^T AP.$$

In particular, any real symmetric matrix can be orthogonally diagonalized. The columns of P *are eigenvectors for* A *and form an orthonormal basis for* $\mathbb{R}^n$.

Remark 3.4.7 To gain some appreciation of why there is a difference in hypotheses between the real and complex versions of the spectral theorem, consider the matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, and note that A is orthogonal (hence normal), but not symmetric. One immediately checks that that characteristic polynomial of A, $\chi_A = x^2 + 1$, has no real roots which means A has no real eigenvalues so cannot possibly be diagonalized to say nothing of orthogonally diagonalized. Clearly one important element of the spectral theorem is that the characteristic polynomial must split completely (factor into all linear factors) over the field. This is given for the complex numbers since they are algebraically closed, but not so for the real numbers. So in the real case, we must somehow guarantee that a real symmetric matrix has only real eigenvalues.

We state the following proposition for complex Hermitian matrices, but it also applies to real symmetric matrices since for a real matrix, $A^T = A^*$. Also note that every real or complex matrix has all its eigenvalues in $\mathbb{C}$.

Proposition 3.4.8 *Let* A *be a complex Hermitian matrix, and* λ *an eigenvalue for* A. *Then* λ *is necessarily a real number.*

Proof. Let λ be an eigenvalue of A, and let v be an eigenvector for λ. Then

$$\lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Av, v\rangle = \langle v, A^* v\rangle = \langle v, Av\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle,$$

where we have used the Hermitian property ($A^* = A$) and the sesquilinearity of the inner product. Since $v \neq 0$, we know that $\langle v, v\rangle \neq 0$, from which we conclude $\overline{\lambda} = \lambda$, hence λ is real. ∎

Analogous to Proposition 1.5.5 for arbitrary matrices, we have

Proposition 3.4.9 *Let* $A \in M_n(\mathbb{C})$ *be Hermitian matrix. Then eigenspaces corresponding to distinct eigenvalues are orthogonal.*

Proof. Suppose that λ and μ are distinct eigenvalues for A. Let v be an eigenvector with eigenvalue λ and w be an eigenvector with eigenvalue μ. Then

$$\begin{aligned} \lambda\langle v, w\rangle = \langle \lambda v, w\rangle = \langle Av, w\rangle &\overset{(1)}{=} \langle v, A^* w\rangle \\ &\overset{(2)}{=} \langle v, Aw\rangle = \langle v, \mu w\rangle = \overline{\mu}\langle v, w\rangle \overset{(3)}{=} \mu\langle v, w\rangle, \end{aligned}$$

where (1) is true by Proposition 3.2.16, (2) is true since A is Hermitian, (3) is true by Proposition 3.4.8 and the remaining equalities hold using standard properties of the inner product. Rewriting the expression, we have

$$(\lambda - \mu)\langle v, w\rangle = 0,$$

and since $\lambda \neq \mu$, we conclude $\langle v, w\rangle = 0$ as desired. ■

The proof of the spectral theorems is rather involved. Of course any matrix over $\mathbb{C}$ will have the property that its characteristic polynomial splits, but we have also shown this for real symmetric matrices. The hard part is showing that each eigenspace has dimension equal to the algebraic multiplicity of the eigenvalue. For this something like Schur's theorem is used as a starting point. See Theorem 6.21 of [1].

We would like to use the spectral theorems to advance the proof of the singular value decomposition (SVD) of a matrix, though it is interesting to note that other authors do the reverse, see section 5.4 of [3].

Remark 3.4.10 We conclude this section with another interpretation of the spectral theorem, giving a spectral decomposition which will be mirrored in the next section on the singular value decomposition.

We restrict our attention to $n \times n$ matrices A over the real or complex number which are Hermitian (i.e., symmetric for a real matrix), and consequently for which all the eigenvalues are real. We list the eigenvalues $\lambda_1, \dots, \lambda_n$, though this does not mean they need be all distinct. By Theorem 3.4.5, there exists a unitary matrix U whose columns $\{u_1, \dots, u_n\}$ form an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors for A so that

$$\operatorname{diag}(\lambda_1, \dots, \lambda_n) = UAU^*.$$

In the discussion preceding Corollary 3.3.2 we used the column-row rule for matrix multiplication to show that

$$UU^* = u_1u_1^* + \cdots + u_nu_n^*$$

which is the orthogonal projection into the column space of A (all of $\mathbb{C}^n$ in this case), but viewed as the sum of one-dimensional orthogonal projections onto the spaces spanned by each u_i. It follows that

Proposition 3.4.11 Spectral decomposition of a Hermitian matrix. *Let $A \in M_n(\mathbb{C})$ be a Hermitian matrix with (real) eigenvalues $\lambda_1, \dots, \lambda_n$. Let U be the unitary matrix whose orthonormal columns u_i are eigenvectors for the λ_i. Then*

$$A = U \operatorname{diag}(\lambda_1, \dots, \lambda_n)U^* = \lambda_1u_1u_1^* + \cdots + \lambda_nu_nu_n^*.$$

Proof. By the spectral theorem,

$$\operatorname{diag}(\lambda_1, \dots, \lambda_n) = U^*AU$$

or

$$A = (U^*)^{-1} \operatorname{diag}(\lambda_1, \dots, \lambda_n)U^{-1} = U \operatorname{diag}(\lambda_1, \dots, \lambda_n)U^*,$$

since U is unitary, so $U^{-1} = U^*$, and the result follows. ■

3.5 Adjoint Maps and properties

In Proposition 3.2.16, we have seen how a complex $m \times n$ matrix and its conjugate transpose have a natural relation with respect to inner products, and in Subsection 3.2.5 took a first look at the four fundamental subspaces. In this section we develop the corresponding notions for linear maps between inner product spaces.

3.5.1 Basic Properties

Let V, W be inner product spaces and $T : V \to W$ be a linear map. We can ask if there exists a linear map $S : W \to V$ so that

$$\langle T(v), w\rangle_W = \langle v, S(w)\rangle_V.$$

Let's look at a few examples.

Example 3.5.1 $T(x) = Ax$**.** If A is an $m \times n$ complex matrix, then $T(x) = Ax$ defines a linear transformation $T : \mathbb{C}^n \to \mathbb{C}^m$. In Proposition 3.2.16, we saw that the linear map $S : \mathbb{C}^m \to \mathbb{C}^n$ given by $S(X) = A^*x$ satisfies the requisite property that

$$\langle T(v), w\rangle_W = \langle v, S(w)\rangle_V.$$

□

Example 3.5.2 Orthogonal Projections. Let V be an inner product space and W a subspace with orthonormal basis $\{w_1, \ldots, w_r\}$. As we have seen in Subsection 3.2.4, the orthogonal projection of V onto W is given by

$$\mathrm{proj}_W(v) = \langle v, w_1\rangle w_1 + \cdots + \langle v, w_r\rangle w_r.$$

By Theorem 3.2.10 and Corollary 3.2.14, the projection map satisfies $v^\perp := v - \mathrm{proj}_W(v) \in W^\perp$ and $\mathrm{proj}_W^2 = \mathrm{proj}_W$. We wish to show that the projection is **self-adjoint,** that is,

$$\langle \mathrm{proj}_W v, u\rangle = \langle v, \mathrm{proj}_W u\rangle.$$

□

Proof. To show that

$$\langle \mathrm{proj}_W v, u\rangle = \langle v, \mathrm{proj}_W u\rangle$$

for all $u, v \in V$, we write $v = \mathrm{proj}_W v + v^\perp$ and $u = \mathrm{proj}_W u + u^\perp$ (with $v^\perp, u^\perp \in W^\perp$). Then

$$\langle \mathrm{proj}_W v, u\rangle = \langle \mathrm{proj}_W v, \mathrm{proj}_W u\rangle + \langle \mathrm{proj}_W v, u^\perp\rangle = \langle \mathrm{proj}_W v, \mathrm{proj}_W u\rangle,$$

and

$$\langle v, \mathrm{proj}_W u\rangle = \langle \mathrm{proj}_W v, \mathrm{proj}_W u\rangle + \langle v^\perp, \mathrm{proj}_W u\rangle = \langle \mathrm{proj}_W v, \mathrm{proj}_W u\rangle$$

which establishes the equality. ■

Example 3.5.3 Hyperplane Reflections. Let V be a finite-dimensional inner product space, u a unit vector, and W the hyperplane (through the origin) normal to u. Geometrically, we want to reflect a vector v across the hyperplane W. One way to describe this is to write $v = \operatorname{proj}_W v + v^\perp$ and define $H(v) = \operatorname{proj}_W v - v^\perp$, but we recognize that $v^\perp = \langle v, u\rangle u$, so we can simply write

$$H(v) = v - 2\langle v, u\rangle u.$$

The map H is often called a **Householder transformation.** We show that it too is self-adjoint. □

Proof. As before, we compute both sides of the desired equality: $\langle Hv, z\rangle = \langle v, Hz\rangle$ and show they are equal.
One the one hand,

$$\langle Hv, z\rangle = \langle v - 2\langle v, u\rangle u, z\rangle = \langle v, z\rangle - 2\langle v, u\rangle\langle u, z\rangle.$$

On the other hand,

$$\langle v, Hz\rangle = \langle v, z - 2\langle z, u\rangle u\rangle = \langle v, z\rangle - 2\overline{\langle z, u\rangle}\langle v, u\rangle,$$

and since $\overline{\langle z, u\rangle} = \langle u, z\rangle$, we have the desired equality. ■

It is straightforward to show than if an adjoint exists, it is unique:

Proof. If for all $v \in V, w \in W$

$$\langle Tv, w\rangle = \langle v, Sw\rangle = \langle v, S'w\rangle,$$

then

$$\langle v, (S - S')w\rangle = 0$$

for all v, w, which implies $S = S'$. ■

We denote the unique adjoint of the linear map T as T^*. As a consequence of uniqueness it is immediate to check that

$$(\lambda T)^* = \overline{\lambda}T^*, (S + T)^* = S^* + T^*. \text{ and } (T^*)^* = T.$$

If V is a finite-dimensional inner product space, it is easy to show that every linear map $T : V \to W$ has an adjoint.

Proposition 3.5.4 *Let V be a finite-dimensional inner product space with orthonormal basis $\{e_1, \ldots, e_n\}$. Then the adjoint map $T^* : W \to V$ is linear and given by*

$$T^*(w) = \sum_{k=1}^{n} \langle w, T(e_k)\rangle e_k.$$

Proof. Recall that by Theorem 3.2.3, every vector $v \in V$ has a unique representation as $v = \sum_{k=1}^n \langle v, e_k\rangle e_k$. As a consequence,

$$\begin{aligned}\langle Tv, w\rangle &= \langle T(\sum_{k=1}^n \langle v, e_k\rangle e_k), w\rangle = \sum_{k=1}^n \langle v, e_k\rangle \langle T(e_k), w\rangle \\ &= \sum_{k=1}^n \overline{\langle w, T(e_k)\rangle}\langle v, e_k\rangle = \sum_{k=1}^n \langle v, \langle w, T(e_k) e_k\rangle e_k\rangle \\ &= \langle v, \sum_{k=1}^n \langle w, T(e_k)\rangle e_k\rangle = \langle v, T^*(w)\rangle.\end{aligned}$$

It follows from this definition and properties of the inner product that T^* is linear. ■

As a means of connecting this notion of adjoint with the properties of the conjugate transpose of a matrix given in Proposition 3.2.16, we have the following proposition.

Proposition 3.5.5 *Let V, W be finite-dimensional inner product spaces with orthonormal bases $\mathcal{B}_V$ and $\mathcal{B}_W$. Then the matrix of the adjoint T^* of a linear map $T : V \to W$ is the conjugate transpose of the matrix of T, precisely*

$$[T^*]_{\mathcal{B}_W}^{\mathcal{B}_V} = \left([T]_{\mathcal{B}_V}^{\mathcal{B}_W}\right)^*.$$

Proof. Let the orthonormal bases be given by $\mathcal{B}_V = \{e_1, \ldots, e_n\}$ and $\mathcal{B}_W = \{f_1, \ldots, f_m\}$. If $A = [T]_{\mathcal{B}_V}^{\mathcal{B}_W}$ and $B = [T^*]_{\mathcal{B}_W}^{\mathcal{B}_V}$ then by Theorem 3.2.3, $B_{ij} = \langle T^*(f_j), e_i\rangle$ and

$$A_{ij} = \langle T(e_j), f_i\rangle = \langle e_j, T^*(f_i)\rangle = \overline{\langle T^*(f_i), e_j\rangle} = \overline{B_{ji}}$$

which establishes the result. ■

3.5.2 A second look at the four fundamental subspaces

In the previous section, we established the existence and uniqueness of the adjoint of a linear map defined on a finite-dimensional inner product space, and connections with the matrix of the linear transformation. Here we look at a few more properties including a second look at the four fundamental subspaces which lie at the heart of the singular value decomposition.

A linear operator $T : V \to V$ on an inner product space is called **self-adjoint** or **Hermitian** if $T^* = T$. We saw that both the Householder transformation and the orthogonal projection were examples of self-adjoint operators.

Proposition 3.5.6 *Suppose that U, V, W are finite-dimensional inner product spaces, and $S : U \to V$ and $T : V \to W$ are linear maps. Then*

- $S^* T^* = (TS)^*$
- *In particular, if T is invertible, then so is T^*, and $(T^*)^{-1} = (T^{-1})^*$.*

Proof. For all $u \in U$ and $w \in W$ we have

$$\langle u, S^*T^*w\rangle = \langle Su, T^*w\rangle = \langle TS(u), w\rangle = \langle u, (TS)^*w\rangle$$

which yields $S^*T^* = (TS)^*$.
For the second, it is immediate that the adjoint of the identity map is again an identity map. As a consequence (and using the first part),

$$(T^{-1}T)^* = (Id_V)^* = Id_V = T^*(T^{-1})^*,$$

which together with the identity with the operators reversed gives the result. Actually, being operators on finite dimensional vector spaces, one such identity yields the other by rank-nullity. ∎

Another important class of linear map between inner product spaces is the notion of an **isometry**, a linear map $T : V \to W$ which satisfies

$$\langle Tv_1, Tv_2\rangle_W = \langle v_1, v_2\rangle_V$$

for all $v_1, v_2 \in V$.

Proposition 3.5.7 *If $T : V \to W$ is an isometry, then T is injective. Moreover, if T is surjective, then $T^* = T^{-1}$.*

Proof. To show that T is injective, we note that the kernel is trivial: If $T(v) = 0$, then

$$\langle T(v), T(v)\rangle = \langle v, v\rangle = 0$$

which can happen if and only if $v = 0$.
Now suppose that T is surjective. Let $v \in V$ and $w \in W$ be arbitrary. Choose $v' \in V$ with $T(v') = w$. Then

$$\langle T(v), w\rangle = \langle T(v), T(v')\rangle = \langle v, v'\rangle = \langle v, T^{-1}(w)\rangle,$$

so by uniqueness of the adjoint, $T^* = T^{-1}$. ∎

The following theorem should be compared to Theorem 3.2.17 and its corollary.

Theorem 3.5.8 *Let V, W be finite dimensional inner product spaces and $T : V \to W$ a linear map. Then*

- $\ker T^* = (\operatorname{Im} T)^\perp$,
- $\operatorname{Im}(T^*) = (\ker T)^\perp$.

Proof. Let $w \in \ker(T^*)$. Then $T^*(w) = 0$, so $\langle v, T^*(w)\rangle = 0 = \langle T(v), w$ for all $v \in V$. Thus w is orthogonal to the image of T, i.e., $\ker(T^*) \subseteq (\operatorname{Im} T)^\perp$. Conversely, if $w \in (\operatorname{Im} T)^\perp$, then for all $v \in V$,

$$0 = \langle T(v), w\rangle = \langle v, T^*(w)\rangle.$$

In particular, choosing $v = T^*(w)$ shows that $T^*(w) = 0$, hence $(\operatorname{Im} T)^\perp \subseteq$

$\ker T^*$, giving the first equality.
Since the first statement is true for any linear map and finite-dimensional inner product spaces, we replace T by T^* and use $(T^*)^* = T$ to conclude

$$\ker T = (\operatorname{Im} T^*)^\perp.$$

Finally, using Corollary 3.2.12 yields

$$(\ker T)^\perp = \operatorname{Im} T^*.$$

■

3.6 Singular Value Decomposition

We show how the spectral decomposition for Hermitian matrices gives rise to an analogous, but very special decomposition for an arbitrary matrix, called the **singular value decomposition (SVD).**

We shall state without proof the version of the SVD which holds for linear transformations between finite-dimensional inner product spaces, but its statement is so elegant, it's depth of importance is almost lost.

Then we state and prove the matrix version, providing some examples to demonstrate its utility.

3.6.1 SVD for linear maps

We begin with a statement of the singular value decomposition for linear maps as paraphrased from Theorem 6.26 of [1].

Theorem 3.6.1 *Let V, W be finite-dimensional inner product spaces and $T : V \to W$ a linear map having rank r. Then there exist orthonormal bases $\{v_1, \ldots, v_n\}$ for V and $\{w_1, \ldots, w_m\}$ for W, and positive scalars $\sigma_1 \geq \cdots \geq \sigma_r$ so that*

$$T(v_i) = \begin{cases} \sigma_i u_i & \text{if } 1 \leq i \leq r \\ 0 & \text{if } i > r. \end{cases}$$

*Moreover, the σ_i are uniquely determined by T, and are called the **singular values** of T.*

Another way to state the main part of this result is the if the orthonormal bases are $\mathcal{B}_V = \{v_1, \ldots, v_n\}$ and $\mathcal{B}_W = \{w_1, \ldots, w_m\}$, then the matrix of T with respect to these bases has the form

$$[T]_{\mathcal{B}_V}^{\mathcal{B}_W} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

where $D = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}$ and the lower right block of zeros of $[T]_{\mathcal{B}_V}^{\mathcal{B}_W}$ has size $(m-r) \times (n-r)$.

Remark 3.6.2 Staring at the form of the matrix above, does it really seem all that special or new? Indeed, we know that given an $m \times n$ matrix A, we can perform elementary row and column operations on A, represented by invertible matrices P, Q so that

$$PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Now the matrices P, Q just represent a change of basis as happens in the theorem. Specifically (and assuming for convenience of notation that $V = \mathbb{C}^n$, $W = \mathbb{C}^m$ with standard bases $\mathcal{E}_n$ and $\mathcal{E}_m$), the matrices P and Q give rise to bases $\mathcal{B}_V$ for V, and $\mathcal{B}_W$ for W, so that

$$PAQ = [I]_{\mathcal{E}_m}^{\mathcal{B}_W}[T]_{\mathcal{E}_n}^{\mathcal{E}_m}[I]_{\mathcal{B}_V}^{\mathcal{E}_n} = [T]_{\mathcal{B}_V}^{\mathcal{B}_W} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

so that with the obvious enumeration of the bases, the map T acts by $v_i \mapsto 1 \cdot w_i$ for $1 \le i \le r$, and $v_i \mapsto 0$ for $i > r$.

But then we look a bit more carefully. The elementary row and column operations just hand us new bases with no special properties. We could make both bases orthogonal via Gram-Schmidt, but then would have no hope that $v_i \mapsto 1 \cdot w_i$ for $1 \le i \le r$, and $v_i \mapsto 0$ for $i > r$. In addition, we know that orthogonal and unitary matrices are very special since they preserve inner products, so the geometric transformations that are taking place in $\mathbb{C}^n$ and $\mathbb{C}^m$ are rigid motions with the only stretching effect given by the singular values. In other words, there is actually a great deal going on in this theorem which we shall now explore.

3.6.2 SVD for matrices

We begin with an arbitrary $m \times n$ matrix A with complex entries. Let $B = A^*A$, and note that $B^* = B$, so B is an $n \times n$ Hermitian matrix and the Spectral Theorem implies that there is an orthonormal basis for $\mathbb{C}^n$ consisting of eigenvectors for $B = A^*A$ having (not necessarily distinct) eigenvalues $\lambda_1, \dots, \lambda_n$.

We have already seen in Proposition 3.4.8 that Hermitian matrices have real eigenvalues, but we can say more for A^*A. Using the eigenvectors and eigenvalues from above, we compute:

$$\begin{aligned} \|Av_i\|^2 &= \langle Av_i, Av_i \rangle = (Av_i)^*(Av_i) = v_i^* A^* A v_i \\ &= v_i^*(A^*A)v_i \overset{(1)}{=} v_i^* \lambda_i v_i \overset{(2)}{=} \lambda_i v_i^* v_i \overset{(3)}{=} \lambda_i, \end{aligned}$$

where (1) is true since v_i is an eigenvector for A^*A, (2) is true since in a vector space scalars commute with vectors, and (3) is true since the v_i are unit vectors.

Thus in addition to the eigenvalues of A^*A being real numbers, the computation shows that they are *non-negative* real numbers.

We let $\sigma_i = \sqrt{\lambda_i}$. The σ_i are called the **singular values of** A, and from the computation above, we see that

$$\sigma_i = \|Av_i\|.$$

We may assume that the eigenvalues are labeled in such a way that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

Usually we introduce the notation that $\sigma_1, \ldots, \sigma_r > 0$, and $\sigma_i = 0$ for $i > r$. We shall show now that $r = \operatorname{rank} A$ so that $r = n$ if and only if $\operatorname{rank} A = n$.

Proposition 3.6.3 *The number of positive singular values of a matrix A equals its rank.*

Proof 1. Proposition 3.4.3 shows that $\operatorname{rank} A = \operatorname{rank} A^*A$, so we need only show that $r = \operatorname{rank} A^*A$. Now recall that $\{v_1, \ldots, v_n\}$ is an orthonormal basis of $\mathbb{C}^n$ consisting is eigenvectors for A^*A.
Now the rank of A^*A is the dimension of its range, the number of linearly independent vectors in $\{A^*Av_1, \ldots, A^*Av_n\}$, and the rank-nullity theorem says that since $A^*Av_i = 0$ for $i > r$, we know the nullity is at least $(n-r)$ and the rank at most r. We need only show that $\{A^*Av_1, \ldots, A^*Av_r\}$ is a linearly independent set to guarantee the rank is r.
Suppose that

$$\sum_{i=1}^{r} \alpha_i A^*Av_i = 0.$$

Since the v_i are eigenvectors for A^*A, we deduce

$$\sum_{i=1}^{r} \alpha_i \lambda_i v_i = 0,$$

and since the v_i are themselves linearly independent, each coefficient $\alpha_i\lambda_i = 0$. Since we are assuming that $\lambda_i > 0$ for $i = 1, \ldots, r$, we conclude all the $\alpha_i = 0$, making the desired set linearly independent, which establishes the result. ∎

Proof 2. A slightly more direct proof that $r = \operatorname{rank} A$ begins by recalling that $\sigma_i = \|Av_i\|$, so we know that $Av_i = 0$ for $i > r$. Again by rank-nullity, we deduce the rank is at most r and precisely is the number of linearly independent vectors in $\{Av_1, \ldots, Av_r\}$. In fact, we show that this is an orthogonal set of vectors, so linearly independent by Proposition 3.2.2. Since $\{v_1, \ldots, v_n\}$ is an orthonormal set of vectors, for $j \neq k$ we know that v_j and $\lambda_k v_k$ are orthogonal. We compute

$$\langle Av_k, Av_j\rangle = (Av_j)^*(Av_k) = v_j^* A^*Av_k = v_j^*(\lambda_k v_k) = \lambda_k\langle v_k, v_j\rangle = 0,$$

which gives the result. ∎

We summarize what is implicit in the second proof given above.

Corollary 3.6.4 *Suppose that $\{v_1, \ldots, v_n\}$ is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors for A^*A arranged so that the corresponding eigenvalues satisfy $\lambda_1 \geq \cdots \geq \lambda_n$. Further suppose that A has r nonzero singular values. Then $\{Av_1, \ldots, Av_r\}$ is an orthogonal basis for the column space of A, hence* $\operatorname{rank} A = r$.

We are now only a few steps away from our main theorem:

Theorem 3.6.5 *Let $A \in M_{m\times n}(\mathbb{C})$ with rank r and having singular values $\sigma_1 \geq \cdots \geq \sigma_n$. Then there exists an $m \times n$ matrix*

$$\Sigma = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \text{ where } D = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}$$

and unitary matrices $U \in M_m(\mathbb{C})$ and $V \in M_n(\mathbb{C})$, so that

$$A = U\Sigma V^*.$$

Proof. Given A, we construct an orthonormal basis of $\mathbb{C}^n$, $\{v_1, \ldots, v_n\}$, consisting of eigenvectors for A^*A arranged so that the corresponding eigenvalues satisfy $\lambda_1 \geq \cdots \geq \lambda_n$. Note that the matrix $V = [v_1\ v_2\ \cdots\ v_n]$ with the v_i as its columns is a unitary matrix.
By Corollary 3.6.4, we know that $\{Av_1, \ldots. Av_r\}$ is an orthogonal basis for the column space of A and we have observed that $\sigma_i = \|Av_i\|$, so let

$$u_i = \frac{1}{\sigma_i} Av_i,\ i = 1, \ldots, r.$$

Then $\{u_1, \ldots, u_r\}$ is an orthonormal basis for the column space of A which we extend to an orthonormal basis $\{u_1, \ldots, u_m\}$ of $\mathbb{C}^m$. We let U be the unitary matrix with orthonormal columns u_i. We now claim that

$$A = U\Sigma V^*$$

where

$$\Sigma = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \text{ and } D = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}.$$

Note that

$$AV = [Av_1\ \cdots\ Av_n] = [Av_1\ \cdots\ Av_r\ \mathbf{0}\ \cdots\ \mathbf{0}] = [\sigma_1 u_1\ \cdots\ \sigma_r u_r\ \mathbf{0} \cdots\ \mathbf{0}]$$

and also that

$$U\Sigma = [u_1, \ldots\ u_m] = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & \mathbf{0} \\ 0 & & \sigma_r & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right] = [\sigma_1 u_1\ \ldots \sigma_r u_r\ \mathbf{0}\ \ldots\ \mathbf{0}]$$

Thus

$$AV = U\Sigma,$$

and since V is a unitary matrix, multiplying both sides of the above equation on the right by $V^{-1} = V^*$ yields

$$A = U\Sigma V^*.$$

■

Remark 3.6.6 In complete analogy with Proposition 3.4.11, we have a spectral-like decomposition of A :

$$A = U\Sigma V^* = \sigma_1 u_1 v_1^* + \cdots + \sigma_r u_r v_r^*. \tag{3.6.1}$$

Remark 3.6.7 A few things to notice about the SVD. First, let's pause to note how the linear maps version of the SVD is implicit in what we have done above. We constructed an orthonormal basis $\{v_1, \ldots, v_n\}$ of $\mathbb{C}^n$, defined $\sigma_i = \|Av_i\|$, and for $i \leq r = \operatorname{rank} A$ set $u_= \dfrac{1}{\sigma_i} v_i$. We noted $\{u_1, \ldots, u_r\}$ is an orthonormal subset of $\mathbb{C}^m$ which we extended to an orthonormal subset of $\mathbb{C}^m$. So just from what we have seen above, we have

$$Av_i = \begin{cases} \sigma_i u_i & \text{if } 1 \leq i \leq r \\ 0 & \text{if } i > r. \end{cases}.$$

What we shall see below is even more remarkable in that there is a duality between A and A^*. We shall see that with the same bases and σ_i,

$$A^* u_i = \begin{cases} \sigma_i v_i & \text{if } 1 \leq i \leq r \\ 0 & \text{if } i > r. \end{cases}.$$

There are some other important and useful things to notice about the construction of the SVD. First is that matrices U, V are not uniquely determined though the singular values are. In light of this, a matrix can have many singular value decompositions all of equal utility.

Perhaps more interesting from a computational perspective and evident from Equation (3.6.1) is that adding the vectors $u_{r+1}, \ldots, u_m$ to form an orthonormal basis of $\mathbb{C}^m$ is completely unnecessary in practice. One only uses $u_1, \ldots, u_r$.

Now we are in desperate need of some examples. Let's start with computing an SVD of $A = \begin{bmatrix} 7 & 1 \\ 5 & 5 \\ 0 & 0 \end{bmatrix}$.

Example 3.6.8 A computation of a simple SVD. The process of computing an SVD is very algorithmic, and we follow the steps of the proof.

Let A be the 3×2 matrix $A = \begin{bmatrix} 7 & 1 \\ 5 & 5 \\ 0 & 0 \end{bmatrix}$. Then $A^*A = A^TA$ is 2×2, so in the notation of the theorem, $m = 3$ and $n = 2$. It is also evident from inspecting A that it has rank $r = 2$, so in this case we will have $r = n = 2$, so the general form of Σ will be "degenerate" with the last $n - r$ columns missing.

We compute

$$A^*A = \begin{bmatrix} 74 & 32 \\ 32 & 26 \end{bmatrix}$$

which has characteristic polynomial $\chi_A = (x - 10)(x - 90)$. The singular values are $\sigma_1 = \sqrt{90} \geq \sigma_2 = \sqrt{10}$. Thus the matrix Σ has the form

$$\Sigma = \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & \sqrt{10} \\ 0 & 0 \end{bmatrix}.$$

It follows from the spectral theorem that eigenspaces of a Hermitian matrix associated to different eigenvalues are orthogonal, so we can find any unit vectors v_1, v_2 which span the one-dimensional eigenspaces, and together they will form an orthonormal basis for $\mathbb{C}^2$. We compute eigenvectors for A^*A by row reducing $A^*A - \lambda_i I_2$, and obtain:

$$\text{Eigenvectors} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\} \mapsto \{v_1, v_2\} = \left\{ \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \right\}.$$

$$\text{So, } V = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}.$$

For the matrix U, we first look at

$$Av_1 = \begin{bmatrix} 7 & 1 \\ 5 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 15 \\ 15 \\ 0 \end{bmatrix} \mapsto u_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix},$$

$$Av_2 = \begin{bmatrix} 7 & 1 \\ 5 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -5 \\ 5 \\ 0 \end{bmatrix} \mapsto u_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}.$$

Finally we extend the orthonormal set $\{u_1, u_2\}$ to an orthonormal basis for $\mathbb{C}^3$, say $\{u_1, u_2, u_3\} = \left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$.

Then with $U = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $V = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$, $\Sigma = \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & \sqrt{10} \\ 0 & 0 \end{bmatrix}$, we have

$$A = \begin{bmatrix} 7 & 1 \\ 5 & 5 \\ 0 & 0 \end{bmatrix} = U\Sigma V^*$$

$$= \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & \sqrt{10} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}.$$

$\square$

Having computed the SVD:

$$A = \begin{bmatrix} 7 & 1 \\ 5 & 5 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & \sqrt{10} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix},$$

let's see how Equation (3.6.1) is rendered:

$$\begin{aligned} A &= U\Sigma V^* = \sigma_1 u_1 v_1^* + \cdots + \sigma_r u_r v_r^* \\ &= \sqrt{90} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} + \sqrt{10} \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \\ &= \sqrt{90} \begin{bmatrix} 2/\sqrt{10} & 1/\sqrt{10} \\ 2/\sqrt{10} & 1/\sqrt{10} \\ 0 & 0 \end{bmatrix} + \sqrt{10} \begin{bmatrix} 1/\sqrt{10} & -2/\sqrt{10} \\ -1/\sqrt{10} & 2/\sqrt{10} \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 3 \\ 6 & 3 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 7 & 1 \\ 5 & 5 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

We shall explore the significance of this kind of decomposition when we look at an application of the SVD to image compression.

Before that, let's summarize creating an SVD algorithmically, and then take a look at what the decomposition can tell us.

3.6.3 An algorithm for producing an SVD

Given an $m \times n$ matrix with real or complex entries, we want to write $A = U\Sigma V^*$, where U, V are appropriately sized unitary matrices (orthogonal if A has all real entries), and Σ is a block matrix which encodes the singular values of A. We proceed as follows:

1. The matrix A^*A is $n \times n$ and Hermitian (resp. symmetric if A is real), so it can be unitarily (resp. orthogonally) diagonalized. So find an orthonormal basis of eigenvectors $\{v_1, \ldots, v_n\}$ of A^*A labeled in such a way that the corresponding (real) eigenvalues satisfy $\lambda_1 \geq \cdots \geq \lambda_n$. Set V to be the matrix whose columns are the v_i. Then we know that

$$\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = V^*(A^*A)V.$$

This step probably involves the most work. It involves finding the characteristic polynomial of A^*A, and for each eigenvalue λ, finding a basis for the eigenspace for λ (i.e., the nullspace of $(A^*A - \lambda I_n)$), then using Gram-Schmidt to produce an orthogonal basis for the eigenspace, and finally normalizing to produce unit vectors. Note that by Proposition 3.4.9, eigenspaces corresponding to different eigenvalues of a Hermitian matrix are automatically orthogonal, so working on each eigenspace separately will produce the desired basis.

We shall review how to use Sage to help with some of these computations in the section below.

2. Let $\sigma_i = \sqrt{\lambda_i}$ and assume $\sigma_1 \geq \cdots \geq \sigma_r > 0$, $\sigma_{r+1} = \cdots = \sigma_n = 0$, knowing that it is possible for r to equal n.

3. Remember that $\{Av_1, \ldots, Av_r\}$ is an orthogonal basis for the column space of A, so in particular, $r = \operatorname{rank} A$. Normalize that set via $u_i = \dfrac{1}{\sigma_i} Av_i$ and complete to an orthonormal basis $\{u_1, \ldots, u_r, \ldots, u_m\}$ of $\mathbb{C}^m$. Put $U = [u_1 \ \ldots \ u_m]$, the matrix with the u_i as column vectors.

4. Then
$$A = U\Sigma V^* = U \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & \mathbf{0}_{r\times(n-r)} \\ 0 & & \sigma_r & \\ \hline & \mathbf{0}_{(m-r)\times r} & & \mathbf{0}_{(m-r)\times(n-r)} \end{array}\right] V^*.$$

3.6.4 Can an SVD for a matrix A be computed from AA^* instead?

This is a very important question, but why? Well, suppose that A is $m \times n$. Then A^*A is $n \times n$ while AA^* is $m \times m$, but both matrices are Hermitian and the first step of the SVD algorithm is to unitarily diagonalize a Hermitian matrix. If m and n differ in size, it would be nice to do the hard work on the smaller matrix. But we really did develop our algorithm based on using A^*A, so let's see if we can figure out how to use AA^* instead.

We know that using the Hermitian matrix A^*A, we deduce an SVD of the form
$$A = U\Sigma V^* = U \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & \mathbf{0}_{r\times(n-r)} \\ 0 & & \sigma_r & \\ \hline & \mathbf{0}_{(m-r)\times r} & & \mathbf{0}_{(m-r)\times(n-r)} \end{array}\right] V^*,$$

with U, V unitary matrices. It follows that

$$A^* = V\Sigma^* U^* = V \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & \mathbf{0}_{r\times(m-r)} \\ 0 & & \sigma_r & \\ \hline & \mathbf{0}_{(n-r)\times r} & & \mathbf{0}_{(n-r)\times(m-r)} \end{array}\right] U^*$$

where we note that upper $r \times r$ block of Σ^* is the same as that of Σ since the only nonzero entries in Σ are on the diagonal and are real.

Recall from Proposition 3.4.3 that the nonzero eigenvalues of A^*A and AA^* are the same, which means the singular values (and hence the matrix Σ or Σ^*) can be determined from either A^*A or AA^*. Also both U and V are unitary matrices which means that

$$A^* = V\Sigma^* U^*$$

is a singular value decomposition for A^*.

More precisely, if we put $B = A^*$ and compute an SVD for B, our algorithm would have us start with the matrix $B^*B = AA^*$, and we would deduce something like

$$B = A^* = U_1 \Sigma_1 V_1^*.$$

Taking conjugate transposes would give

$$A = V_1 \Sigma_1^* U_1^*,$$

providing an SVD for A.

Example 3.6.9 Compute an SVD for a 2×3 matrix. To compute an SVD for $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$, we have two choices: work with A^*A which is 3×3 or work with AA^* which is 2×2. Since we are doing the work by hand, we choose the smaller example, but remember that in working with AA^* we are computing an SVD for $B = A^*$ and will have to reinterpet as above.

We check that $B^*B = AA^* = \begin{bmatrix} 14 & 10 \\ 10 & 14 \end{bmatrix}$, which has characteristic polynomial

$$\begin{vmatrix} 14 - x & 10 \\ 10 & 14 - x \end{vmatrix} = (14 - x)^2 - 100 = (x - 4)(x - 24).$$

So ordered in descending order, we have

$$\lambda_1 = 24 \geq \lambda_2 = 4, \text{ (and) } \sigma_1 = 2\sqrt{6} \geq \sigma_2 = 2$$

so the rank $r = 2$. It is easy to see that

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

are corresponding eigenvectors which we normalize as

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

We now compute

$$u_1 = \frac{1}{\sigma_1} B v_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ and } u_2 = \frac{1}{\sigma_2} B v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix},$$

which we complete to an orthonormal basis for $\mathbb{C}^3$ with $u_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$.

Thus if we put

$$U = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \\ \sqrt{2} & \sqrt{3} & 1 \end{bmatrix} \text{ and } V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

we check that

$$B = U \begin{bmatrix} 2\sqrt{6} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} V^* = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \\ \sqrt{2} & \sqrt{3} & 1 \end{bmatrix} \begin{bmatrix} 2\sqrt{6} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

so that

$$A = B^* = V\Sigma^* U^* = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2\sqrt{6} & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ -\sqrt{3} & 0 & \sqrt{3} \\ 1 & -2 & 1 \end{bmatrix}.$$

□

3.6.5 Some Sage computations for an SVD

In the example above we computed an SVD for $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$ by computing an SVD for $A^* = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix}$ and converting, but since all the explicit work is for $B = A^*$, we do our Sage examples using that matrix.

First we parallel our computations done above following our algorithm, and then we switch to using Sage's builtin SVD algorithm. Why waste time if we can just go to the answer? It is probably better to judge for yourself.

First off, set up for pretty output and square brackets for delimeters, a style choice. Next, enter and print the matrix B.

```
%display latex
latex.matrix_delimiters("[", "]")
B=matrix(QQbar,[[1,3],[2,2],[3,1]])
```

```
B
```

Form $C = B^*B$, our Hermitian matrix.

```
C=B.conjugate_transpose()*B;C
```

Find the characteristic polynomial of B^*B, and factor it. Remember that all the eigenvalues are guaranteed to be real and the eigenspaces will have dimension equal to the algebraic multiplicities.

```
C.characteristic_polynomial().factor()
```

Ask Sage to give us the eigenvectors which, when normalized, will form the columns of the matrix V. The output of the `eigenmatrix_right()` command is a pair, the first entry is the diagonalized matrix, and the second the matrix whose columns are the corresponding eigenvectors. It is useful to see both so as to be sure the eigenvectors are listed in descending order of eigenvalues. Ours are fine, so we let V be the matrix of (unnormalized) eigenvectors.

```
C.eigenmatrix_right()
```

Next we grab the second entry in the above pair, the matrix of eigenvectors.

```
V=C.eigenmatrix_right()[1]
V
```

Now we normalize the column vectors:

```
for j in range(V.ncols()):
    w=V.column(j)
    if w.norm() != 0 :
            V[:,j] = w/w.norm()
V
```

Next we think about the U matrix. Technically, we have the orthonormal vectors v_i and need to find Bv_i and normalize them by dividing by $\sigma_i = \sqrt{\lambda_i}$. However, especially if doing pieces of the computation by hand so as to produce exact arithmetic, we can simply apply B to the unnormalized eigenvectors and normalize the result since the arithmetic (which we perform by hand) will be prettier.

We already have eigenvectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for $\lambda_1 = 24$, and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ for eigenvalue $\lambda_2 = 4$, but we need to apply B to them, normalize the result and complete that set to an orthogonal basis for $\mathbb{C}^3$. So we fast forward and have two orthogonal vectors $Bv_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $Bv_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. How do we find a vector orthogonal to the given two?

The orthogonal bit is easy; we can Gram-Schmidt our way to an orthogonal basis, but first we should choose a vector not in the span of the first two. Again since we have a small example, this is easy, but the method we suggest is to build a matrix with the first two ***rows*** the given orthogonal vectors, add a (reasonable) third row and ask for the rank. Really, we need only invoke Gram-Schmidt, and either we will have a third orthogonal vector or only the original two. We show what happens in both cases.

We build a container for the orthogonal vectors.

```
%display latex
latex.matrix_delimiters("[", "]")
D= matrix(QQbar,[[1,1,1],[-1,0,1],[0,0,0]]);D
```

First we add a row we know to be in the span of the first two; it is the sum of the first two, and Gram-Schmidt kicks it out.

```
D[2]=[0,1,2];D
```

We see Gram-Schmidt knew the third row was in the span of the first two.

```
G,M=D.gram_schmidt();G
```

Then we add a more reasonable row, and Gram-Schmidts produces an orthogonal basis.

```
D[2]=[1,0,2];D
```

```
G,M=D.gram_schmidt();G
```

To produce the orthogonal (unitary) matrix U, we must normalize the vectors and take the transpose to have the vectors as columns.

Sage also has the ability to compute an SVD directly once the entries of the matrix have been converted to RDF or CDF (Real or Complex double precision). This conversion can be done on the fly or by direct definition; we show both methods. The algorithm outputs the triple (U, Σ, V).

```
B=matrix(QQ,[[1,3],[2,2],[3,1]])
B,B.change_ring(RDF).SVD()
```

```
B=matrix(RDF,[[1,3],[2,2],[3,1]])
B,B.SVD()
```

3.6.6 Deductions from seeing an SVD

Suppose that A is a 2×3 real matrix and that A has the singular value decomposition

$$A = U\Sigma V^* = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \end{bmatrix}.$$

Question 3.6.10 What is the rank of A?

Answer. This is too easy. The rank r is the number of nonzero singular values, so $\operatorname{rank} A = 1$. □

Question 3.6.11 What is a basis for the column space of A?

Answer. Recall that $\{Av_1, \ldots, Av_r\}$ is a basis for the column space of A, and normalized, those vectors are $u_1, \ldots, u_r$, the first r columns of U. Since $r = 1$, the set $\{u_1\} = \left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}$ is a basis. □

Question 3.6.12 What is a basis for the kernel (nullspace) of A?

Answer. Hmmm. A bit trickier, or is it? The matrix A is 2×3, meaning the linear map L_A defined by $L_A(x) = Ax$ is a map from $\mathbb{C}^3 \to \mathbb{C}^2$. By rank-nullity, we deduce that nullity $A = 2$, and how conviently (recall the singular values), we have $Av_2 = Av_3 = 0$, which means $\{v_2, v_3\}$ is a basis for the nullspace. □

3.6.7 SVD and image processing

Matlab was used to render a (personal) photograph into a matrix whose entries are the gray-scale values 0-255 (black to white) of the corresponding pixels in the original jpeg image. The photo-rendered matrix A has size 2216×1463, and most likely is not something we want to treat by hand, but that is what computers are for.

But suppose I hand the matrix A to some nice software and it returns an SVD for A, say

$$A = U\Sigma V^* = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots + \sigma_r u_r v_r^*.$$

Recall that $\sigma_1 \geq \sigma_2 \geq \cdots$, so that the most significant features of the image (matrix) are conveyed by the early summands $\sigma_i u_i v_i^*$ each of which is an $m \times n$ matrix of rank 1. Now it turns out that the rank of our matrix A is $r = 1463$, so that is a long sum. What is impressive about the SVD is how quickly the early partial sums reveal the majority of the critical features we seek to infer.

So let's take a look at the renderings of some of these partial sums recalling that it takes 1463 summands to recover the original jpeg image.

Here is the rendering of the first summand. Notice how all the rows (and columns) are multiples of each other reflecting that the matrix corresponding to this image has rank 1.

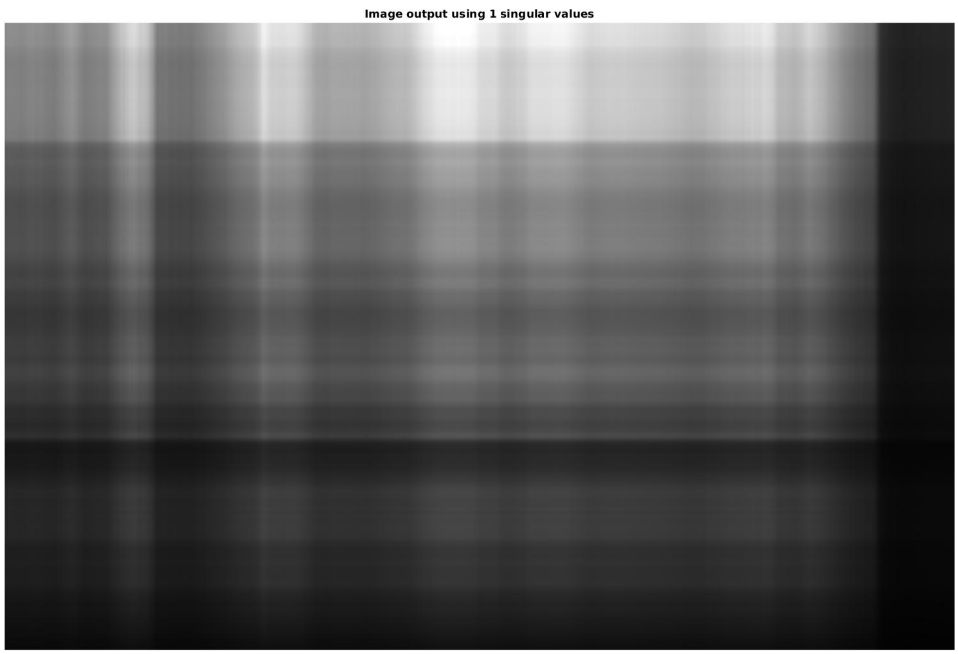

Figure 3.6.13 Image output from first summand of SVD

Here is the rendering of the partial sum of the three summands. Hard to know what this image is.

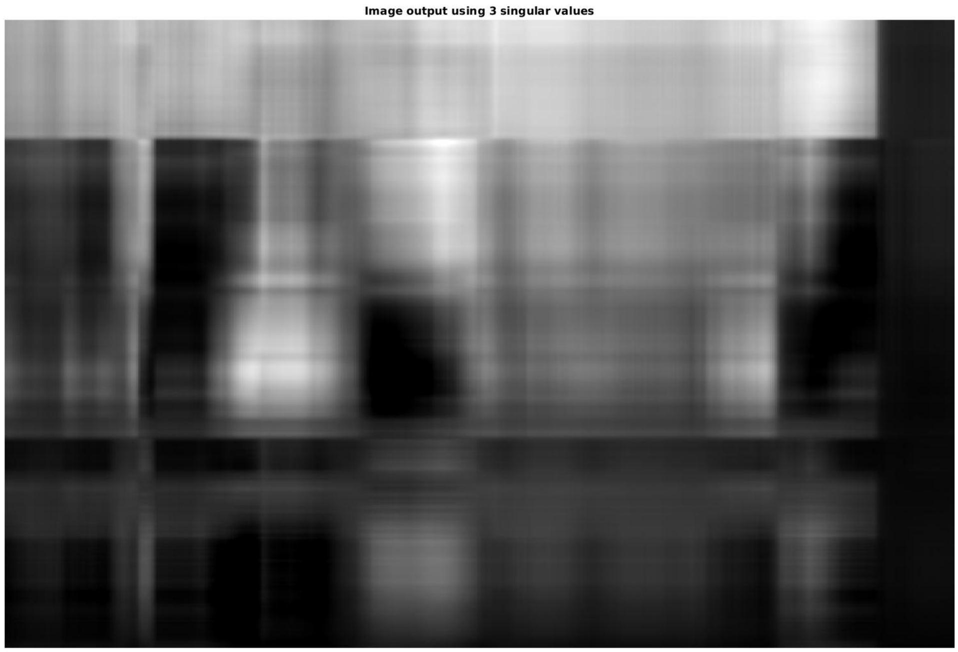

Figure 3.6.14 Image output using first three summands of SVD

Even with the partial sum of 5 summands, interpreting the image is problematic, but remember it takes 1463 to render all the detail. But also, once you know what the image is, you will come back to this rendering and already see essential features.

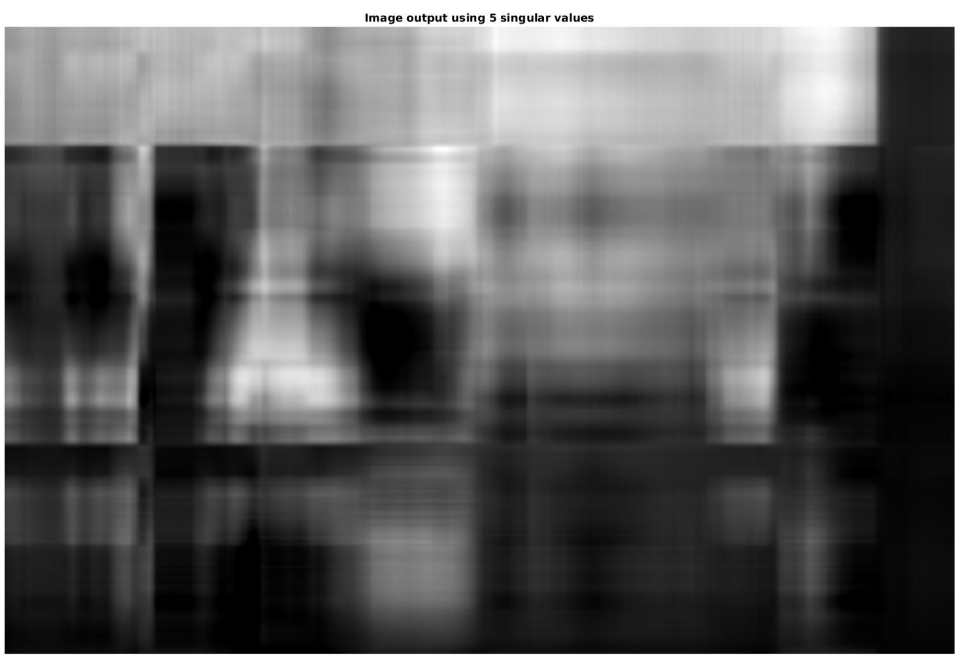

Figure 3.6.15 Image output using first five summands of SVD

Below are the renderings of partial sums with 10, 15, 25, 50, 100, 200, 500, 1000, and all 1463 summands. Look at successive images to see how (and at what stage) the finer detail is layered in. Surely with only 10 summands rendered, there can be no question of what the image is.

Figure 3.6.16 Image output using first 10 summands of SVD

Figure 3.6.17 Image output using first 15 summands of SVD

Figure 3.6.18 Image output using first 25 summands of SVD

Figure 3.6.19 Image output using first 50 summands of SVD

Figure 3.6.20 Image output using first 100 summands of SVD

Figure 3.6.21 Image output using first 200 summands of SVD

Figure 3.6.22 Image output using first 500 summands of SVD

Figure 3.6.23 Image output using first 1000 summands of SVD

Figure 3.6.24 Original image (all 1463 summands)

3.6.8 Some parting observations on the SVD

Back in Theorem 3.2.17 and Corollary 3.2.18 we defined the so-called **four fundamental subspaces.** Let us see how they are connected via the singular value decomposition of a matrix.

We started with an $m \times n$ matrix A having rank r, and an SVD of the form $A = U\Sigma V^*$:

$$\begin{bmatrix} \underbrace{u_1 \ \cdots u_r}_{\text{Col } A} & \underbrace{u_{r+1} \cdots u_m}_{\ker A^*} \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & \mathbf{0} \\ 0 & & \sigma_r & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right] \begin{bmatrix} v_1^* \\ \vdots \\ v_r^* \\ v_{r+1}^* \\ \vdots \\ v_n^* \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix} v_1^* \\ \vdots \\ v_r^* \end{matrix}}\right\} \text{Col } A^* \\ \left.\vphantom{\begin{matrix} v_{r+1}^* \\ \vdots \\ v_n^* \end{matrix}}\right\} \ker A \end{array} \tag{3.6.2}$$

We have the orthonormal basis $\{u_1, \ldots, u_m\}$for $\mathbb{C}^m$ of which $\{u_1, \ldots, u_r\}$ is an orthonormal basis for $\operatorname{Col} A$, the column space of A. So that means that $\{u_{r+1}, \ldots, u_m\}$ is an orthogonal subset of $(\operatorname{Col} A)^\perp = \ker A^*$ by Theorem 3.2.17.

By Corollary 3.2.11, we know that $\mathbb{C}^m = \operatorname{Col} A \boxplus (\operatorname{Col} A)^\perp$, so

$$m = \dim \mathbb{C}^m = \dim \operatorname{Col} A + \dim(\operatorname{Col} A)^\perp,$$

so $\dim(\operatorname{Col} A)^\perp = m - r$, and it follows that $\{u_{r+1}, \ldots, u_m\}$ is an orthonormal basis for $(\operatorname{Col} A)^\perp = \ker A^*$.

Turning to the right side of the SVD, we know that

$$\|Av_i\| = \sigma_i \text{ for } i = 1, \dots, n,$$

and by the choice of r, we know that

$$Av_{r+1} = \cdots = Av_n = 0.$$

Since the rank $A = r$, the nullity $A = n - r$ which means that $\{v_{r+1}, \dots, v_n\}$ is an orthonormal basis for $\ker A$.

Finally it follows that $\{v_1, \dots, v_r\}$ is an orthonormal basis for $(\ker A)^\perp = \operatorname{Col} A^*$. Note that when A is a real matrix, $\operatorname{Col} A^* = \operatorname{Col} A^T = \operatorname{Row} A$.

In display (3.6.2), we have seen a certain symmetry between the kernels and images of A and A^*, and in part we saw that above in Subsection 3.6.4 where we used the SVD for A^* to obtain one for A. We connect the dots a bit more with the following observations.

In constructing an SVD for $A = U\Sigma V^*$, we had an orthonormal basis $\{v_1, \dots, v_n\}$ which were eigenvectors for A^*A with eigenvalues $\lambda_i = \sigma_i^2$. Noting that $\|Av_i\| = \sigma_i$, we set $u_i = \dfrac{1}{\sigma_i}Av_i$ for $i = 1, \dots, r$, observed it was an orthormal set and extended in to an orthonormal basis $\{u_1, \dots, u_m\}$ for $\mathbb{C}^m$.

From the definiton, $u_i = \frac{1}{\sigma_i}Av_i$ we see that $Av_i = \sigma_i u_i$. What do you think A^*u_i should equal?

We compute

$$A^*u_i = \frac{1}{\sigma_i}A^*(Av_i) = \frac{1}{\sigma_i}(A^*A)v_i = \frac{1}{\sigma_i}\lambda_i v_i = \sigma_i v_i.$$

Thus we have the wonderfully symmetric relation:

$$Av_i = \sigma_i u_i \text{ and } A^*u_i = \sigma_i v_i \text{ for } i = 1, \dots, r.$$

Typically in a given singular value decompostion, $A = U\Sigma V^*$, the columns of U are called the **left singular vectors** of A, while the columns of V are called the **right singular vectors.**

3.7 Exercises (with solutions)

Exercises

1. Let $W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in \mathbb{R}^4 \;\middle|\; x_1 + 2x_2 + 3x_3 + 4x_4 = 0 \right\}$.

(a) Find bases for W and $W^\perp$.

Solution. W is the solution space to $Ax = 0$ where A is the 1×4

matrix $A = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$; it is a hyperplane in $\mathbb{R}^4$. We easily read a set of independent solutions from the matrix A which is already in reduced row-echelon form. Taking x_2, x_3, x_4 as free variables, we may take as a basis:

$$\{w_1, w_2, w_3\} = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Thinking the four fundamental subspaces (Theorem 3.2.17), we know that the

$$W^\perp = (\ker A)^\perp = C(A^*) = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \right\}.$$

If you did not recall that fact, it is clear that this vector is in $W^\perp$, but since

$$4 = \dim \mathbb{R}^4 = \dim W + \dim W^\perp,$$

we see we already have a spanning set.

(b) Find orthogonal bases for W and $W^\perp$.

Solution. Since $W^\perp = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \right\}$ is one-dimensional, the given basis is automatically an orthogonal basis.

For W, we use Gram-Schmidt: We take $v_1 = w_1 = \begin{bmatrix} -2 \\ 0 \\ 0 \\ 1 \end{bmatrix}$, and compute

$$v_2 = w_2 - \frac{\langle w_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = \begin{bmatrix} -3/5 \\ -6/5 \\ 1 \\ 0 \end{bmatrix}$$

and

$$v_3 = w_3 - \cdots = \begin{bmatrix} -2/7 \\ -4/7 \\ -6/7 \\ 1 \end{bmatrix}.$$

(c) Find the orthogonal projection of $b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ onto the subspace W.

Hint. It is definitely worth noting that $\mathbb{R}^4 = W \boxplus W^\perp$. The question is, how to leverage that fact.

Solution. The issue we want to leverage is that

$$\mathrm{proj}_W = I_V - \mathrm{proj}_{W^\perp}.$$

Since we know that $W^\perp = \mathrm{Span}\{e\}$ where $e = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$, we compute

$$\mathrm{proj}_{W^\perp}(b) = \frac{\langle b, e\rangle}{\langle e, e\rangle} e = \frac{10}{30} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}.$$

Now using the observation, we compute

$$\mathrm{proj}_W(b) = b - \mathrm{proj}_{W^\perp}(b) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}.$$

One alternative is that having gone to the trouble of finding an orthogonal basis for W, we could brute force the answer from Definition 3.2.13.

Other alternatives: if we made our orthogonal basis for W into an orthonormal one, we could use Corollary 3.3.2. Or perhaps with a bit less fuss, we could simply take advantage of Proposition 3.3.3 as follows: Let

$$A = \begin{bmatrix} -2 & -3 & -4 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then

$$A(A^*A)^{-1}A^* = \begin{bmatrix} \frac{29}{30} & -\frac{1}{15} & -\frac{1}{10} & -\frac{2}{15} \\ -\frac{1}{15} & \frac{13}{15} & -\frac{1}{5} & -\frac{4}{15} \\ -\frac{1}{10} & -\frac{1}{5} & \frac{7}{10} & -\frac{2}{5} \\ -\frac{2}{15} & -\frac{4}{15} & -\frac{2}{5} & \frac{7}{15} \end{bmatrix}$$

is the matrix of the projection map $[\text{proj}_W]$ with respect to the standard basis, so that

$$\text{proj}_W(b) = A \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}.$$

I am pretty sure which method I prefer!

2. Let $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix} \in M_3(\mathbb{R})$.

(a) What observation tells you that A is diagonalizable without any computation?

Solution. It is a real, symmetric matrix so not only is it diagonalizable, it is orthogonally diagonalizable.

(b) Compute the characteristic polynomial.

Solution.

$$\begin{aligned}\chi_A &= \det(xI - A) = \det\left(\begin{bmatrix} x-3 & 0 & 0 \\ 0 & x-1 & -2 \\ 0 & -2 & x-1 \end{bmatrix}\right) = (x-3)[(x-1)^2 - 4] \\ &= (x-3)(x^2 - 2x - 3) = (x-3)^3(x+1).\end{aligned}$$

(c) Determine a basis for each eigenspace.

Solution.

$$A + I = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \mapsto v_1 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

$$A - 3I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & 2 & -2 \end{bmatrix} \mapsto \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mapsto v_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \; v_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Note that $v_3' = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is another obvious choice for an independent eigenvector, though not as useful for a later part (since v_2 and v_3 are orthogonal).

(d) Find a matrix P so that $P^{-1}AP$ is diagonal.

Solution. The matrix P is any matrix with the eigenvectors as columns. For example, if we want the diagonal matrix to be

$$\begin{bmatrix} 3 & & \\ & 3 & \\ & & -1 \end{bmatrix} \text{ choose } P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix},$$

or if we want the diagonal matrix to be

$$\begin{bmatrix} -1 & & \\ & 3 & \\ & & 3 \end{bmatrix} \text{ choose } P = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Other matrices are certainly possible.

(e) Determine whether the matrix A is orthogonally diagonalizable. If not, why; if so, find an orthogonal matrix Q so that $Q^T AQ$ is diagonal.

Solution. Since A is a real symmetric matrix, we know it is orthogonally diagonalizable. The columns of the matrices P above have orthogonal columns. We need only normalize the columns, say

$$Q = \begin{bmatrix} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}.$$

3. View $\mathbb{R}^7$ as an inner product space with the usual inner product.

(a) let $T : \mathbb{R}^7 \to \mathbb{R}^7$ be a linear map with the property that $\langle T(v), v\rangle = 0$ for all $v \in \mathbb{R}^7$. Show that T is not invertible.

Hint. Calculus tells you that a polynomial of degree 7 and real coefficients has at least one real root.

Solution. Let χ_T be the characteristic polynomial of T. Since the degree is odd, the hint says χ_T has a real root, that is, T has a real eigenvalue λ. Let v be a (nonzero) eigenvector with $T(v) = \lambda v$. We now consider the requirement that $\langle T(v), v\rangle = 0$.

$$\langle T(v), v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle = 0.$$

Since $v \neq 0$, we cannot have $\langle v, v\rangle = 0$, so we must have $\lambda = 0$, which says zero is an eigenvalue, and hence the nullspace is nontrivial. This means that T is not invertible.

(b) Show by example that there exist linear maps $T : \mathbb{R}^2 \to \mathbb{R}^2$ with $\langle T(v), v\rangle = 0$ for all $v \in \mathbb{R}^2$, but with T invertible. Verify that your T satisfies the required conditions.

Hint. If we consider the previous part, the dimension only mattered to produce a real eigenvalue, so that provides a direction to look.

Solution. Let $[T]_{\mathcal{E}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ where $\mathcal{E}$ is the standard basis for $\mathbb{R}^2$. Then $\chi_T = x^2 + 1$ which has no real roots. In particular, 0 is not an eigenvalue which means the nullspace is zero, so T is invertible. We claim that $\langle T(v), v\rangle = 0$ for all $v \in V$. We can read off the action of T from the matrix:

$$T(e_1) = -e_2 \text{ and } T(e_2) = e_1, \text{ so } T(ae_1 + be_2) = be_1 - ae_2.$$

We check

$$\langle T(ae_1 + be_2), ae_1 + be_2\rangle = \langle be_1 - ae_2, ae_1 + be_2\rangle = ba - ab = 0,$$

for all a, b.

4. Let V be a finite-dimensional real inner product space, and $T : V \to V$ a linear operator satisfying $T^2 = T$, that is $T(T(v)) = T(v)$ for all $v \in V$. To eliminate trivial situations, assume that T is neither the zero transformation, nor the identity operator.

(a) Show that the only possible eigenvalues of T are zero and one.

Solution. Suppose that $T(v) = \lambda v$ for some nonzero vector v. Then

$$T(v) = T^2(v) = T(T(v)) = T(\lambda v) = \lambda T(v),$$

so $(\lambda - 1)T(v) = 0$, which means either $\lambda = 1$ (so one is an eigenvalue), or $T(v) = 0$ which means the nullspace is not zero, hence zero is an eigenvalue.

(b) Let E_λ denote the λ-eigenspace. Show that $E_0 = N(T)$, the nullspace of T, and that E_1 is the image of T.

Solution. That $E_0 = N(T)$ is the definition of $E_0 = \{v \in V \mid T(v) = 0 = 0v\}$.

If $v \in E_1$, then $T(v) = 1 \cdot v$, but then $T(v) = v$ which says that $v \in R(T)$. Conversely if $w = T(v') \in R(T)$, then $T(w) = T^2(v') = T(v') = w$, so $w \in E_1$. Thus the image is precisely E_1.

(c) Show that T is diagonalizable.

Solution. $\dim E_0$ equals the nullity of T, and from above $\dim E_1$ is the rank, so by rank-nullity, the sum of the sizes of the eigenspaces (which have trivial intersection) is the dimension of the space, so V has a basis of eigenvectors for T.

(d) Let W be a subspace of V, and let $S = \mathrm{proj}_W$ be the orthogonal projection onto the subspace W. Show that $S^2 = S$, so that the orthogonal projection is one linear map satisfying the given property.

Solution. By definition, we take an orthonormal basis for W (say having dimension r), and extend it to an orthonomal basis $\mathcal{B} = \{v_i\}$ for V. Then $S(v) = \sum_{i=1}^{r} \langle v, v_i \rangle v_i = w$ and by Theorem 3.2.10 we know that $v = w^\perp + w$ for unique $w^\perp \in W^\perp$. Since $S(v) = w \in W$ and $w = w + 0$, $S(w) = w$ (Corollary 3.2.14), that is $S^2(v) = S(v)$.

5. Let $A = \begin{bmatrix} 1 & 0 & -1 \\ -4 & 1 & 6 \\ 0 & -5 & -9 \\ 1 & 5 & 8 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$.

In answering the questions below, you may find some of the information below of use. By rref(X) we mean the reduced row-echelon form of the matrix X.

$$\text{rref}(A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{rref}(A|b) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 18 & 1 & -17 \\ 1 & 51 & 91 \\ -17 & 91 & 182 \end{bmatrix}, \quad A^T b = \begin{bmatrix} -3 \\ 7 \\ 16 \end{bmatrix}, \quad \text{rref}(A^T A | A^T b) = \begin{bmatrix} 1 & 0 & 0 & 74 \\ 0 & 1 & 0 & -128 \\ 0 & 0 & 1 & 71 \end{bmatrix}$$

$$AA^T = \begin{bmatrix} 2 & -10 & 9 & -7 \\ -10 & 53 & -59 & 49 \\ 9 & -59 & 106 & -97 \\ -7 & 49 & -97 & 90 \end{bmatrix}, \; AA^T b = \begin{bmatrix} -19 \\ 115 \\ -179 \\ 160 \end{bmatrix}, \; \text{rref}(AA^T | AA^T b) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(a) Show that the system $Ax = b$ is inconsistent.

Solution. We see that rref$(A|b)$ has a pivot in the augmented column, meaning the system is inconsistent.

(b) Find a least squares solution to the system $Ax = b$.

Solution. A least squares solution to $Ax = b$ is obtained by solving the consistent system $A^T Ax = A^T b$. From the work above, we read off the solution $x = \begin{bmatrix} 74 \\ -128 \\ 71 \end{bmatrix}$.

6. Suppose a real matrix has SVD given by $A = U\Sigma V^T$:

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

(a) Using only your knowledge of the SVD (and no compuation), determine rank A.

Solution. The rank is two since there are precisely two nonzero singular values, $\sqrt{3}$ and $\sqrt{2}$.

(b) Using only your knowledge of the SVD, give a basis for the kernel (nullspace) of A; explain your process.

Solution. The SVD process begins by finding an orthonormal basis $\{v_1, \dots, v_4\}$ for A^TA. With $\sigma_i = \|Av_i\|$ and the rank of A equaling 2, we know the nullity of A is also two, and since $Av_3 = Av_4 = 0$, $\{v_3, v_3\}$ gives an orthogonal basis for the kernel.

(c) Using only your knowledge of the SVD, give a basis for the column space of A, explaining your process.

Solution. The column space is spanned by $\{Av_1, \dots, Av_r\}$ where $r = \operatorname{rank} A = 2$, so $\{Av_1, Av_2\}$ is an (orthogonal) basis for the column space.

7. Let A have singular value decomposition

$$A = U\Sigma V^T = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & -2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix}.$$

(a) Prove that A is invertible.

Solution. A is a 2×2 matrix with two nonzero singular values, so has rank 2, and so is invertible. Alternatively, it is easy to show that $\det A \neq 0$.

(b) Using the given SVD, find an expression for A^{-1}.

Solution. $A = U\Sigma V^T$ implies that $A^{-1} = (V^T)^{-1}\Sigma^{-1}U^{-1} = V\begin{bmatrix} 1/8 & 0 \\ 0 & 1/2 \end{bmatrix}U^T$ since both U and $V$are orthogonal matrices.

(c) The goal of this part is to find an SVD for A^{-1}. You should express your answer (confidently) as an appropriate product of matrices without multiplying things out, though you should explain why the expression you write represents an SVD for A^{-1}. In particular, a couple of warm up exercises will help in this endeavor, and no, the answer in part b is not the correct answer.

- First show that the product of two orthogonal matrices in $M_n(\mathbb{R})$ is orthogonal.
- Next show that the diagonal matrices (with real entries) $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ and $\begin{bmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{bmatrix}$ are orthogonally equivalent, i.e., that there exists an orthogonal matrix P so that

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = P\begin{bmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{bmatrix}P^T.$$

- Now you should be able to proceed using your answer from part b as a starting point.

Solution.

- For the first warm up, suppose that $AA^T = I_n = BB^T$. Then

$$(AB)(AB)^T = ABB^TA^T = AI_nA^T = AA^T = I_n.$$

- For the second warm up, one can choose

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

but an explanation would be nice. It should be clear that the standard basis vectors e_1, e_2 for $\mathbb{R}^2$ are eigenvectors for the matrix. $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. It follows that the matrix P with columns e_2, e_1 is also a matrix of eigenvectors, but which reverses the order of appearance of the eigenvalues.

- Now for the main event: The expression for A^{-1} in the previous part would be an SVD for A^{-1} but for the fact that the singular values do not satisfy $\sigma_1 > \sigma_2$. Fortunately the warm up exercises come to the rescue! We see that

$$\begin{bmatrix} 1/2 & 0 \\ 0 & 1/8 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1/8 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

and that $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is an orthogonal matrix with $Q^T = Q$, hence by the warm ups, so are the matrices $QU^T = (UQ^T)^T = (UQ)^T$ and VQ. Thus

$$A^{-1} = (V^T)^{-1}\Sigma^{-1}U^{-1} = V \begin{bmatrix} 1/8 & 0 \\ 0 & 1/2 \end{bmatrix} U^T = (VQ) \begin{bmatrix} 1/2 & 0 \\ 0 & 1/8 \end{bmatrix} (UQ)^T$$

is an SVD for A^{-1}.

Chapter 4

Definitions and Examples

Here we accumulate basic definitions and examples from a standard first course in linear algebra.

4.1 Basic Definitions and Examples

Listed in alphabetical order.

Definition 4.1.1 Given an $n \times n$ matrix A with eigenvalue λ, the **algebraic multiplicity** of the eigenvalue is the degree d to which the term $(x - \lambda)^d$ occurs in the factorization of the characteristic polynomial for A. $\diamondsuit$

Definition 4.1.2 An **basis** for a vector space is a linearly independent subset of the vector space whose span is the entire space. $\diamondsuit$

Example 4.1.3 Some standard bases for familiar vector spaces.

- The standard basis for F^n is $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ where $\mathbf{e}_i$ is the column vector in F^n with a 1 in the ith coordinate and zeroes in the remaining coordinates.
- A standard basis for $M_{m\times n}(F)$ is

  $$\mathcal{B} = \{\mathbf{E}_{ij} \mid 1 \le i \le m, 1 \le j \le n\},$$

  where $\mathbf{E}_{ij}$ is the $m \times n$ matrix with a 1 in row i and column j, and zeroes in all other entries.
- A standard basis for $P_n(F)$ is $\mathcal{B} = \{1, x, x^2, \ldots, x^n\}$, and a standard basis for $P(F) = F[x]$ is $\mathcal{B} = \{1, x, x^2, x^3, \ldots\}$.

$\square$

Definition 4.1.4 The **characteristic polynomial** of a square matrix $A \in M_n(F)$ is $\chi_A = \det(xI_n - A)$. One can show that χ_A is a monic polynomial of degree n with coefficients in the field F.

Note that some authors define the characteristic polynomial as $\det(A - xI_n)$ in which case the leading coefficient is $(-1)^n$, but since the interest is only in the factorization of χ_A (in particular any roots it may have), it does not really matter which definition one uses. $\Diamond$

Definition 4.1.5 The **column space** of an $m \times n$ matrix A is the span of the columns of A. As such, it is a subspace of F^m. $\Diamond$

Definition 4.1.6 Given an $m \times n$ matrix A with complex entries, the **conjugate transpose** of A is the $n \times m$ matrix A^* whose ij-entry is given by

$$(A^*)_{ij} = \overline{A_{ji}} = \overline{(A^T)_{ij}}.$$

$\Diamond$

Definition 4.1.7 The **dimension** of a vector space is the cardinality (size) of any basis for the vector space.

Implicit in the definition of dimension are theorems which prove that every vector space has a basis, and that any two bases for a given vector space have the same cardinality. In other words, the dimension is a well-defined term not depending upon which basis is chosen to consider. When a vector space has a basis with a finite number of elements, it is called **finite-dimensional**. $\Diamond$

Definition 4.1.8 An **elementary matrix** is a matrix obtained by performing a single elementary row (or column) operation to an identity matrix. $\Diamond$

Definition 4.1.9 Elementary row (respectively column) operations on a matrix are one of the following:

- Interchange two rows (resp. columns) of A.
- Multiply a row (resp. column) of A by a nonzero scalar.
- Replace a given row (resp. column) of A by the sum of the given row (resp. column) and a multiple of a different row (resp. column).

$\Diamond$

Definition 4.1.10 Given an $n \times n$ matrix A with eigenvalue λ, the **geometric multiplicity** of the eigenvalue is the dimension of the eigenspace associated to λ. $\Diamond$

Definition 4.1.11 A complex matrix A is called **Hermitian** if $A = A^*$. Necessarily the matrix needs to be square. $\Diamond$

Definition 4.1.12 The **image** of a linear map $T : V \to W$ is

$$\operatorname{Im}(T) := \{w \in W \mid w = T(v) \text{ for some } v \in V\}.$$

The image of T is a subspace of W; T is surjective if and only if $W = \operatorname{Im}(T)$. $\Diamond$

Definition 4.1.13 A function $f : X \to Y$ between sets X and Y is **injective** if for every $x, x' \in X$, $f(x) = f(x')$ implies $x = x'$. $\diamondsuit$

Definition 4.1.14 Let F denote the field of real or complex numbers. For $z = a + bi \in \mathbb{C}$ ($a, b \in \mathbb{R}$ and $i^2 = -1$), we have the notion of the **complex conjugate** of z, denoted $\overline{z} = a - bi$. Note that when $z \in \mathbb{R}$, that is $z = a = a + 0i \in \mathbb{C}$, we have $z = \overline{z}$. The **magnitude** (**norm**, *absolutevalue*) of $z = a + bi$ is $|z| = \sqrt{a^2 + b^2}$.

Let V be a vector space over the field F. An **inner product** is a function:

$$\langle \cdot, \cdot \rangle : V \times V \to F$$

so that for all $u, v, w \in V$ and $\lambda \in F$:

1. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
2. $\langle \lambda v, w \rangle = \lambda \langle v, w \rangle$
3. $\overline{\langle v, w \rangle} = \langle w, v \rangle$, where the bar denotes complex conjugate.
4. $\langle v, v \rangle$ is a positive real number for all $v \neq 0$.

$\diamondsuit$

Definition 4.1.15 An **inner product space** is a vector space V defined over a field $F = \mathbb{R}$ or $\mathbb{C}$ to which is associated an inner product. If $F = \mathbb{R}$, V is called a **real inner product space**, and if $F = \mathbb{C}$, then V is called a **complex inner product space.** $\diamondsuit$

Definition 4.1.16 An **isomorphism** is a linear map which is bijective (one-to-one and onto; injective and surjective). $\diamondsuit$

Definition 4.1.17 The **Kronecker delta** is defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

$\diamondsuit$

Definition 4.1.18 A **linear combination** of vectors $v_1, \ldots, v_r \in V$ is any vector of the form $a_1 v_1 + \cdots + a_r v_r$ for scalars $a_i \in F$. $\diamondsuit$

Definition 4.1.19 Let $S \subseteq V$ be a subset of vectors in a vector space V (finite or infinite). The set S is a **linearly dependent** subset of V if it is not linearly independent, that is there exists a *finite* subset $\{v_1, \ldots, v_r\} \subseteq S$, and nonzero scalars $a_1, \ldots, a_r$ so that

$$a_1 v_1 + \cdots + a_r v_r = \mathbf{0}.$$

$\diamondsuit$

Definition 4.1.20 Let $S \subseteq V$ be a subset of vectors in a vector space V (finite or infinite). The set S is a **linearly independent** subset of V if for every *finite* subset $\{v_1, \ldots, v_r\} \subseteq S$, a linear dependence relation of the form

$$a_1v_1 + \cdots + a_rv_r = \mathbf{0}$$

forces all the scalars $a_i = 0$. $\diamondsuit$

Definition 4.1.21 Given two vector spaces V and W (defined over the same field F), a **linear map** (or **linear transformation**) from V to W is a function $T : V \to W$ which is

- additive: $T(v + v') = T(v) + T(v')$ for all $v, v' \in V$, and
- preserves scalar multiplication: $T(\lambda v) = \lambda T(v)$ for all vectors $v \in V$ and scalars λ.

$\diamondsuit$

Definition 4.1.22 The **minimal polynomial** of a square matrix $A \in M_n(F)$ is the monic polynomial, μ_A, of least degree with coefficients in the field F so that $\mu_A(A) = 0$. The **Cayley-Hamilton** theorem implies that the minimal polynomial divides the characteristic polynomial. $\diamondsuit$

Definition 4.1.23 A matrix $A \in M_n(\mathbb{C})$ is **normal** if it commutes with its conjuate transpose: $AA^* = A^*A$. $\diamondsuit$

Definition 4.1.24 The **nullity** of a linear transformation $T : V \to W$ is the dimension of $ker(T)$, that is, the dimension of its nullspace.

If $T : F^n \to F^m$ is given by $T(x) = Ax$ for an $m \times n$ matrix A, then the nullity of T is the dimension of the set of solutions of $Ax = 0$. $\diamondsuit$

Definition 4.1.25 The **nullspace** of a linear transformation $T : V \to W$ is the kernel of T that is,

$$ker(T) = \{v \in V \mid T(v) = \mathbf{0}_W\}.$$

If $T : F^n \to F^m$ is given by $T(x) = Ax$ for an $m \times n$ matrix A, then the nullspace of T is often called the **nullspace of** A, the set of solutions of $Ax = 0$. $\diamondsuit$

Definition 4.1.26 A matrix $A \in M_n(\mathbb{R})$ is an **orthogonal** matrix if

$$A^TA = AA^T = I_n.$$

Note that the condition $A^TA = I_n$ is equivalent to saying that the columns of A form an orthonormal basis for $\mathbb{R}^n$, while the condition AA^T makes the analogous statement about the rows of A. $\diamondsuit$

Definition 4.1.27 The **pivot positions** of a matrix are the positions (row,column) which correspond to a leading one in the reduced row-echelon form of the matrix. The **pivots** are the actual entry of the given matrix at the pivot position.

The **pivot columns** are the columns of the ***original*** matrix corresponding to the columns of the RREF containing a leading one. $\lozenge$

Definition 4.1.28 The **rank** of a linear transformation $T : V \to W$ is the dimension of its image, $\mathrm{Im}(T)$.

If $T : F^n \to F^m$ is given by $T(x) = Ax$ for an $m \times n$ matrix A, then the rank of T is the dimension of the column space of A.

By theorem, it is also equal to the dimension of the row space which is the number of nonzero rows in the RREF form of the matrix A. $\lozenge$

Definition 4.1.29 The **row space** of an $m \times n$ matrix A is the span of the rows of A. As such, it is a subspace of F^n. $\lozenge$

Definition 4.1.30 Let $A, B \in M_n(F)$. The matrix B is said to be **similar** (or **conjugate**) to A if there exists an invertible matrix $P \in M_n(F)$ so that $B = P^{-1}AP$. Note that if we put $Q = P^{-1}$, then $B = QAQ^{-1}$, so it does not matter which side carries the inverse. Also note that this is a symmetric relationship, so that B is similar to A if and only if A is similar to B. Indeed similarity (conjugacy) is an equivalence relation. $\lozenge$

Definition 4.1.31 Let $S \subseteq V$ be a subset of vectors in a vector space V (finite or infinite). The **span** of the set S, denoted $\mathrm{Span}(S)$, is the set of all finite linear combinations of the elements of S. That is to say

$$\mathrm{Span}(S) = \{a_1v_1 + \cdots + a_rv_r \mid r \geq 1, a_i \in F, v_i \in S\}$$

$\lozenge$

Definition 4.1.32 Let V be a vector space over a field F, and let $W \subseteq V$. W is called a **subspace** of V if W is itself a vector space with the operations of vector addition and scalar multiplication **inherited** from V.

Of course checking all the vector space axioms can be quite tedious, but as a theorem you prove much easier criteria to check. Recall that you already know that V is a vector space, so many of the axioms (associativity, distributive laws etc) are inherited from V. Indeed, you prove that to show that W is a subspace of V, it is enough to show that the additive identity of V is in W, and that W is **closed** under the inherited operations of vector addition and scalar multiplication, i.e, whenever $w, w' \in W$ and $\lambda \in F$, we must have $w + w' \in W$, and $\lambda w \in W$. $\lozenge$

Definition 4.1.33 A function $f : X \to Y$ between sets X and Y is **surjective** if for every $y \in Y$, there exists an $x \in X$ such that $f(x) = y$. $\lozenge$

Definition 4.1.34 A matrix A is called **symmetric** if $A = A^T$. Necessarily the matrix needs to be square. $\lozenge$

Definition 4.1.35 Given a square matrix $A \in M_n(F)$, we define its **trace** to be the scalar

$$\mathrm{tr}(A) := \sum_{i=1}^{n} A_{ii}.$$

♢

Definition 4.1.36 A matrix $A \in M_n(\mathbb{C})$ is an **unitary** matrix if

$$A^* A = AA^* = I_n.$$

Note that the condition $A^* A = I_n$ is equivalent to saying that the columns of A form an orthonormal basis for $\mathbb{C}^n$, while the condition AA^* makes the analogous statement about the rows of A. ♢

Definition 4.1.37 A **vector space** is a non-empty set V and an associated field of scalars F, having operations of vector addition, denoted $+$, and scalar multiplication, denoted by juxtaposition, satisfying the following properties: For all vectors $u, v, w \in V$, and scalars $\lambda, \mu \in F$

- **closure under vector addition**
- $u + v \in V$
- **addition is commutative**
- $u + v = v + u$
- **addition is associative**
- $(u + v) + w = u + (v + w)$
- **additive identity**
- There is a vector $\mathbf{0} \in V$ so that $\mathbf{0} + u = u$.
- **additive inverses**
- For each $u \in V$, there is a vector denoted $-u \in V$ so that $u + -u = \mathbf{0}$.
- **closure under scalar multiplication**
- $\lambda u \in V$.
- **scalar multiplication distributes across vector addition**
- $\lambda(u + v) = \lambda u + \lambda v$
- **distributes over scalar addition**
- $(\lambda + \mu)v = \lambda v + \mu v$
- **scalar associativity**
- $(\lambda \mu)v = \lambda(\mu v)$
- V **is unital**
- The field element $1 \in F$ satisfies $1v = v$.

♢

References and Suggested Readings

[**1**] Friedberg, S., Insel, A., and Spence, L. *Linear Algebra.* 4th ed. Pearson Education, Upper Saddle River, NJ, 2003.

[**2**] Lay, D., Lay, S, and McDonald, J. *Linear Algebra and its Applications.* 5th ed. Pearson, Boston, 2016.

[**3**] Meckes, E. and Meckes, M. *Linear Algebra.* 1st edition, Cambridge University Press, 2018.