Linear Algebra

Lucas Brito

December 28, 2020

Contents

Preface

These lecture notes are based on MIT's course 18.06 - Linear Algebra, as taught by Professor Gilbert Strang in the spring of 2005, in addition to the book *Linear Algebra Done Right* by Sheldon Axler. The notes' purpose is twofold—to facilitate picking up linear algebra for a second time after some time away from it, and to catalog the various pedagogies and analogies I've (personally) found useful in understanding the topic. As such, the depth of its content ranges from introductory to advanced. Often a section will start with a first-course level of intuition building, then transition to a more rigorous treatment. I would not recommend learning linear algebra from these alone; rather, I believe these lecture notes to be useful as a purely supplemental tool. I recommend reading them alongside a primary text.

If you have any suggestions for improvement, corrections, or overall comments, please reach me at `lucas_brito@brown.edu`.

Introduction

> Mathematics requires a small dose, not of genius, but of an imaginative freedom which, in a larger dose, would be insanity.
>
> Angus K. Rodgers

The Mental Preparation for Linear Algebra

...Another thing that I really want to mention is that linear algebra is a pretty big meal for most newcomers, so having "guiding principles" in your mind as you learn it will speed up the process. Here are the 3 that I found most useful for my students and for myself:

1. Keep a list of "equivalent things" to invertibility. You'll realize that a lot of concepts neatly focus this way, from eigenvalues to solving equations to rank to taking matrix inverses. People who are experienced do these things fairly automatically, but it may take some time to get used to this if you are new.

2. Always, always think about how to take the "skeleton" of a matrix that captures most of its essence. The clearest example of this is a dirty secret: mathematicians (and especially physicists) like to "squint their eye" at a matrix and only see a diagonal matrix with the eigenvalues on the diagonal, then they make smart-sounding statements about the matrix and then justify it later with similarity transformations and (heaven forbid) the Jordan form.

3. (Slightly more relevant for abstract thinking, but useful for both.) Matrices always have two interpretations: one as a transformation that eats vectors and spits out vectors; the other as a bilinear form that eats two vectors and spits out a number. To kill your confusion by 80%, always mentally remind yourself which world you are in. This also helps you to remember when you need to wrap your matrix by something and its inverse (the former) and when you need to wrap your matrix by something and its transpose (the latter).

- Yan Zhang

Chapter 1

Solving Linear Systems

1.1 The Geometry of Linear Equations

We begin linear algebra by asking ourselves how we can solve a system of linear equations. For example, consider all x and y for which twice x minus one y is zero, as well as all x and y for which twice y minus one x is three. Most students that are yet to be versed in the language of linear algebra might set up this problem by considering a pair of equations

$$\begin{cases} 2x - y = 0 \\ -x + 2y = 3 \end{cases}$$

then solving for an x in terms of y and plugging in then continuing to narrow in on a solution. This is fine for many problems, the above included. However, for reasons that will eventually become abundantly clear, linear algebra sheds this set up—and even this approach—for one centered around matrices. We choose to express the system above as

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

often abbreviated $Ax = b$. The solution (x, y) then can be interpreted as an x and y which satisfy

$$x \begin{bmatrix} 2 \\ -1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

In words, we are taking multiplicatives of the vector $(2, -1)$ and adding those to multiplicatives of the vector $(-1, 2)$ to see if we can obtain the vector $(0, 3)$. These are two operations: scalar multiplication and addition. This will become a pivotal idea, perhaps the most pivotal of linear algebra.

Note: I proceed assuming the reader has a cursory understanding of vector operations such as the dot product, as well as matrix multiplication. If you don't, look them up.

1.2 Elimination with Matrices

When solving linear systems with equations, we use the process of elimination. We may eliminate by adding multiples of rows to other rows. Our goal is to turn the matrix A into an **echelon matrix** U, or to put it in echelon form. Specifically, we do this by trying eliminate values from the bottom left corner. Observe:

$$pivots \Rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 3-3 & 8-6 & 1-3 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix}$$

we add negative three times the first row to the second row, and produce a zero on the second row. We can then add negative two times the second row to the third row to produce U:

$$U = \begin{bmatrix} \mathbf{1} & 2 & 1 \\ 0 & \mathbf{2} & -2 \\ 0 & 0 & \mathbf{5} \end{bmatrix}$$

The bolded numbers, constituting the "edge" of our U, are the **pivots**. Pivots are the entries of a matrix with zeroes below them. This definition is important; remember it.

1.2.1 Reduced Row Echelon Form

We can further *reduce* this matrix by using row three to eliminate in row two, and row two to eliminate in row one. We then divide rows by scalars so that every pivot entry is one, putting it in **reduced row echelon form** R:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \Rightarrow R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

1.2.2 Augmented Matrices

If we desire to keep track of a right side b in $Ax = b$ (so that we may solve for x), we can construct an **augmented matrix** $[\, A \;\; b \,]$, where we insert the b column to the right of the matrix. For example,

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad [\, A \;\; b \,] = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 3 & 8 & 1 & 2 \\ 0 & 4 & 1 & 3 \end{bmatrix}$$

As we eliminate, we perform the same operations on b as if it was part of the same matrix. That way, when we get to an echelon form, we can easily put the matrix back in equation form to obtain some equalities and back substitute appropriately. In the above example, for $b = (1, 2, 3)$, we obtain

$$U = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 2 & -2 & 1 \\ 0 & 0 & 5 & 5 \end{bmatrix} \Rightarrow \begin{cases} x + 2y + z = 1 \\ 2y - 2z = 1 \\ 5z = 5 \end{cases}$$

where we can easily solve for x, y and z.

1.2.3 Elimination Matrices

If we want to keep track of the eliminations we've made, we can use an elimination matrix E. For example, the action of adding negative two times row one to row two can be expressed as

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

(see 1.3). If we are to continue making eliminations, we add similar matrices to the left, as matrix multiplication is to occur from right to left. We obtain something like $E_3E_2E_1A = R$. We can then multiply these elimination matrices together in order to obtain *one* elimination matrix which, when multiplied with A, gets us to the reduced row echelon form of A.

1.3 Interpreting Matrices by Rows and Columns

Multiplying matrices becomes easier once we start interpreting matrices and vectors as operations on rows and columns of other matrices. Starting with vectors, we already know from section 1.1 that a vector $x = (x_1, x_2, x_3)$ produces a combination of the columns a matrix. This is literally what is going on when we multiply a vector by a matrix.

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} a \\ d \\ g \end{bmatrix} + x_2 \begin{bmatrix} b \\ e \\ h \end{bmatrix} + x_3 \begin{bmatrix} c \\ f \\ i \end{bmatrix}$$

1.3.1 Extracting Rows and Columns

By this logic, we can extract the second column of a matrix if we multiply by the column vector $x = (0, 1, 0)$ on the *right*. We can likewise extract the first and third columns by multiplying by the column vectors $(1, 0, 0)$ and $(0, 0, 1)$ respectively. This can be generalized to any matrix with n columns.

We can also extract rows of a matrix by multiplying on the $left$ of a matrix by row vectors such as $(1, 0, 0)$. This hints at the crucial property:

Multiplications on the **left** act on rows. Multiplications on the **right** act on columns.

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} b \\ e \\ h \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} d & e & f \end{bmatrix}$$

1.3.2 Constructing New Matrices

Any matrix multiplication can be considered a series of extractions of rows or columns, multiplications by scalars, then additions. We continue to follow the property that multiplications on the left act on rows and multiplications on the right act on columns. If we have more than one row on, say, a matrix on the left, we're simply constructing a new matrix where the rows are those combinations. For example, the matrix multiplication

$$EA = \begin{bmatrix} 1 & -3 & 2 \\ 1 & 2 & 3 \\ 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

is *really* telling you to construct a new matrix from A where the first row is: 1 times the first row of A plus -3 times the second row of A plus 2 times the third row of A. The second row is 1 times the first row plus 2 times the second row plus 3 times the third row, and the third row follows a similar pattern. If E only had two rows, then oh well! We're only told to construct two rows from A and its rows.

When we multiply on the right, we perform these operations on columns.

$$AB = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} 1 & -3 & 2 \\ 1 & 2 & 3 \\ 4 & 1 & 1 \end{bmatrix}$$

tells you to construct a matrix where the first *column* is one times (a, d, g) plus one times (b, e, h) plus 4 times (c, f, i).

1.4 Permutation Matrices

A permutation matrix P rearranges the rows or columns of a matrix without changing their entries. Following our column/row interpretation from section 1.3, it is easy to see that a permutation matrix is constructed by simply making the desired row changes to an *identity matrix*. For example, if we want to construct a permutation matrix P_{12} which switches rows one and two, we simply switch rows one and two of the identity matrix,

$$P_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and multiply on the left. Notice that there are $n!$ permutation matrices of size n by n, a fact which you can easily test for yourself.

1.5 Inverse Matrices

The inverse of A is a matrix A^{-1} such that $A^{-1}A = I$. A matrix that cannot be inverted is called singular or non-invertible. The "invertibility" of a matrix is a quite telling property:

- For square matrices, it means that we can row reduce A into I, or that the reduced row echelon form of A *is* I.
- For square matrices n by n, an invertible matrix's column space will be $\mathbb{R}^n$. In other words, all of its column vectors are independent.
- An invertible matrix has solutions for every b in $Ax = b$. This is also equivalent to the above.
- An invertible matrix's rows are independent.
- An invertible matrix has a nonzero determinant.
- $Ax = 0$ will have one solution, $x = 0$, for an invertible matrix. Singular matrices have nonzero solutions to $Ax = 0$.
- $Ax = b$ will have one solution for an invertible matrix.
- An invertible matrix has full rank.

Two properties of inverse matrices:

$$(AB)^{-1} = B^{-1}A^{-1}, \;\; (ABC)^{-1} = C^{-1}B^{-1}A^{-1}, \;\; (A^T)^{-1} = (A^{-1})^T$$

1.5.1 Finding Inverse Matrices

To find an inverse matrix, we start by augmenting A with I: $[\,A\;\;I\,]$. It's easy to see that multiplying $A^{-1}[\,A\;\;I\,]$ gives us $\left[\,I\;\;A^{-1}\,\right]$ by the distributive property of block matrices. As such, if we reduce the left portion of the augmented matrix $[\,A\;\;I\,]$ to I by elimination, the eliminations will yield the inverse matrix on the right.

$$E\,[\,A\;\;I\,] \Rightarrow \left[\,I\;\;A^{-1}\,\right]$$

1.6 Transposes

We transpose a matrix by turning its columns into rows.

$$A = \begin{bmatrix} 1 & 4 & 2 & 5 \\ 6 & 3 & 5 & 1 \\ 0 & 6 & 5 & 4 \end{bmatrix},\; A^T = \begin{bmatrix} 1 & 6 & 0 \\ 4 & 3 & 6 \\ 2 & 5 & 5 \\ 5 & 1 & 4 \end{bmatrix}$$

This is a useful operation for a variety of reasons. We can easily analyze the row space of a matrix by considering the column space of the transpose of a matrix:

$$R(A) = C(A^T)$$

The product of a matrix and its inverse will always produce a **symmetric matrix** $A^T = A$:

$$(R^T R)^T = R^T R^{TT} = R^T R$$

where we use $(AB)^T = B^T A^T$ and $(R^T)^T = R$.

1.7 Factorization A = LU

Given an elimination matrix E which puts a matrix A into an upper triangular form U, we can take the inverse of E to factor A into upper and lower triangular forms L and U:

$$EA = U \Rightarrow A = E^{-1}U \Rightarrow A = LU \text{ where } L = E^{-1}$$

As an example, consider the following matrix operation of the form $EA = U$

$$\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$$

We can multiply both sides by $E^{-1} = L$,

$$\begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} = LU$$

Note that L has the constant diagonal of ones and U does not. We can further factor this matrix into $A = LDU$ where D is the diagonal of U

In some cases, we might not be able to factorize a matrix in this way without permuting some rows. In this event, our factorization will instead be $PA = LU$.

1.7.1 Solving Using A = LU

For $Ax = b$, we can leverage our factorization LU. Multiplying both sides by L^{-1} brings A to U (as $A = LU$ and therefore $L^{-1}A = U$):

$$Ux = L^{-1}b$$

We can set $L^{-1}b$ equal to some c and multiply both sides by L to get

$$Lc = b$$

which we then use to solve for c. Once we've solved for c, we simply use

$$Ux = c, \quad Lc = b$$

to solve for x.

1.8 Solving $Ax = 0$

It is obvious that for matrices with sufficient pivots, the only solution to $Ax = 0$ is the zero vector. Matrices with less pivots than columns, however, actually have many solutions. We find them as follows.

Assuming we have already performed elimination and put the matrix in echelon form, we might observe there are fewer pivots than columns, as is the case in

$$\begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The columns to the right of the pivot columns are called free columns. The presence of these free columns indicates that we have more unknowns than solutions. At first we might throw our hands up and say that there's no way to find all of them. However, we can use linear independence to our advantage. When solving $Ax = 0$, we can choose one free column to be multiplied by one, and the rest of our free columns to be multiplied by zero. Each time we do this, we are finding a **special solution**. In the matrix above, we can use the special solutions $x = (-, 1, -, 0)$ and $x = (-, 0, -, 1)$ where the blanks can be solved for:

$$\begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} - \\ 1 \\ - \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} x_1 + 2 + 2x_3 + 2(0) = 0 \\ 2x_3 + 4(0) = 0 \end{cases}$$

$$\begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} - \\ 0 \\ - \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} x_1 + 2(0) + 2x_3 + 2 = 0 \\ 2x_3 + 4 = 0 \end{cases}$$

We can then fully describe the solutions to $Ax = 0$ by taking all linear combinations of the special solutions:

$$\text{Solutions to } Ax = 0\text{: } c_1 \begin{bmatrix} -2 \\ 0 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$$

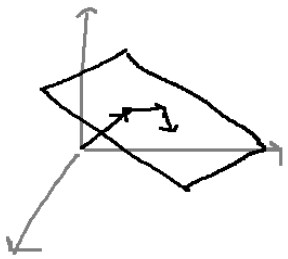

Figure 1.1: An example of the complete solution.

1.9 Particular Solutions

In section 1.8 we found a way to solve a system with more unknowns than equations, but only for a right-hand side zero. We can use a similar strategy to solve for $Ax = b$; our approach is to set all free variables equal to zero. That way we get one *particular* valid solution to our system (hence the name).

$$\begin{bmatrix}1 & 2 & 2 & 2\\ 0 & 0 & 2 & 4\\ 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}-\\ 0\\ -\\ 0\end{bmatrix} = \begin{bmatrix}1\\ 3\\ 0\end{bmatrix} \Rightarrow \begin{cases}x_1 + 2(0) + 2x_3 + 2(0) = 1\\ 2x_3 + 4(0) = 3\end{cases}$$

Where substituting and solving gives us $x = (-2, 0, 3/2, 0)$.

1.10 The Complete Solution to $Ax = b$

We can easily put the particular solution and the solutions to $Ax = 0$ together by noticing that

$$(Ax_p = b) + (Ax_n = 0) \Rightarrow Ax = b$$

or, more precisely, if a set of vectors x_n gives the zero vector, we can add them to the particular solution to $Ax = b$ without consequence; no matter what those vectors are, they'll contribute zero to the final solution.

As such, the complete solution to a system $Ax = b$ becomes something like

$$x = \begin{bmatrix}p_1\\ p_2\\ p_3\end{bmatrix} + c_1\begin{bmatrix}s_1\\ 0\\ 1\end{bmatrix} + c_2\begin{bmatrix}s_2\\ 1\\ 0\end{bmatrix}$$

where we *cannot* multiply the particular solution by an arbitrary constant because this was, after all, a particular solution which yielded the particular right hand side b.

Graphically, we can think of this as representing, say, a plane. We first move according to the particular solution. Then, at the tip of that vector, we are free to add any combination of the special solutions. These linear combinations clearly constitute a plane. See figure 1.1.

1.10.1 Different Scenarios

The appearance of the complete solution is dependent on the rank of the matrix, or how many pivot columns there are. Here are a handful of different scenarios for an m by n matrix.

$$r = n < m$$

In this case we have a matrix that looks somewhat like this:

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

Note that this matrix will only have solutions if the value of b in the second and third rows is zero. This is equivalent to saying that the solution must be in the matrix's column space. If this is true, then there exists a solution. However, this solution is only one: the particular solution. There are no free columns to use to explore the nullspace. The nullspace is only the zero vector.

* * *

$$r = m < n$$

In this case we have a matrix such as

$$\begin{bmatrix} 1 & 0 & a & b \\ 0 & 1 & c & d \end{bmatrix}$$

Where column space covers the entirety of $\mathbb{R}^2$. More broadly, it will cover the entirety of $\mathbb{R}^m$. This means we can solve this system for any right hand side. However, it also means that we have free variables which may be used to determine solutions to $Ax = 0$. Therefore we have infinitely many solutions.

* * *

$$r < m,\ r < n$$

This case combines both above cases. We can either have no solutions, or infinitely many; this depends on whether the right-hand side has zeroes in the appropriate entries. If it does, we note only have a particular solution, but also several solutions that yield zero and can therefore be included. Hence, either zero or infinitely many solutions.

* * *

Summarizing,

$r = m = n$	$r = n < m$	$r = m < n$	$r < m,\ r < n$
One solution	Zero or one solution	∞ solutions	zero or ∞ solutions
$R = I$	$R = \begin{bmatrix} I \\ 0 \end{bmatrix}$	$R = \begin{bmatrix} I & F \end{bmatrix}$	$R = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$

1.10.2 Steps to Solving a Linear System

1. Reduce $[\,A \;\; b\,]$ to $[\,U \;\; c\,]$ so that $Ax = b$ becomes $Ux = c$
2. Find the condition on b_1, b_2, b_3 for $Ax = b$ to have a solution.
3. Describe the column space of A.
4. Find the nullspace of A by setting $Ax = 0$ then choosing independent free variables. Describe the nullspace of A.
5. Find the particular solution by plugging zero into the free variables.

Chapter 2

Vector Spaces

2.1 Introduction to Vector Spaces

Now we begin leaving our comfort zone of linear systems and dipping a toe in more abstract waters. This is, in the opinion of the author, where linear algebra gets interesting, and where it becomes more of a study and less of a collection of computational techniques. The following defines a *vector space*, a crucial idea in this subject.

> Consider a set of objects on which you can perform two operations, addition and scalar multiplication. A **vector space** is the set of all possible combinations of these operations on these objects.

There remain a handful of conceptual ambiguities to be clarified. The terminology seems to suggest that vector spaces are comprised of vectors as we have been defining them: objects with n components in $\mathbb{R}$, objects with "magnitudes" and "directions." Notice that nowhere in the definition of a vector space is this specification made. Instead, the "vectors" of a vector space are any mathematical objects we can add together and perform scalar multiplications on (as long as those operations satisfy a set of rules). This includes matrices, functions, vectors with complex components, so on.

If we *cannot* perform a certain operation on a member of of a set without leaving the set, then that set is not a vector space. For example, the set of all vectors with positive components is not a vector space, for multiplying by a negative scalar produces a vector with negative components.

We refer to the set of all linear combinations given by a set of vectors as the vectors' **span**. Span and vector spaces are two connected ideas; the span of a set of vectors is a vector space.

Linear algebra texts typically denote vector spaces by the capital letters V, U, or W. As such, the statement $u, v \in V$ says that the vectors u and v are in the vector space V.

2.1.1 Subspaces

A subspace of a vector space is a subset of that vector space that is, itself, a vector space. An example of a subset of $\mathbb{R}^2$ is any line going through the origin. The line must go through the origin because we can multiply any vector on the line by the scalar zero.

$\mathbb{R}^3$ has a handful of subspaces. As with $\mathbb{R}^2$, we can consider lines through the origin. However, we can also consider any planes through the origin. Such a subspace would be the set of all scalar multiplications and additions of *two* vectors. We also consider $\mathbb{R}^3$ to be a subspace of itself, but only because it fits the definition. Here is a complete list of the subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$:

$\mathbb{R}^2$	$\mathbb{R}^3$
The zero vector	The zero vector
Any line through the origin	Any line through the origin
$\mathbb{R}^2$	Any plane through the origin
	$\mathbb{R}^3$

2.2 Definition of a Vector Space

The definition in section 2.1 is a bit of a lie. It only holds if the following properties are also true for that space. A set V is a vector space if (where $v, u, w \in V$, and $\mathbb{F}$ is the field of V):

- **Commutativity**: $u + v = v + u$.
- **Associativity**: $(u + v) + w = v + (u + w)$ and $(ab)v = a(bv)$ for all $a, b \in \mathbb{F}$.
- **Additive identity**: There exists an element $0 \in V$ such that $v + 0 = v$.
- **Additive inverse**: For every v, there exists w such that $v + w = 0$.
- **Multiplicative identity**: $1v = v$.
- **Distributive properties**: $a(u + v) = au + bv$ and $(a + b)v = av + bv$ for all $a, b \in \mathbb{F}$.

Example The set of all polynomials is a vector space:

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

where x can be complex. We interpret the *coefficients* as being "entries" of a vector. For example, the zero vector is $a_0, \ldots, a_n = 0$ and $p(x) = 0$. A basis for this vector space is $1, x, x^2, \ldots$.

$\mathcal{P}(\mathbb{F})$ denotes the set of all polynomials with coefficients in the field $\mathbb{F}$. This is an infinite-dimensional vector space! $\mathcal{P}_n(\mathbb{F})$ denotes the set of all polynomials with coefficients in $\mathbb{F}$ with degrees up to n. This vector space has dimension $n + 1$ (counting x^0).

2.3 Column Space

Having defined vector spaces, we can go on to analyze a matrix with more abstraction.

We can think of the columns of a matrix as a set of vectors. The **column space** of a matrix is the set of all linear combinations of those vectors. For example, the column space of

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

is all of $\mathbb{R}^3$. We can take combinations of the three vectors $(1, 0, 0)$, $(2, 3, 0)$, and $(1, 0, 4)$ to reach any point in $\mathbb{R}^3$.

The column space of

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

on the other hand, is a plane in $\mathbb{R}^3$. Here we see the connection between the ability to solve a system and the number of pivot columns. If the third row of b is zero, that means that b just happens to be in the column space; if the case is otherwise, we cannot reach that right-hand side because it is outside of the column space.

In order words, for a system to be solvable, augmenting to $\begin{bmatrix} A & b \end{bmatrix}$ must *not* enlarge the column space of A. If adding the column b produces a larger column space, that implies that the column space of A did not contain b and thus that no linear combination of A's columns would produce b. This is equivalent to the observation made in section 1.10 that if there is a zero row in R, that same row of b must have zero in its entry.

2.4 Nullspace

The nullspace (or kernel) of a matrix consists of the solutions to $Ax = 0$. In solving $Ax = 0$ (finding the special solutions, finding free variables, etc.) we have already extensively described nullspaces.

For example, the matrix

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 2 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$$

is zero if multiplied by

$$c \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

Its nullspace is thus a line through the origin, or all scalar multiplications of the vector $(1, 1, -1)$.

Strang's Review of Key Ideas

1. The nullspace $N(A)$ is a subspace of $\mathbb{R}^n$ it contains all solutions to $Ax = 0$.
2. Elimination produces an echelon matrix U, then a row reduced echelon form R with pivot columns and free columns.
3. Every free column of R leads to a special solution. The free variable equals 1 and the other free variables equal 0. Back substitution solves $Ax = 0$.
4. The complete solution to $Ax = 0$ is a combination of the special solutions.
5. For a $m \times n$ matrix, if $n > m$ (i.e. if A is a horizontal rectangular matrix), then A has at least one column without pivots, giving a special solution. So there must be nonzero vectors in the nullspace of this rectangular A.

2.5 Linear Independence

The vectors $x_1, x_2, x_3, \ldots, x_n$ are independent if no combination of them gives zero expect the zero combination.

In other words, we can only get the zero vector if we multiply every vector by zero. Consequentially, for a vector to be independent of a preexisting vector, it must enlarge the vector space. For example, if your subspace consists of a line, a vector also on that line would *not* be independent; we can imagine that that vector could be added to another vector in the subspace to produce zero. However, if that vector was to point in a direction that is not parallel to the line, there's no way to get to the zero vector without adding *zero* times both vectors. See figure 2.1.

Another way to think of independence: a vector is not independent if it is a linear combination of the preexisting vectors.

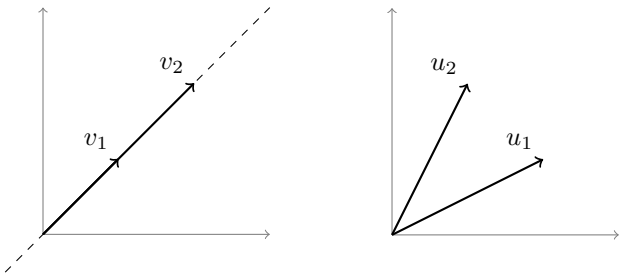

Figure 2.1: Examples of linear independence. There exist a combination of v_1 and v_2 that leads to the zero vector, whereas no combination of u_1 and u_2 leads to the zero vector.

2.5.1 Showing Independence

The requirement for linear independence is that the linear combination

$$c_1 v_1 + c_2 v_2 + \cdots + c_i v_i = 0$$

for no combination other than $c_1, \ldots, c_i = 0$. Therefore we can demonstrate that a set of vectors v_i is independent if we can show that, for the above linear combination, $c_i = 0$.

Example: We can show that $v_1 = \langle 1, 0, 0 \rangle$, $v_2 = \langle 1, 1, 0 \rangle$, and $v_3 = \langle 1, 1, 1 \rangle$ are independent by setting their linear combinations equal to zero:

$$c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0$$

Performing the scalar multiplications by c_i and adding the vectors, we get

$$\begin{bmatrix} c_1 + c_2 + c_3 \\ c_2 + c_3 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

We can treat this as a system of equations where $c_3 = 0$. Solving further reveals that c_1 and c_2 are also zero, and thus that the only linear combination that produces the zero vector is one where $c_i = 0$. Therefore these vectors are linearly independent.

Note that this is equivalent to setting up a matrix the columns of which are the vectors. As a matter of fact, if this matrix's nullspace is larger than the set $\{0\}$, we can conclude the column vectors of which it is comprised are not linearly independent.

2.6 Basis

The basis for a vector space is a set of vectors that are independent but which span the entire space. The basis is a balancing act between having few enough vectors that none are dependent, but enough vectors so that the entire space is covered.

As an example, let's find a basis for $\mathbb{R}^3$. A general strategy is to choose a vector with 1 in one component and zero in the others, then do this for as many vectors as necessary to fill the space (to have vectors which "point" in all the necessary directions). For $\mathbb{R}^3$, a basis—the **standard basis**—is

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

though there are infinitely many other valid bases. Note that these are the columns of the 3 by 3 identity matrix!

There can be many different bases, but there will always be the same number of vectors in a basis. We cannot add a fourth vector to a basis of $\mathbb{R}^3$ which is not a linear combination of the preexisting vectors.

The number of vectors in basis for a space is the **dimension** of that space!

2.7 The Four Fundamental Subspaces

For an m by n matrix, there are four distinct subspaces we can analyze:

1. The **column space**. This has been extensively discussed in section 2.3.
2. The **nullspace** of solutions to $Ax = 0$. This has also been extensively discussed in section 2.4
3. The **row space**, or the column space of A^T.
4. The **left nullspace**, or the nullspace of A^T.

Some properties of these subspaces are the same for A and its reduced counterpart R. As eliminations consist of scalar multiplications and additions between rows, the row spaces of A and R are the same. The column spaces are different but have the same dimension, as the rank of the matrix remains the same. The nullspace remains the same. The left nullspace has the same dimension.

2.7.1 Finding the Left Nullspace

The left nullspace with have more vectors than just the zero vector if some combinations of the rows yields all zeroes. This is the same as with the typical nullspace, only with rows instead of columns.

We take advantage of elimination! If some combination during elimination yielded a row of all zeroes, we know that combination is a vector in the left nullspace. Using the augmented matrix $[\,A\ \ I\,]$ to keep track of eliminations, a basis for the left nullspace will be the row vector to the right of the zero row:

$$\left[\begin{array}{ccc|cc}1&2&4&1&0\\2&4&8&0&1\end{array}\right] \rightarrow \left[\begin{array}{ccc|cc}1&2&4&1&0\\0&0&0&-2&1\end{array}\right]$$

Clearly, the combination of one of the second row plus -2 times the second row yielded zero. Evidently $(-2, 1)$ is a special solution.

2.7.2 The Fundamental Theorem of Linear Algebra, Part. 1

We now take a look at the relationship between a matrix's rank and its subspaces.

A matrix's **column space** is of dimension r. Rank is synonymous with linear independence: if a column of a matrix is not linearly independent, it will become a free column. Thus, the column space will have a basis with as many vectors as there are linearly independent pivot columns.

A matrix's **nullspace** is of dimension $n - r$. The free columns of R determine how many special solutions we will have in the nullspace, or how many vectors there are in the nullspace's basis.

A matrix's **row space** is of dimension r. If $r < m$, or the number of rows, then there will be some zero rows. These rows contribute nothing to the dimension of the row space.

Finally, a matrix's **left nullspace** is of dimension $m - r$. We can either think of this as being the number of free columns in A^T, or the number of zero rows in R.

Here is a succinct summary of the above ideas:

The column space and row space both have dimension r. The left and right nullspaces have dimensions $n - r$ and $m - r$.

2.7.3 Rank One Matrices

Here's a fun property of rank one matrices. For the following rank one matrix, the column space has a basis of $(1, 2)$ and the row space has a basis $(1, 4, 5)$. We can then express the matrix itself as a multiplication of these two vectors!

$$\begin{bmatrix} 1 & 4 & 5 \\ 2 & 8 & 10 \end{bmatrix} \rightarrow \begin{cases} \text{Basis for col. space: } (1,2) \\ \text{Basis for row space: } (1,4,5) \end{cases} \rightarrow \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \end{bmatrix}$$

2.8 Matrix Vector Spaces

The standard basis is composed of matrices with a 1 in one position and 0's elsewhere. It is easy to see that matrices of this form are independent and span all M3. So to show both (a) and (b) it suffices to show that all of these matrices are in the space spanned by either invertible matrices or rank one matrices.

2.9 Infinite-Dimensional Vector Spaces

The set of all polynomials with degree at most m and coefficients in $\mathbb{F}$ is denoted $\mathcal{P}_m(\mathbb{F})$. This is a finite dimensional vector space.

For example, a basis for $\mathcal{P}_2(\mathbb{R})$ is $1, x, x^2$. A vector in this vector space is any

$$1a_0 + a_1x + a_2x^2$$

The set of all polynomials with any arbitrary degree is $\mathcal{P}(\mathbb{F})$. For any x^m, there is an x^{m+1}. Thus, no list of $1, \ldots, x^m$ spans this vector space, and $\mathcal{P}(\mathbb{F})$ is infinite-dimensional.

2.10 Affine Subsets and Quotient Spaces

An **affine subset** of V is a linear subspace "transposed" by addition of a vector v. That is, given a subspace $U \subset V$, an affine subset is $\{v + u : u \in U\}$ for an arbitrary $v \in V$.

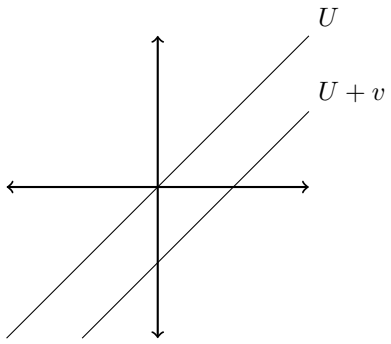

Figure 2.2: The subspace U and the affine subset $U + v$.

The quotient space V/U is the set of all affine subsets of V:

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

2.11 Application: Incidence Matrices

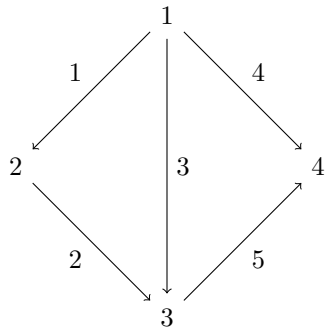

Figure 2.3: A graph with numbered edges and nodes.

An incidence matrix can be used to describe a mathematical *graph* with edges and nodes. The rows of the matrix are edges and the columns are nodes. The edge leads into a node if its value at that column is positive one. It "exits" a node if its value is -1. The graph in figure 2.3, for example, can be described by

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ x_3 - x_1 \\ x_4 - x_1 \\ x_4 - x_3 \end{bmatrix}$$

where x_n are the potentials at each node, the right side giving the potential difference at each edge. If we look at each column in isolation, we can see what edges lead in and out of a node. Column/node one, for example, is $(-1, 0, -1, -1, 0)$, showing that edges one, three, and four lead out of node one.

We can interpret the **subspaces of incidence matrices** to derive some meaningful results.

1. **Nullspace:** these are solutions to $Ax = 0$; in other words, they're the potentials at each node so that all the potential differences are zero.
2. **Row space:** these are all the combinations of the rows, i.e. you're adding the edges of the graphs. After r edges, there's no way to choose another independent edge, meaning that there's no way to *not* form a loop after r edges.
3. **Column space:** All combinations of the columns.
4. **Left Nullspace:** Contains solutions to $A^T y = 0$. We find that the solutions to the left nullspace are loops. This is where we obtain Kirchhoff's voltage law for circuits: the sum of voltages around any closed loop must be zero.

2.11.1 Euler's Formula for Graphs

As the left nullspace has dimension $m - r$, we can assert that there are $m - r$ independent loops. As there are $r = n - 1$ independent rows (there will always be more edges than nodes, and thus not all edge rows are linearly independent), we conclude that, for any graph,

$$m - r = m - n + 1$$

where we plug $n - 1$ into r. In words:

$$(\text{number of nodes}) - (\text{number of edges}) + (\text{number of small loops}) = 1$$

2.11.2 Adding an Outside Source

We can use Ohm's Law to obtain some more results from our incidence matrix. Ohm's Law states that the current is proportional to the voltage and inversely proportional to the resistance: $I = V/R$. We express the inverse of resistance as *conductance*: $C = 1/R$. In matrices, this is $y = -CAx$ where Ax gives potential differences and C is the conductance. Combining this with Kirchhoff's law $A^T y = 0$, we get

$$A^T CAx = 0$$

With an external current source f this becomes

$$A^T CAx = f$$

.

If the conductences are all one, $C = I$.

Example: For the following incidence matrix, find the currents along the six edges if a current source f goes into node 1. Take the conductance across each edge to be one.

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

With $C = I$, A^TCAx becomes A^TAx, which is

$$\begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} S \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

This matrix has rank three. We must find a particular solution by setting an x equal to zero; in this case we choose $x_4 = 0$. This yields

$$x = \begin{bmatrix} S/2 \\ S/4 \\ S/4 \end{bmatrix}$$

Which we can then plug into Ohm's Law $y = -CAx$ to find the currents.

2.12 Exam Review Examples

$$C = \begin{bmatrix} U & U \\ U & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} U & U \\ 0 & -U \end{bmatrix} \Rightarrow \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}$$

For a matrix A with

$$Ax = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

and complete solutions

$$x = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + c \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + d \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

3 columns, dimension of nullspace is two, rank is one. If A times the above complete solution's particualr solution, first column is $(1, 2, 1)$. From nullspace we can determine the other two columns. The sum of the two first columns is the zero vector, and the last column alone is the zero vector. Therefore

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

For what right-hand side b can A be solved? Don't just claim that this can be solved if b is in the column space. In other words, we are asking what the column space of the matrix is. In this case, the column space is

$$b = c \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

EXAMPLE For a matrix B, if $B^2 = 0$ is B zero? Not necessarily.

Example: For a square b by n matrix with independent columns, is $Ax = b$ always solvable? Yes. The rank is n and therefore the column space fills the entirety of $\mathbb{R}^n$. For a rectangular matrix? Not always. We could have a matrix $\begin{bmatrix} I \\ 0 \end{bmatrix}$.

Example: For the matrix B

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -7 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Basis for $N(B) \subseteq \mathbb{R}^4$. Note that for the product CD if C is invertible the nullspace of $N(CD)$ is $N(D)$ because you can always do the inverses steps of C. Complete solutions to $Bx = (1, 0, 1)$? Well $(1, 0, 1)$ is the first column of B, therefore our particular solution could be $(1, 0, 0, 0)$.

Example: If A and B have the same four subspaces, is $A = cB$? Not necessarily. A and B could be any invertible 6 by 6 matrices. Their ranks, however, must be the same.

Example: Why can't $v = (1, 2, 3)$ be in a nullspace and be a row? This is a question of orthogonality!

The rank of AB is not greater than the rank of A or the rank of B. This is because B can be thought of as taking linear combinations of the rows of A.

If you find, by quick analysis, a basis for the nullspace, you don't need to solve $Ax = 0$; by definition all combinations of that basis will suffice.

Chapter 3

Orthogonality

3.1 Orthogonality

Two vectors are orthogonal if they form a 90° angle. Using the Pythagorean theorem, we can claim that two orthogonal vectors x and y have magnitudes such that

$$|x|^2 + |y|^2 = |x + y|^2$$

but we know that the inner product $x^T x$ (or the dot product of x with itself) is $x^T x = |x|^2$. We substitute those inner products, obtaining

$$\begin{aligned} x^T x + y^T y &= (x+y)^T (x+y) \\ x^T x + y^T y &= \left(x^T + y^T\right)(x+y) \\ x^T x + y^T y &= x^T x + y^T y + x^T y + x^T y \end{aligned}$$

We can cancel out $x^T x + y^T y$ and sum the $x^T y$, obtaining

$$\begin{aligned} 0 &= 2x^T y \\ 0 &= x^T y \end{aligned}$$

proving that two orthogonal vectors will have an inner product $x^T y = 0$.

Two vectors x and y are orthogonal if their inner product is $x^T y = 0$.

By this theorem, the zero vector is perpendicular to every vector.

3.1.1 Orthogonal Subspaces

A subspace S is orthogonal to a subspace T if every vector in S is orthogonal to every vector in T. For example, two planes that form a 90° angle are *not* orthogonal subspaces because there are vectors in both planes that are parallel to each other.

In contrast, a plane and a line perpendicular to that plane *are* orthogonal subspaces. They only share the zero vector, and the zero vector is orthogonal to itself, as $0^T 0 = 0$.

If the columns of A are a basis for a subspace orthogonal to the subspace formed by x, we see that

$$\begin{bmatrix} \text{row } 1 \cdots \\ \text{row } 2 \cdots \\ \text{row } 3 \cdots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \end{bmatrix}$$

where it's easy to see that the above matrix multiplication involves the dot product of the rows of A and x. We also arrive at an important property of subspaces of a matrix, the **fundamental theorem of linear algebra, part two**:

> The row space is orthogonal to the null space, and
> the column space is orthogonal to the left null space.

We also claim that two orthogonal spaces in $\mathbb{R}^n$ are **complements** if $\dim(S_1) + \dim(S_2) = n$. The left null space and the column space are complements.

3.2 Projections

3.2.1 Projections onto Vectors

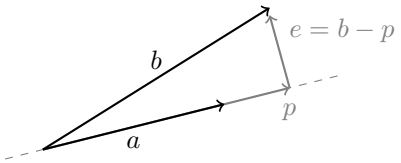

Figure 3.1: b, a, the projection vector p, and the error vector e.

Consider projecting a vector b onto the line formed by a vector a. We term the projection p, and the error (the difference between the vector and the projection) $e = b - p$. The vector e will always be perpendicular to the projection vector. We take $p = xa$, where the projection is just a scaled by a factor x. The dot product of a and the error e is

$$\begin{aligned} a^T(b - xa) &= 0 \\ xa^Ta &= a^Tb \\ x &= \frac{a^Tb}{a^Ta} \end{aligned}$$

And therefore the projection of b onto a is

$$p = a \cdot \frac{a^Tb}{a^Ta}$$

Notice the parallel between this formula for projection and the dot product formula:

$$a \cdot \frac{a^Tb}{a^Ta} \Rightarrow \vec{a} \cdot \frac{\vec{a} \cdot \vec{b}}{\vec{a} \cdot \vec{a}} \Rightarrow \vec{a} \cdot \frac{\vec{a} \cdot \vec{b}}{|\vec{a}|^2} \Rightarrow \hat{a}\,|\vec{a}| \cdot \frac{\vec{a} \cdot \vec{b}}{|\vec{a}|^2} \Rightarrow \hat{a} \cdot \frac{\vec{a} \cdot \vec{b}}{|\vec{a}|}$$

3.2.2 Projections onto Subspaces

$Ax = b$ may have no solutions because b is not in A's column space. However, we can project b onto that column space and solve for the projection: $A\hat{x} = b$. p is some linear combination of the bases of the plane,

$$p = \hat{x}_1 a_1 + \hat{x}_2 a_2 + \cdots \Rightarrow p = A\hat{x}$$

now we just find the $\hat{x}$ that leads to that projection. From $e = b - p$ we know that $b - A\hat{x}$ is perpendicular to A's column space:

$$a_1^T(b - A\hat{x}) = 0,\ a_2^T(b - A\hat{x}) = 0, \ldots$$

Where a_i are A's columns. We then put these transposed columns into the transposed matrix A^T

$$\begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \end{bmatrix} (b - A\hat{x}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \end{bmatrix} \ \Rightarrow\ A^T(b - A\hat{x}) = 0$$

Solving for $\hat{x}$:

$$\hat{x} = \left(A^T A\right)^{-1} A^T b$$

At last, we multiply by A, as $\hat{x}$ is not the projection but a solution to $A\hat{x} = p$:

$$p = A\hat{x} = A\left(A^T A\right)^{-1} A^T b$$

which, in fact, resembles our formula for projection onto a vector, $aa^T/a^T a$. The equation we will often solve is then

$$A^T A\hat{x} = A^T b$$

In summary, a **projection matrix** P is

$$P = A\left(A^T A\right)^{-1} A^T$$

3.3 Least Squares

One application of projection matrices is least squares interpolation. In least squares problems, we have an equation $C + Dt = y$ for which not all given (t, y) are solutions. For example we can translate the following system into a matrix equation:

$$\begin{cases} C + 1D = 1 \\ C + 2D = 2 \\ C + 3D = 2 \end{cases} \Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

This is an $Ax = b$ with no solution! We must project b onto the column space of A to obtain a $\hat{x}$ which is closest to our desired b. In other words, we must solve $A^T A\hat{x} = A^T b$.

$$A^TA = \begin{bmatrix}1 & 1 & 1\\ 1 & 2 & 3\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 2\\ 1 & 3\end{bmatrix} = \begin{bmatrix}3 & 6\\ 6 & 14\end{bmatrix}$$

$$A^Tb = \begin{bmatrix}1 & 1 & 1\\ 1 & 2 & 3\end{bmatrix}\begin{bmatrix}1\\ 2\\ 2\end{bmatrix} = \begin{bmatrix}5\\ 11\end{bmatrix}$$

$$\begin{bmatrix}3 & 6\\ 6 & 14\end{bmatrix}\begin{bmatrix}\hat{C}\\ \hat{D}\end{bmatrix} = \begin{bmatrix}5\\ 11\end{bmatrix} \Rightarrow \hat{x} = \begin{bmatrix}2/3\\ 1/2\end{bmatrix}$$

3.3.1 Least Squares With Polynomials

Occasionally we might want to perform a least squares interpolation on a non-linear phenomenon. We may have some system such as

$$\begin{cases}C + Dt_1 + Et_1^2 = y_1\\ \vdots\\ C + Dt_n + Et_n^2 = y_n\end{cases} \Rightarrow \begin{bmatrix}1 & t_1 & t_1^2\\ \vdots & \vdots & \vdots\\ 1 & t_n & t_n^2\end{bmatrix}\begin{bmatrix}C\\ D\\ E\end{bmatrix} = \begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$$

where we simply use $A^TA\hat{x} = A^Tb$ once again!

Example Fit a parabola to the data points $(-1, 0)$, $(-1, 0)$, $(0, 1)$, $(1, 0)$, $(2, 0)$. We begin by expressing this system $C + Dt_n + Et_n^2 = y_n$ as

$$\begin{bmatrix}1 & -2 & (-2)^2\\ 1 & -1 & (-1)^2\\ 1 & 0 & 0^2\\ 1 & 1 & 1^2\\ 1 & 2 & 2^2\end{bmatrix}$$

$$A^TA = \begin{bmatrix}5 & 0 & 10\\ 0 & 10 & 0\\ 10 & 0 & 34\end{bmatrix}, \quad A^Tb = \begin{bmatrix}1\\ 0\\ 0\end{bmatrix}$$

$$\begin{bmatrix}5 & 0 & 10\\ 0 & 10 & 0\\ 10 & 0 & 34\end{bmatrix}\begin{bmatrix}C\\ D\\ E\end{bmatrix} = \begin{bmatrix}1\\ 0\\ 0\end{bmatrix}$$

Leading us to coefficients $C = 34/70$, $D = 0$, $E = 10/70$, and a polynomial $(34/70) + (10/70)t^2 = y$.

3.4 Orthogonal Bases/Matrices

A basis with vectors $q_1, q_2, \ldots q_n$ is **orthonormal** if

$$q_i^Tq_j = \begin{cases}0 \text{ if } i \neq j\\ 1 \text{ if } i = j\end{cases} \tag{3.1}$$

i.e., every vector is of length one, and every vector is orthogonal to all other vectors.

An **orthogonal matrix** Q is a matrix with orthonormal columns. If Q is also square, $Q^TQ = I$ by equation 3.1; in other words $Q^{-1} = Q^T$ for square matrices.

The operation Qx **does not change the length** of x. Therefore, orthogonal matrices are isometries (section 8.6). Proof: take the dot product of the vector Qx: $(Qx)^T(Qx) = x^TQ^TQx$. By $Q^TQ = I$, we then get $(Qx)^T(Qx) = x^Tx$. Therefore our vectors Qx and x have the same length.

Examples of orthogonal matrices include any permutation matrices, and the rotation matrix

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

Note the projection matrix for a square orthogonal matrix Q is simply I; as the basis must span the whole space $\mathbb{R}^n$, the projection is simply the vector in question.

We will later realize that orthogonal matrices have **eigenvalues with absolute values one**. Proof: take the lengths of the vectors on both sides:

$$\begin{aligned} |Qx| &= |\lambda||x| \\ |x| &= |\lambda||x| \\ 1 &= |\lambda| \end{aligned}$$

where we use the fact that Qx has the same length as x.

3.5 Projecting onto Orthogonal Bases

We project onto the subspace spanned by an orthogonal basis; the orthogonality of the basis can be leveraged, and we can find the projection without using a projection matrix.

The vector we are projecting, w, can be expressed as a linear combination of an orthogonal basis:

$$w = c_1v_1 + c_2v_2 + c_3v_3$$

We would like to find the projection onto the subspace spanned by v_1, v_2. We know v_1 and v_2—they are givens. Take the dot product of both sides with v_1.

$$v_1^Tw = c_1v_1^Tv_1 + c_2v_1^Tv_2 + c_3v_1^Tv_3$$

With orthogonality, we know $v_1^Tv_2$ and $v_1^Tv_3$ are both zero. Thus,

$\frac{v_1^Tw}{v_1^Tv_1} = c_1$

The same can be done for v_2. Finding c_1 and c_2, it just remains to add $c_1v_1 + c_2v_2$ to find the projection

Example Find the projection of $v = (1, 3, 5, 7)$ onto the orthogonal basis $v_1 = (1, 1, 1, 1), v_2 = (-1, 0, 1, 0)$.

$$c_1 = \frac{v_1^Tw}{v_1^Tv_1} = \frac{16}{4} = 4$$

$$c_2 = \frac{v_2^Tw}{v_2^Tv_2} = \frac{4}{2}$$

Thus, the projection is

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

3.6 Gram-Schmidt Process

The Gram-Schmidt process is a method for orthonormalizing a matrix, or a set of vectors. The orthonormalized result *spans the same subspace*. The key idea is leveraging our error vector $e = b - p$. We know e is perpendicular to a projection; all that remains is subtracting $b - p$ and normalizing the result.

Let us illustrate the technique with two vectors a and b. We denote the orthogonalized vectors a' and b'. Setting $a = a'$, we use

$$\begin{aligned} b' &= b - p \\ b' &= b - \frac{a'a'^T}{a'^T a'}b \end{aligned}$$

Then we normalize the vectors by

$$q_1 = \frac{a'}{|a'|} \quad q_2 = \frac{b'}{|b'|}$$

EXAMPLE Orthonormalize the vectors $a = \langle 1,1,1\rangle$, $b = \langle 1,0,2\rangle$, and $c = \langle 2,0,0\rangle$. We set $a' = a$ and

$$b' = b - \frac{a^T b}{a^T a}a \Rightarrow \begin{bmatrix}1\\0\\2\end{bmatrix} - \frac{3}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}0\\1\\-1\end{bmatrix}$$

$$c' = c - \frac{a^T c}{a^T a}a - \frac{b'^T c}{b'^T b'}b' \Rightarrow \begin{bmatrix}2\\0\\0\end{bmatrix} - \frac{2}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} - 0\begin{bmatrix}0\\1\\-1\end{bmatrix} = \begin{bmatrix}4/3\\-2/3\\-2/3\end{bmatrix}$$

Where c' is obtained by subtracting its projections onto b' and a', thereby obtaining a vector perpendicular to the other two vectors. Normalizing these vectors,

$$q_1 = \frac{a'}{|a'|} = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}$$

$$q_2 = \frac{b'}{|b'|} = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\-1\\1\end{bmatrix}$$

$$q_3 = \frac{c'}{|c'|} = \frac{\sqrt{3}}{2\sqrt{2}}\begin{bmatrix}4/3\\-2/3\\2/3\end{bmatrix}$$

Putting these vectors into a matrix, we obtain the orthogonal matrix

$$Q = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & \sqrt{\frac{2}{3}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{bmatrix}$$

We can check our work by verifying that $Q^TQ = I$.

3.6.1 Factorization A=QR

We can bring an orthogonalized matrix back to its original A by $A = QR$, where R happens to be a triangular matrix.

EXAMPLE Below is a matrix and its Gram-Schmidtted counterpart.

$$A = \begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix} \quad Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

How can we "restore" Q to A? The original vectors are the columns of A. By $a' = a$, we know we simply normalized a, and its direction wasn't changed; therefore, we multiply by $|a|$.

To bring b back, we must likewise un-normalize it and multiply it by $|b'|$, but we must also add the projection of b onto a that we subtracted when performing Gram-Schmidt. We do this by bringing a back to its original magnitude, hence $|a|$, then multplying a by the projection of b. These two then multiply the first column of Q, the normalized a. Putting these steps into a matrix,

$$R = \begin{bmatrix} |a| & |a| \cdot \text{proj}_a b \\ 0 & |b'| \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix}$$

Where a^Tb/a^Ta is 2, $|a|$ is $\sqrt{2}$, and $|b|$ is $2\sqrt{2}$. In conclusion,

$$\begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix}$$

3.7 Cauchy-Schwarz Inequality

Suppose $u, v \in V$. Then

$$|\langle u, v \rangle| \leq |u||v|$$

where $\langle u, v \rangle$ is the inner product of u and v. $|\langle u, v \rangle| = |u||v|$ if and only if (3.2) is an equality.

Proof: Perform an orthogonal decomposition (Gram-Schmidt process):

$$u = \frac{\langle u, v \rangle}{|v|^2} v + w$$

where w is orthogonal to v. By the Pythagorean theorem,

$$|u|^2 = \left| \frac{\langle u, v \rangle}{|v|^2} v \right|^2 + |w|^2$$

The $|v|^2$ on the first term cancel, giving us

$$|u|^2 = \frac{|\langle u, v\rangle|^2}{|v|^2} + |w|^2$$

As $|u|^2$ is $|\langle u, v\rangle|^2$ plus some quantity, we can conclude that

$$|u|^2 \geq \frac{|\langle u, v\rangle|^2}{|v|^2} \tag{3.2}$$

which is only an equality if $w = 0$ (i.e., u is a multiple of v). We then multiply both sides by $|v|^2$ and take square roots to obtain the inequality.

3.8 Triangle Inequality

This theorem has the following geometric interpretation: the length of each side of a triangle is less than the sum of the lengths of the two other sides.

Suppose $u, v \in V$. Then

$$|u + v| \leq |u| + |v|$$

Proof:

$$\begin{aligned}
|u+v|^2 &= \langle u+v, u+v\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \langle v, u\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \overline{\langle v, u\rangle} \\
&= |u|^2 + |v|^2 + 2\mathrm{Re}\langle u, v\rangle \\
&\leq |u|^2 + |v|^2 + 2|\langle u, v\rangle| \\
&\leq |u|^2 + |v|^2 + 2|u||v| \\
&\leq (|u| + |v|)^2
\end{aligned}$$

3.9 Orthogonal Complements

If U is a subset of V, then the orthogonal complement of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$\{v \in V : \langle v, u\rangle = 0 \text{ for every } u \in U\}$$

Some properties of $U^\perp$:

- U is a subset of V, $U^\perp$ is a subspace of V.
- $\{0\}^\perp$ is V.
- $V^\perp = \{0\}$.
- If U is a subset of V, then $U \cap U^\perp \subset \{0\}$.

- If U and W are subsets of V, and $U \subset W$, then $W^\perp \subset U^\perp$.
- $\dim U^\perp = \dim V - \dim U$
- $U = \left(U^\perp\right)^\perp$

Additionally, $V = U \oplus U^\perp$. I.e., a subspace and its orthogonal complement form the entire vector space.

We can define projections in terms of orthogonal complements. Take $U \subset V$. The orthogonal projection is the operator $P_U \in \mathcal{L}(V)$ defined as: for $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then $P_U v = u$. We only keep the component *in* the subspace U.

3.10 Application: Fourier Series

Some background: we can separate a vector v into an orthonormal basis times some constants. Any v can become

$$v = x_1 q_1 + x_2 q_2 + \cdots + x_n q_n$$

If we were to get take the product of an arbitrary member of the orthonormal basis, say, q_1, with every one of these vectors, we would get

$$q_1^T v = x_1 q_1^T q_1 + \cdots + x_n q_1^T q_n$$

By orthonormality, every one of these terms becomes 0 except the dot product of q_1 with itself, which becomes 1. Thus we get

$$q_1^T v = x_1$$

Now we transition into function spaces. To take the dot product of two functions, we must add the products of the corresponding "components" of the functions. We can multiply two functions together, of course, but since functions are continuous, we "add" these products by taking the integral:

$$f^T g = \int f(x) g(x) \, \mathrm{d}x$$

Functions which exhibit "orthonormality" are $\cos x$, $\sin x$, $\cos 2x$, $\sin 2x$, so on. Just as we separated a v above, we separate an $f(x)$ into

$$f(x) = a_0 + a_1 \cos x + b_2 \sin x + \cdots + a_n \sin nx + b_n \cos nx$$

Also just as above, we take the dot product of one of these functions with every other function. Let us choose $\cos x$.

$$\begin{aligned}\int_0^{2\pi} f(x) \cos x \, \mathrm{d}x = a_0 \int_0^{2\pi} \cos x \, \mathrm{d}x + a_1 \int_0^{2\pi} (\cos x)^2 \, \mathrm{d}x \\ + b_1 \int_0^{2\pi} \sin x \cos x \, \mathrm{d}x + a_2 \int_0^{2\pi} \cos x \cos 2x \, \mathrm{d}x + \cdots\end{aligned}$$

All terms but the second become zero, a fact which you can check for yourself. We can then solve for the Fourier coefficient by

$$\int_0^{2\pi} f(x)\cos x\,\mathrm{d}x = a_1 \int_0^{2\pi} (\cos x)^2\,\mathrm{d}x$$

Chapter 4

Determinants

4.1 Properties of Determinants

Following Gilbert Strang's pedagogy, we introduce the properties of determinants before the formula for determinants or even what the determinant is. We denote the determinant either by "det$A$" or $|A|$.

1. $\det I = 1$.
2. Exchanging rows reverses the sign of the determinant.
3. Multiplying a row by a scalar multiplies the whole determinant.

$$\begin{vmatrix} 2a & 2b \\ c & d \end{vmatrix} = 2 \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

4. Determinants are linear in each row.

$$\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

5. Two equal rows $\rightarrow \det A = 0$. I.e., singular matrices have determinant zero.
6. Linear combinations of rows don't change the determinant; if we subtract c (row i) from row k, the determinant remains the same.
7. Row of zeroes $\rightarrow \det A = 0$.
8. The determinant of a triangular matrix is the product of the pivots $\prod d_i$. As taking linear combinations of rows doesn't change matrix, we can turn a triangular matrix into a diagonal matrix. By properties 1 and 3, this is just I times the diagonals, or the the product of the diagonals!

$$\begin{vmatrix} d_1 & \cdot & \cdot \\ 0 & d_2 & \cdot \\ 0 & 0 & d_3 \end{vmatrix} \Rightarrow \begin{vmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{vmatrix} \Rightarrow d_3 d_2 d_1 \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

9. $\det(AB) = \det A \cdot \det B$
10. From the above property we obtain $\det A^{-1} = 1/\det A$:

$$\det(A^{-1}A) = \det I \;\Rightarrow\; \left(\det A^{-1}\right)(\det A) = 1$$

11. $\det(cA) = c^m \det A$, as every row is multiplied by c. Think volume!
12. $\det A^T = \det A$. *All properties for rows are true for columns.* Prove this factorizing into LU

$$\begin{aligned} \det A^T &= \det A \\ \left|U^T L^T\right| &= |LU| \\ \det U^T \cdot \det L^T &= \det L \cdot \det U \end{aligned}$$

We know $|L|$ is one as its diagonals are 1. We know the diagonals of U and U^T are the same, therefore

$$\det A^T = \det A \Rightarrow \det U^T = \det U$$

4.2 Formula for the Determinant

Using the above properties of the determinant, we can attempt to derive a formula. Using property 4, we split a generic 2 by 2 matrix into

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix}$$

where we can take the diagonals of these to obtain a sum of determinants. The first and last matrices have diagonals with products zero, the second is ad, and the third, after a row exchange, is bc.

Generalizing this method to three-by-three matrices, we notice that the only matrices with nonzero determinants are ones that can be row exchanged into a diagonal matrix—i.e., "permutation" configurations:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \cdots$$

Some of these permutations must undergo row exchanges to be put in a diagonal form. For odd numbers of row exchanges, the terms become negative; for even numbers of row exchanges, the (-1)s become $+1$. This leads to the pattern seen in the cofactor matrix in section 4.3.

Generalizing, the formula becomes

$$\sum \pm a_{1\alpha} a_{2\beta} \dots a_{n\omega}$$

where $\alpha \neq \beta \neq \cdots \neq \omega$.

4.3 Cofactor Formula for the Determinant

From our lengthy sum of terms $a_{11}a_{22}a_{33} + \cdots = \det A$, we group like terms together and obtain

$$\det A = a_{11}a_{22}a_{33} + \cdots \Rightarrow a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{31}(\dots)$$

where every other term is negative due to the row exchange that must be performed. Upon closer inspection, it becomes clear that this is

$$\sum \pm a_{1i} \cdot \det(\text{smaller matrix})$$

Theses determinants of smaller matrices are termed **cofactors**. They are the determinants of the matrices that result from omitting the row and column of an entry. The cofactor matrix is comprised of these cofactors. Their coefficients follow a checkerboard pattern (arising from the row exchanges that must be performed):

$\det A = a_{11}C_{11} + \cdots + a_{1n}C_{1n}$, where C is the cofactor matrix with coefficients $\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$

4.4 Applications of Determinants

4.4.1 Cofactor Formula for the Inverse

$$A^{-1} = \frac{1}{\det A} C$$

where C is the cofactor matrix.

This is because the multiplication AC^T is

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} c_{11} & \cdots & c_{n1} \\ \vdots & \ddots & \vdots \\ c_{1m} & \cdots & c_{mn} \end{bmatrix}$$

Where the diagonals of the product are just the determinant: $a_{11}c_{11} + a_{12}c_{12} + \ldots$ hence our division by $\det A$. Any other entry is zero, as that would be the determinant of a matrix with two identical rows.

REMARK C^T is called the adjugate or classical adjoint matrix.

4.4.2 Cramer's Rule

Given $Ax = b \Rightarrow x = A^{-1}b$, we claim that

$$x = \frac{1}{\det A} C^T b$$

$C^T b$ is the cofactors of A times the entries of b:

$$x = C^T b = \begin{bmatrix} c_{11} & \cdots & c_{n1} \\ \vdots & \ddots & \vdots \\ c_{1m} & \cdots & c_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} b_1c_{11} + \cdots + b_nc_{n1} \\ \vdots \\ b_1c_{1m} + \cdots + b_nc_{mn} \end{bmatrix}$$

From this we can claim that

$$x_1 = \frac{\det B_1}{\det A},\ x_2 = \frac{\det B_2}{\det A},\ \ldots,\ x_n = \frac{\det B_n}{\det A}$$

where B_n is A with the nth column replaced with b.

4.4.3 Volume

$\det A$ gives the volume of a parallelepiped formed by the columns/rows of the matrix (recall that $\det A^T = \det A$). The determinant could be negative, which is the case for a "left-handed" parallelepiped.

The identity matrix, for example, forms a cube of volume one; the unit cube. Any orthonormal matrix, then, is the unit cube, but rotated.

Chapter 5

Eigenvalues/Eigenvectors

5.1 Introduction to Eigenstuff

The equation of this chapter is

$$Ax = \lambda x$$

the key concept being that Ax outputs x but scaled by a λ.

> For $Ax = \lambda x$, x is the **eigenvector**. It is a vector which, when multiplied by A leads to itself up to a scalar. λ is the **eigenvalue**. It is the scalar which multiplies the eigenvector.

We do *not* consider the zero vector an eigenvector, as it would be the eigenvector for every matrix.

As any cx will suffice as an eigenvector for a specific eigenvalue, these qualify as a vector space. We often call this the eigenspace of a specific λ of A, denoted $\mathcal{E}_A(\lambda)$. We will see that this is the nullspace of $(A - \lambda I)$.

If a nonzero vector is in the nullspace of A, then it is an eigenvector for the eigenvalue 0. Thus *whenever there is a* 0 *eigenvalue*, the matrix is singular. Also, if there are k independent vectors with $\lambda = 0$, we know the rank of the matrix is $n - k$.

Example Consider a projection matrix onto a plane. Any x in the plane is an eigenvector with eigenvalue $\lambda = 1$! In addition, any vector perpendicular to the plane is an eigenvector with $\lambda = 0$. Therefore the projection matrix has $\lambda = 0, 1$.

Example Consider a permutation matrix

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

any vector with the same components is an eigenvector with $\lambda = 1$. Any vector $c\langle -1, 1\rangle$ is an eigenvector with $\lambda = -1$.

5.2 Solving for Eigenstuff

$$Ax = \lambda x \;\Rightarrow\; Ax - \lambda x = 0$$

we factor out the common x, leading to

$$(A - \lambda I)x = 0$$

Now, for there to be any eigenvectors, there must be some vector other than the zero vector to lead to $Ax = \lambda x$. Thus, there must be some nonzero x which leads to $(A - \lambda I)x = 0$; recall that this means that $(A - \lambda x)$ is singular. By property 5 of determinants, we know then that

$$\det(A - \lambda I) = 0$$

and we leverage this fact to solve for λ and x.

EXAMPLE Find the eigenvectors and eigenvalues of

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

Setting $\det(A - \lambda I) = 0$,

$$\begin{aligned} \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 - 1 &= 0 \\ \lambda^2 - 6\lambda + 8 &= 0 \\ (\lambda - 4)(\lambda - 2) = 0 & \end{aligned}$$

Giving us $\lambda = 4, 2$. We now plug these λs into $(A - \lambda I)x = 0$ and solve for x:

$$(A - \lambda_1 I)x_1 = 0 \Rightarrow \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} x_1 = 0 \Rightarrow x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$(A - \lambda_2 I)x_2 = 0 \Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x_2 = 0 \Rightarrow x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

EXAMPLE We find the eigenvalues of a 90° rotation matrix.

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

where $\lambda = i$. This makes sense, as rotating a vector twice (i.e., we get $i \cdot i \cdot x = -x$) leads to an 180° rotation, or flipping the vector to $-x$!

REMARK Notice that the λ quadratic from the above examples follows nicely from the sum of the diagonals (the **trace**) and the determinant:

$$\begin{aligned} \begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} &= (a-\lambda)(d-\lambda) - bc \\ &= ad - (a+d)\lambda + \lambda^2 - bc \\ &= ad - bc - (a+d)\lambda + \lambda^2 \end{aligned}$$

It follows that the quadratic expression for λ is

$$\det A - (\text{trace} A)\lambda + \lambda^2$$

5.3 Properties of Eigenvalues

1. If we add cI to A, the eigenvectors remain the same and the eigenvalues become $\lambda + c$. Distributing x to both A and cI:

$$(A + cI)x = \lambda x + cx = (\lambda + c)x$$

2. λ of $A^{-1} = \frac{1}{\lambda \text{ of } A}$
3. $\prod \lambda_i = \det A$
4. $\sum \lambda_i = \sum a_{ii}$. The right side is the sum of the diagonals, called the **trace**.
5. λ of $A^n = \lambda^n$. The eigenvectors remain the same.
6. The eigenvalues of a triangular matrix are the diagonals.
7. If there is an $\lambda = 0$, the matrix is singular. Additionally, the rank of an n by n matrix is n minus the number of independent vectors in the eigenspace of $\lambda = 0$. Therefore

$$\mathcal{E}_A(0) = N(A)$$

5.3.1 Separating into Eigenvectors

For an invertible matrix, the n eigenvectors, being independent, will form a basis for $\mathbb{R}^n$; thus, any vector in the span of the eigenvectors (in this case, any vector $\mathbb{R}^n$) can be expressed as a linear combination of the eigenvectors.

Example For the following A and eigenvectors x_1, x_2,

$$\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}, \quad x_1 = \begin{bmatrix} .6 \\ .4 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \lambda_1 = 1, \quad \lambda_2 = 1/2$$

We express, say $\langle .8, .2 \rangle$ as a linear combination

$$\begin{bmatrix} .8 \\ .3 \end{bmatrix} = \begin{bmatrix} .6 \\ .4 \end{bmatrix} + (0.2) \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Multiplying $\langle .8, .2 \rangle$ by A then multiplies these eigenvector components by their respective eigenvalues:

$$\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \begin{bmatrix} .8 \\ .3 \end{bmatrix} = (1) \begin{bmatrix} .6 \\ .4 \end{bmatrix} + (1/2) \begin{bmatrix} 0.2 \\ -0.2 \end{bmatrix}$$

We call x_1 a steady state because it does not change; i.e., $\lambda_1 = 1$ It's easy to see that multiplying $\langle .8, .2 \rangle$ by A^n will not change the eigenvectors, and the eigenvalues will be multiplied by themselves n times. Thus the result just becomes

$$A^n \begin{bmatrix} .8 \\ .3 \end{bmatrix} = 1^n \begin{bmatrix} .6 \\ .4 \end{bmatrix} + (1/2)^n \begin{bmatrix} 0.2 \\ -0.2 \end{bmatrix}$$

5.4 Diagonalizing a Matrix

Suppose A has n linearly independent eigenvectors. We put them in the columns of a vector S. AS is then $AS = S\Lambda$, where Λ is a diagonal matrix with the eigenvectors.

$$AS = A\begin{bmatrix} \vdots & \vdots & \vdots \\ x_1 & x_2 & x_3 \\ \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \vdots \\ \lambda_1 x_1 & \lambda_2 x_2 & \lambda_3 x_3 \\ \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \vdots \\ \lambda_1 x_1 & \lambda_2 x_2 & \lambda_3 x_3 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = S\Lambda$$

Where multiplying AS is just taking the linear combinations given by Ax_n and putting the results in a matrix; by $Ax = \lambda x$, this is just S but with each column multiplied by an eigenvalue.

We then solve for A:

$$AS = S\Lambda \;\Rightarrow\; A = S\Lambda S^{-1}$$

A new factorization! By property 5 of eigenvalues, we know that

$$A^n = S\Lambda^n S^{-1}$$

and by this we know that $A^n \to 0$ as $n \to \infty$ if all $|\lambda_i| < 1$.

Diagonalizability

λ can be repeated in two different ways:

1. **Geometric multiplicity:** Number of independent eigenvectors of A for a specific λ_i. Dimension of $\mathcal{E}_A(\lambda_i)$ or the nullspace of $(A - \lambda_i I)$.
2. **Algebraic Multiplicity:** Number of repetitions of λ_i among eigenvalues.

If we have two of the same eigenvalue, we can have two independent eigenvectors for those eigenvalues. However, we can also have two of the same eigenvalue and only one eigenvector for that eigenvalue.

> When (Geometric Multip.) < (Algebraic Multip.), there are not enough vectors for the eigenvalues; A is not diagonalizable.

Example We attempt to diagonalize the following matrix:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \begin{vmatrix} -\lambda & 1 \\ 0 & -\lambda \end{vmatrix} = \lambda^2 = 0 \;\Rightarrow\; \lambda = 0$$

The only eigenvector is $\langle 1, 0 \rangle$:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x = 0 \text{ only for } \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The algebraic multiplicity is two. The geometric multiplicity is one. There's only one eigenvector to put into S! So there's no way to obtain a form

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

5.5 Invariant Subspaces

Suppose the linear transformation (see chapter 8) T is an operator on the vector space V. That is, $T \in \mathcal{L}(V)$.

A subspace $U \subset V$ is called invariant under T if $u \in U$ implies $Tu \in U$. That is, the transformation does not take the vector out of the subspace.

If U is a *one-dimensional* subspace invariant under T, then $Tu \in U$ and thus there exists some λ such that $Tv = \lambda v$. So, if v is an eigenvector of T, then $\operatorname{span} v$ is invariant under T.

EXAMPLE The range of T is invariant under T. If $u \in \operatorname{range} T$, then $Tu \in \operatorname{range} T$.

The null space of T is invariant under T. If $u \in N(T)$, then $Tu = 0$, and $0 \in N(T)$.

5.6 Sequence of Increasing Nullspaces

Suppose $T \in \mathcal{L}(V)$. Then

$$\{0\} = N(T^0) \subset N(T^1) \subset N(T^2) \subset \cdots \subset N(T^k)$$

In other words, the null space of T is a subset of the null space of T^2, which is a subset of the null space of T^3, and so on. This is because T could send a vector to its own null space.

If $n = \dim V$, then

$$\{0\} = N(T^0) \subset N(T^1) \subset N(T^2) \subset \cdots \subset N(T^n) = N(T^{n+1}) = \cdots$$

in other words, the null space of this transformation ceases to increase after applied $\dim T$ times.

Additionally, $V = N(T) \oplus \operatorname{range} T^n$.

5.7 Generalized Eigenvectors

We want to be able to decompose an operator's domain into invariant subspaces. However, we can only do this for operators with proper geometric and algebraic multiplicities (i.e., only for diagonalizeable matrices).

In order to be able to decompose any operator in this fashion, we introduce **generalized eigenvectors**. A vector $v \in V$ is called a generalized eigenvector of T corresponding to λ if $v \neq 0$ and

$$(T - \lambda I)^j v = 0,\ j > 0$$

Every generalized eigenvector satisfies this with $j = \dim V$.

We can prove that for $T \in \mathcal{L}(V)$ and distinct eigenvalues $\lambda_1, \ldots \lambda_m$, the generalized eigenvectors $v_1, \ldots v_m$ are linearly independent. These generalized eigenvectors are thus a basis for V and are used in constructing the *Jordan Form* (section 6.5). The matrix of the generalized eigenvectors is the equivalento of the matrix of eigenvectors in $A = S\Lambda S^{-1}$.

5.7.1 Generalized Eigenspace

Let $G_T(\lambda)$ be the set of all generalized eigenvectors corresponding to T, along with the zero vector. Because every eigenvector is a generalized eigenvector, $\mathcal{E}_T(\lambda) \subset G_T\lambda$.

As such, $V = G_T(\lambda_1) \oplus \cdots \oplus G_T(\lambda_m)$, and each $G_T(\lambda_i)$ is invariant under T.

The algebraic and geometric multiplicities of an eigenvalues can be defined (in addition to its definition given in section 5.4) as

$$\begin{aligned}\text{Algebraic multiplicity: } &\dim N((T-\lambda I)^{\dim V}) = \dim G_T(\lambda)\\ \text{Geometric multiplicity: } &\dim N(T-\lambda I) = \dim \mathcal{E}_T(\lambda)\end{aligned}$$

That is, with this generalization, we will have linearly independent generalized eigenvectors for every repetition of λ.

5.8 Application: Differential Equations

EXAMPLE Consider the following **system of differential equations**:

$$\frac{\mathrm{d}u_1}{\mathrm{d}t} = -u_1 + 2u_2, \quad \frac{\mathrm{d}u_2}{\mathrm{d}t} = u_1 - 2u_2$$

We want a function that has an eigen-like property. The insight from calculus is that $\frac{\mathrm{d}}{\mathrm{d}t}e^{\lambda t} = \lambda e^{\lambda t}$. The exponential function is an *eigenfunction* for the operation $\frac{\mathrm{d}}{\mathrm{d}t}$. We therefore set up our system as follows:

$$u = \begin{bmatrix}u_1\\u_2\end{bmatrix} = ce^{\lambda t}\begin{bmatrix}x_1\\x_2\end{bmatrix} \text{ where } u_i \text{ is } c_i e^{\lambda_i t}x_i \text{ and } x_i \text{ is some eigenvector.}$$

$$\begin{cases}\frac{\mathrm{d}u_1}{\mathrm{d}t} = -u_1 + 2u_2,\\ \frac{\mathrm{d}u_2}{\mathrm{d}t} = u_1 - 2u_2\end{cases} \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}u_1\\u_2\end{bmatrix} = \begin{bmatrix}-1 & 2\\1 & -2\end{bmatrix}\begin{bmatrix}u_1\\u_2\end{bmatrix}$$

Finding the eigenvectors and eigenvalues of the coefficient matrix,

$$\lambda_1 = 0 \quad \begin{bmatrix}-1 & 2\\1 & -2\end{bmatrix}x_1 = 0 \text{ for } x_1 = \begin{bmatrix}2\\1\end{bmatrix}$$

$$\lambda_2 = 0 \quad \begin{bmatrix}2 & 2\\1 & 1\end{bmatrix}x_2 = 0 \text{ for } x_2 = \begin{bmatrix}-1\\1\end{bmatrix}$$

Recalling that we can linearly combine eigenvectors, we see that our solution function is

$$u(t) = c_1 e^{\lambda_1}x_1 + c_2 e^{\lambda_2 t}x_2$$

$$\begin{bmatrix}u_1\\u_2\end{bmatrix} = c_1 e^{\lambda_1 t}\begin{bmatrix}2\\1\end{bmatrix} + c_2 e^{\lambda_2 t}\begin{bmatrix}-1\\1\end{bmatrix}$$

Say we're given $u(0) = \langle 1, 0\rangle$. Now we can solve for the constants c_i. By $e^0 = 1$, we get

$$\begin{bmatrix}1\\0\end{bmatrix} = c_1\begin{bmatrix}2\\1\end{bmatrix} + c_2\begin{bmatrix}-1\\1\end{bmatrix}$$

which yields solutions $c_1 = 1/3$ and $c_2 = 1/3$. Therefore our final solution is

$$\begin{bmatrix}u_1\\u_2\end{bmatrix} = \frac{1}{3}e^{0t}\begin{bmatrix}2\\1\end{bmatrix} + \frac{1}{3}e^{-3t}\begin{bmatrix}-1\\1\end{bmatrix} \Rightarrow \begin{cases}u_1(t) = \frac{2}{3} - \frac{1}{3}e^{-3t}\\ u_2(t) = \frac{1}{3} + \frac{1}{3}e^{-3t}\end{cases}$$

Some observations: Notice that our steady state is $u(\infty) = 1/3\,\langle 2, 1\rangle$, as $\lim_{t\to\infty} \frac{1}{3}e^{-3t} = 0$. Also notice that for some differential equations, we might end up with a complex eigenvalue. This would result in a

$$u(t) = c_1 e^{ift} x_1 + c_2 e^{ift} x_2$$

which, per Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$, is a periodic function with period π/f.

For a **homogeneous second order differential equation**, we can split the equation into two and solve it as if it were a system.

$$y'' + by' + cy = 0 \;\Rightarrow\; y'' = -by' - cy$$

$$\begin{cases} y'' = -by' - cy \\ y' = y' \end{cases} \quad \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} y' \\ y \end{bmatrix} = \begin{bmatrix} -b & -c \\ 1 & 0 \end{bmatrix}\begin{bmatrix} y' \\ y \end{bmatrix}$$

and we solve as above.

5.8.1 Exponential of a Matrix

Here is a quick aside. We consider e^{At} where A is a matrix. Evaluate this by substituting A into the Taylor series for e^x:

$$e^{At} = I + At + \frac{(At)^2}{2} + \frac{(At)^3}{6} + \cdots$$

Substitute $A = S\Lambda S^{-1}$:

$$e^{At} = I + S\Lambda S^{-1}t + \frac{\left(S\Lambda S^{-1}t\right)^2}{2} + \frac{\left(S\Lambda S^{-1}t\right)^3}{6} + \cdots$$

using the fact that $\left(S\Lambda S^{-1}\right)^n = S\Lambda^n S^{-1}$,

$$e^{At} = I + S\Lambda S^{-1}t + \frac{S\Lambda^2 S^{-1}t^2}{2} + \frac{S\Lambda^3 S^{-1}t^3}{6} + \cdots$$

and we can easily factor out S and S^{-1}, leaving us with

$$Se^{\Lambda t}S^{-1}$$

where $e^{\Lambda t}$ is a diagonal matrix with entries $e^{\lambda_i t}$.

5.9 Application: Markov Matrices

A Markov matrix is one such that

1. All entries ≥ 0.
2. All columns add to 1.

The eigenvalues and eigenvectors of Markov matrices have the following properties:

1. $\lambda = 1$ is an eigenvalue.
2. All other $|\lambda_i| < 1$
3. $u_k = A^k u_0 \to c_1 \lambda x$ as $k \to \infty$. I.e., any vector will begin approaching the eigenvector after consecutive A^k. All other λ_i approach zero.
4. Eigenvector x_1 has all positive components.

Markov matrices are typically used to represents fractions or probabilities.

EXAMPLE A is a two by two Markov matrix representing the populations of two states, California and Massachusetts. Each entry represents a fraction of the total population. The columns represent states, and the rows represent the amount of people that either move from one state to another, or stay in their respective state. If .9 of residents of California stay in California and .1 move into Mass., and .8 of residents of Mass. stay in Mass. and .2 move to California,

$$\begin{bmatrix} u_{\text{cal}} \\ u_{\text{mass}} \end{bmatrix}_{t=k+1} = \begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix} \begin{bmatrix} u_{\text{cal}} \\ u_{\text{mass}} \end{bmatrix}_k = \begin{cases} u_c(k+1) = .9\, u_c(k) + .2\, u_m(k) \\ u_m(k+1) = .1\, u_c(k) + .8\, u_m(k) \end{cases}$$

Say we begin with 1000 people in Mass. and 0 people in California. Clearly

$$u_1 = Au_0 = \begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix} \begin{bmatrix} 0 \\ 1000 \end{bmatrix} = \begin{bmatrix} 200 \\ 800 \end{bmatrix}$$

Eigenstuffs make it easy to analyze such a problem. A gives the eigenvalues and eigenvectors

$$\lambda_1 = 1, \quad x_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\lambda_2 = .7, \quad x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Separating u_k into eigenvectors then gives us

$$u_k = c_1 1^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 (.7)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

for which we can obtain $c_1 = 1000/3$ and $c_2 = 2000/3$ for our $k = 0$ populations of 0 and 1000. As $k \to \infty$, we see that the second term dies out, and the steady state solution is $\lambda_1 = 1$ and $x_1 = \langle 2, 1 \rangle$. I.e., the populations approach that state over time.

Chapter 6

Selected Matrices and Additional Topics

6.1 Symmetric Matrices

A symmetric matrix is one such that

$$A^T = A$$

which we will see results in a handful of useful properties.

Symmetric matrices have real eigenvalues. We begin by taking a look at possibly complex eigenvalues of a *real* symmetric matrix A (i.e., $\overline{A} = A$). Let us take the complex conjugate of the operation Ax:

$$Ax = \lambda x \;\Rightarrow\; A\overline{x} = \overline{\lambda}\overline{x} \tag{6.1}$$

Then transpose both sides to

$$\overline{x}^T A = \overline{x}^T \overline{\lambda} \tag{6.2}$$

We now take the dot product of the left side of (6.1) with $\overline{x}$, and multiply (6.2) by x.

$$\overline{x}^T Ax = \lambda \overline{x}^T x, \;\; \overline{x}^T Ax = \overline{\lambda}\overline{x}^T x$$

Where we put our λ in the expression doesn't matter as it is a scalar. The left sides of both equations are equal, so we can set the right sides equal to each other.

$$\lambda \overline{x}^T x = \overline{\lambda}\overline{x}^T x \;\Rightarrow\; \overline{\lambda} = \lambda$$

where we divide both sides by $\overline{x}^T x$ to show that $\lambda = \overline{\lambda}$. In words, λ is its own complex conjugate and therefore it is real.

Symmetric matrices have orthogonal eigenvectors. We suppose $Ax = \lambda_1 x$ and $Ay = \lambda_2 y$. Assume $\lambda_2 \neq \lambda_1$. Take dot products of the first equation with the vector y.

$$(\lambda_1 x)^T y = (Ax)^T y = x^T A^T y = x^T Ay$$

recalling that $A^T = A$. But $Ay = \lambda_2 y$, and therefore

$$\lambda_1 x^T y = \lambda_2 x^T y$$

where if $x^T y$ is nonzero, we have $\lambda_2 = \lambda_1$, which contradicts our assumption. Therefore $x^T y$ is zero and a symmetric matrix's eigenvectors are orthogonal.

Since we know a symmetric matrix has real eigenvalues and orthogonal eigenvectors, our diagonalization

$$A = S\Lambda A^{-}1$$

will have orthonormal eigenvector columns in S (provided we normalize the eigenvectors) and thus can be written as

$$A = Q\Lambda Q^T$$

where we use the fact that an orthonormal matrix's inverse is its transpose. This is called the spectral theorem, or the principal axis theorem:

Spectral Theorem: $A = Q\Lambda Q^T$ for symmetric matrices A.

Separating out the rows of Q^T, the diagonal eigenvalues in Λ, and the columns of Q, we can express any symmetric matrix as a sum

$$A = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots + \lambda_n q_n q_n^T$$

Another important property of symmetric matrices that will not be proved here is that a symmetric A will have the same number of positive eigenvectors and pivots, the same number of negative eigenvectors and pivots, and the same number of zero eigenvectors and pivots.

To summarize, here are the **properties of symmetric matrices**:

1. A will have real eigenvalues.
2. A will have orthogonal eigenvectors.
3. $A = Q\Lambda Q^T$.
4. # of positive pivots = # of positive eigenvalues. Same for negative and zero pivots.

Remark Skew-symmetric matrices $A^T = -A$ have all-complex eigenvalues.

6.2 Positive-Definite Matrices

The idea behind positive-definite matrices is that a if one of these properties is true for a *symmetric* matrix, all of them are true:

1. All pivots are positive.
2. All upper left determinants are positive.
3. All eigenvalues are positive.

4. $x^T x$ is positive except at $x = 0$.
5. $A = R^T R$ for some R with independent columns.

What does property 4 really mean? Well, let us work in two dimensions and choose $x = \langle x, y \rangle$:

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} ax + by \\ bx + cy \end{bmatrix} = ax^2 + 2bxy + cy^2$$

The key equation is the quadratic above. If it positive-definite, it is strictly positive save for at $(0, 0)$; i.e., it curls upward. If it is **positive-semidefinite**, there are nonzero x that lead to $x^T Ax = 0$; i.e., $x^T Ax \geq 0$ for all x.

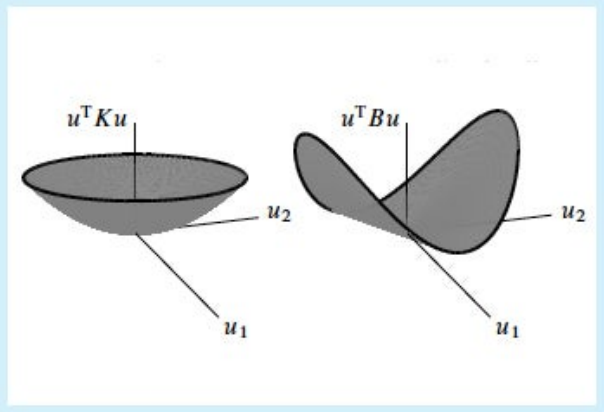

Figure 6.1: $x^T Ax$ for a symmetric positive matrix K and a semidefinite B.

Example Take the following positive definite matrix and find $x^T Ax$.

$$A = \begin{bmatrix} 2 & 6 \\ 2 & 20 \end{bmatrix} \quad x^T Ax = 2x^2 + 12xy + 20y^2$$

Note that we can factor the expression on the right into

$$2(x + 3y)^2 + 2y^2$$

which is a sum of squares; i.e., it is always positive.

6.2.1 Principal Axis Theorem from Positive-Definiteness

We can set $x^T Ax = 1$ for a level curve of this surface. $ax^2 + 2bxy + xy^2 = 1$ is an *ellipse*. Take, for example, the positive-definite

$$\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$$

for which $x^T Ax = 1$ is

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5x^2 + 8xy + 5y^2 = 1$$

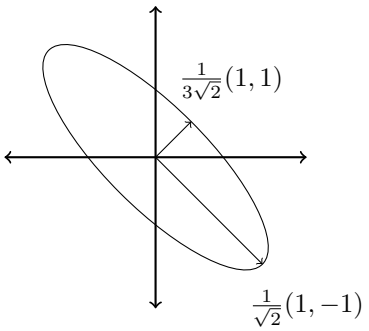

Figure 6.2: The tilted ellipse $5x^2 + 8xy + 5y^2$.

Likewise, we can take the factorization $A = Q\Lambda Q^T$.

$$\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

which, after multiplying $x^T Ax$, will likewise lead us to $5x^2 + 8xy + 5y^2$. We then factor this expression to obtain

$$5x^2 + 6xy + 5y^2 = 9\left(\frac{x+y}{\sqrt{2}}\right)^2 + 1\left(\frac{x-y}{\sqrt{2}}\right)^2 = 1$$

It follows from the formula for an ellipse that the half-lengths of this ellipse's axes are $\sqrt{2}/\sqrt{9}$ and $\sqrt{2}/1$. Thus, generally, an ellipse has axes lengths $1/\sqrt{\lambda_1}$ and $1/\sqrt{\lambda_2}$ (the $\sqrt{2}$'s just serve to normalize the $x + y$ in the numerator).

At last we arrive at the true meaning of the **principal axis theorem**.

$$\begin{bmatrix} x & y \end{bmatrix} Q\Lambda Q^T \begin{bmatrix} x \\ y \end{bmatrix} = \frac{X^2}{\frac{1}{\lambda_1}} + \frac{Y^2}{\frac{1}{\lambda_2}} = 1$$

where X is the first eigenvector as expressed with x and y and Y is the second eigenvector as expressed with x and y.

6.3 Singular Value Decomposition

Our motivation starts with finding an orthonormal basis in the rowspace $v_1, \dots v_n$ and an orthogonal basis in the column space $u_1, \dots u_n$ such that $Av_i = \sigma_i u_i$. We can do this because every vector in the column space comes from one vector in the row space. Putting these vectors in matrices,

$$A \begin{bmatrix} v_1 & \cdots & v_n \\ & & \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_n \\ & & \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \Rightarrow AV = \Sigma U$$

where, using the fact that V is orthonormal and $V^{-1} = V^T$, we obtain

$$A = U\Sigma V^T$$

Extending this notion to a non-invertible matrix,

$$A\begin{bmatrix} v_1 & \cdots & v_r & v_{r+1} & \cdots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & \ddots & & \\ & & & & & 0 \end{bmatrix}$$

where the diagonal σ_i turn to zeroes after we've run out independent vectors in the column space.

The vectors in a singular value decomposition are

- $v_1, \dots v_r$, an orthonormal basis for the row space.
- $u_1, \dots u_r$, an orthonormal basis for the column space.
- $v_{r+1}, \dots v_n$, an orthonormal basis for the nullspace.
- $u_{r+1}, \dots u_n$, an orthonormal basis for the left nullspace.

6.3.1 Obtaining the SVD

V are the eigenvectors of A^TA, U are the eigenvectors of AA^T, and σ_i are the eigenvalues of A^TA, squared. Find U by $u_i = Av_i/\sigma_i$ to ensure the proper u_i.

To show this, let's begin by taking A^TA:

$$A = U\Sigma V^T \;\Rightarrow\; A^TA = V\Sigma^T U^T U\Sigma V^T$$

U is orthonormal, so $U^TU = I$. Σ is diagonal, so $\Sigma^T\Sigma = \Sigma^2$.

$$A^TA = V\Sigma^T\Sigma V^T = V\Sigma^2 V^T$$

As V is orthonormal, this is just the diagonalization $A^TA = S\Lambda S^{-1}$, so the entries of Σ are $\sqrt{\lambda}$. The **singular values are the square roots of the eigenvalues of** A^TA. The exact same process can be carried out for AA^T to find the columns of U; I omit this process here, not just for the sake of space, but because this will yield the eigenvectors with possibly the wrong sign. Instead, you should use $Av_i = \sigma_i u_i$.

Example We take $\left[\begin{smallmatrix} 4 & 4 \\ -3 & 3 \end{smallmatrix}\right]$. Find its singular value decomposition $A = U\Sigma V^T$.

$$A^TA = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix} \text{ has } \lambda = 32, 18, \text{ eigenvectors } \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Normalizing and putting these vectors in V, and taking $\sqrt{\Sigma}$

$$V = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sqrt{32} & \\ & \sqrt{18} \end{bmatrix}$$

We then obtain U using $Av_i = \sigma_i u_i$:

$$u_1 = \frac{1}{\sqrt{32}}\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\frac{1}{\sqrt{2}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$u_2 = \frac{1}{\sqrt{18}}\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}\frac{1}{\sqrt{2}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

which obviously has $\lambda = 32, 18$ (as expected), and eigenvectors $\langle 1, 0\rangle$, $\langle 0, 1\rangle$. Now, putting in the form $A = U\Sigma V^T$:

$$\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{32} & \\ & \sqrt{18} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

6.4 Similar Matrices

Two matrices A and B are similar if there exists some $B = M^{-1}AM$.

All similar matrices have **the same eigenvalues**. They come from the relation $A = M^{-1}AM = S^{-1}\Lambda S$.

Proof: $Ax = \lambda x$, and $B = M^{-1}AM$. Inserting $MM^{-1} = I$:

$$AMM^{-1} = \lambda x$$

then multiplying both sides by M^{-1} on the left:

$$M^{-1}AMM^{-1}x = \lambda M^{-1}x$$

But $M^{-1}AM = B$! Thus,

$$BM^{-1}x = \lambda M^{-1}x$$

Cancelling the M^{-1}, we at last conclude λ is also an eigenvalue of B,

$$Bx = \lambda x$$

Consider the case that there are repeated eigenvalues. The geometric multiplicity will determine which family matrices with these eigenvalues will belong to.

$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ is in one family.

$\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ is in a different family. It only has one eigenvector.

6.5 The Jordan Form

Matrices that can't be diagonalized can still be put in a form *close* to diagonal. This "close to diagonal" form is called the **Jordan form**. It is notorious for being hard to compute.

Jordan forms consist of **Jordan blocks**. A Jordan block is one

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix}$$

where J_i can be any size, as long as it has a repeated eigenvalue along the diagonal and 1s above it. The complete Jordan form matrix is then

$$\begin{bmatrix} J_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_n \end{bmatrix}$$

The number of blocks corresponds to the number of independent eigenvectors. For a diagonalizable matrix, $J = \Lambda$.

6.6 Left and Right Inverses

6.6.1 Left Inverse

Say a matrix has full column rank: $r = n$. This means our reduced row echelon form is

$$R = \begin{bmatrix} I \\ 0 \end{bmatrix}$$

In the past, we deemed this matrix singular. There is a change b is not in its column space. However, A^TA is invertible! So multiplying A^TA by $\left(A^TA\right)^{-1}$ yields

$$\left(A^TA\right)^{-1}A^TA = I$$

So $(A^TA)^{-1}A^T$ multiplying on the *left* of A leads to I, and we conclude that this expression leads to the **left inverse**.

$$A^{-1}_{\text{left}} = (A^TA)^{-1}A^T$$

As a matter of fact, this is the very matrix we use to project b onto the column space of A back in section 3.2.2! We use it to find $\hat{x}$, then multiply A again for the projection:

$$\hat{x} = (A^TA)^{-1}A^Tb \;\Rightarrow\; A\hat{x} = A(A^TA)^{-1}A^Tb$$

In other words, the left inverse gives us the $\hat{x}$ which is the closest solution to an $Ax = b$, where b is not in the column space.

6.6.2 Right Inverse

The right inverse does effectively the same thing, but for the row space. The matrix, in this case, has full *row* rank: $r = m$. We instead focus on AA^T, as that product is invertible in this case.

$$AA^T\left(AA^T\right)^{-1} = I$$

Multiplying $A^T\left(AA^T\right)^{-1}$ on the right of A leads to I, so

$$A^{-1}_{\text{right}} = A^T\left(AA^T\right)^{-1}$$

Transposing the expression A^{-1}_{right} leads us to

$$\begin{aligned}\left(A_{\text{right}}^{-1}\right)^T &= \left(A^T\left(AA^T\right)^{-1}\right)^T \\ &= (AA^T)^{-1}A\end{aligned}$$

Which, multiplied on the left by A^T, gives us the projection matrix onto the column space of A^T, or the projection matrix onto the row space of A.

6.7 Pseudoinverses

We extend the left and right inverses to a matrix with $r < m$, $r < n$. We claim that, since dimension of row space = dimension of column space, the linear transformation expressed by the matrix A is a one-to-one function between the row space and the column space.

Proof: We claim $Ax \neq Ay$ when $x \neq y$. Then $A(x-y) = 0$. Then $x-y$ is in the null space. But it evidently is also in the column space. Therefore it must be the zero vector. So $x = y$! But we just said $x \neq y$. By contradiction, we prove that Ax must not be Ay for $x \neq y$.

So, even for a singular matrix, every vector that is *outputted* by Ax must come from one unique vector x. It sounds like there should be some sort of inverse! The only reason we can't invert every A is there are some b that are not in the column space of A.

We introduce the pseudoinverse, notated A^+.

The driving idea behind the pseudoinverse is that we'll break such a b up into a component that is not in the column space, and a component that *is* in the column space; $b = b_n + b_c$. The component not in the column space goes to zero, so $A^+(b_n + b_c) = \hat{x} \Rightarrow A\hat{x} = b_c$. We essentially find $\hat{x}$, the vector in the row space that leads to the closest p, the projection of b onto the column space.

Computing the pseudoinverse is actually quite intuitive! We start with the singular value decomposition of A. Inverting it,

$$\begin{aligned}A^+ &= \left(U\Sigma V^T\right)^{-1} \\ &= V\Sigma^+U^T\end{aligned}$$

where Σ^+ is

$$\Sigma^+ = \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}$$

I.e., we take the inverses all singular values, save for singular values of zero, which we keep as zeroes. This makes sense, as we're loosely speaking, going backwards. In other words, we go from $Av_i = \sigma_i u_i$ to $\frac{1}{\sigma_i}v_i = A^+u_i$.

We can also look at the pseudoinverse from a change of basis point of view (see section 8.2):

$$\text{Out: standard basis} \leftarrow V \leftarrow u_1, u_2 \leftarrow \Sigma^+ \leftarrow u_1, u_2 \leftarrow U^T \leftarrow \text{ In: standard basis}$$

where we start with the standard basis, change to the orthonormal basis in the column space $u_1, \ldots u_n$, change from u_i to the basis $v_1, \ldots v_n$ by dividing by our σ_i, then changing back to the standard basis. In this process, by virtue of some $\sigma_{r+1} = 0$, any components of our input bases that are outside the column

space of A go to zero. I.e., there are some $Av_i = 0u_i$, and these v_i are clearly in the null space. If we give the pseudoinverse a u_i that came from $Av_i = 0u_i$, and that are thus not in the column space, those become zero.

Example Take the following SVD:

$$A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Using our above formula for the pseudoinverse, we obtain

$$A^+ = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{10} & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$$

If we multiply A^+b with a b orthogonal to the column space, we expect to have the zero vector.

$$\frac{1}{10} \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

6.7.1 Pseudoinverses and Projections

By taking the pseudoinverse, we're effectively projecting b onto the column space, then finding the $\hat{x}$ that leads to that projection. As such, multiplying by A on the left again should lead us to the projection p.

Therefore the projection of b onto the **column space** of A is

$$AA^+b$$

The advantage of the pseudoinverse is that we can do the same thing for the row space. If we multiply bA^+, we find the $\hat{y}$ in $\hat{y}A$. Therefore, the projection of b onto the **row space** of A is

$$bA^+A$$

6.8 Nilpotent Operators/Matrices

An operator is called **nilpotent** if some power of it is zero:

$$N^j = 0,\ j > 0$$

A nilpotent operator raised to the dimension of the domain is always zero (see section 5.6). $N^{\dim V} = 0$.

The matrix of a nilpotent operator has zeroes along the diagonal, and zeroes below the diagonal:

$$\begin{bmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0sz \end{bmatrix}$$

Chapter 7

Complex Matrices and Vectors

7.1 Conjugate Transposes

We transition out of $\mathbb{R}^n$ and into $\mathbb{C}^n$. A couple of changes have to be made to our operations.

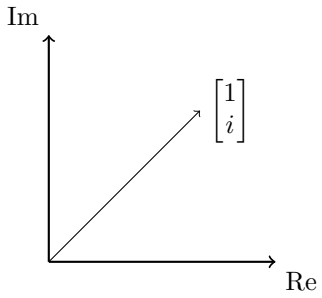

Figure 7.1: The length of this complex vector is obviously not zero.

Firstly, the dot product/inner product is no good. The dot product of the complex vector $(1, i)$ with itself should be its length squared. However,

$$\begin{bmatrix}1 & i\end{bmatrix}\begin{bmatrix}1\\i\end{bmatrix} = 1 + (-1) = 0$$

which is nonsensical. Instead, we define the transpose differently. For complex vectors and matrices, we use the conjugate transpose

$$z^H = \overline{z}^T$$

where $\overline{z}$ is the complex conjugate of z. Recall that the complex conjugate is

$$z = a + bi, \;\; \overline{z} = a - bi$$

where we simply flip the sign of the complex component. This leads to $\overline{z}z = |z|^2 = a^2 + b^2$—which you can work out for yourself—which itself leads nicely to

$$z^H z = \overline{z}^T z = \overline{z}_1 z_1 + \overline{z}_2 z_2 + \cdots \overline{z}_n z_n = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2$$

We similarly redefine orthogonality to mean $x^H y = 0$. A matrix with orthogonal complex columns is called **unitary**.

7.2 Hermitian Matrices

The equivalent of a symmetric matrix for a matrix with complex entries is the **Hermitian**, or self-adjoint matrix.

A **Hermitian** matrix is one $A^H = A$.

For $A^H = A$, the number $z^H Az$ is real for any z. $z^H Az$ is one by one, so its transpose is itself. Take its conjugate transpose:

$$\left(z^H Az\right)^H = z^H A^H (z^H)^H = z^H Az$$

therefore $\overline{z^H Az} = z^H Az$.

Every eigenvalue of a Hermitian is real. Suppose $Az = \lambda z$. Multiply both sides by z^H to get

$$z^H Az = \lambda z^H z$$

$z^H Az$ is real, as has been proved above. On the right, $z^H z$ is the length squared—it is real and positive. Therefore

$$\lambda = \frac{z^H Az}{z^H z}$$

The eigenvectors of a Hermitian are orthogonal. If $Az = \lambda z$ and $Ay = \beta y$, multiply $Az = \lambda z$ on the left by y^H, and multiply the conjugate transpose of $Ay = \beta y$ on the right by z:

$$y^H Az = \lambda y^H z, \;\; y^H A^H z = \beta y^H z$$

By $A^H = A$, we know that $y^H A^H z = y^H Az$. Therefore

$$y^H Az = \lambda y^H z, \;\; y^H Az = \beta y^H z$$

and therefore

$$\lambda y^H z = \beta y^H z$$

Assuming $\lambda \neq \beta$, $y^H z$ must be zero.

7.2.1 Self-Adjoint Operators

The linear transformation associated with a Hermitian matrix is a self-adjoint operator, or a Hermitian operator.

A **self-adjoint** operator is an operator $T \in \mathcal{L}(V)$ such that $T = T^*$. That is,

$$\langle Tv, w\rangle = \langle v, Tw\rangle$$

for all $v, w \in V$.

Here are some results, which correspond to the properties of Hermitian matrices listed above:

- Eigenvalues of self-adjoint operators are *real*.
- Over $\mathbb{C}$, $\langle Tv, v\rangle$ is real for all v only for self-adjoint operators.
- If $T = T^*$ and $\langle Tv, v\rangle = 0$ for all v, then $T = 0$.
- The sum of two self-adjoint operators in self-adjoint, and the product of a real scalar and a self-adjoint operator is self-adjoint.

7.3 Unitary Matrices

Unitary matrices are complex square matrices with orthogonal columns (where x and y are orthogonal if $x^H y = 0$). We typically name them U.

As expected, $U^H U = I$. Therefore, $U^{-1} = U^H$. Suppose $Uz = b$; $|z| = |b|$. We notice this by taking the dot product of this equation with itself

$$z^H U^H U z = z^H z$$

i.e., the length of Uz is the length of z. Therefore, if z is an eigenvector, we notice that all $|\lambda| = 1$.

Unitary matrices have the same properties as orthogonal matrices. Additionally, unitary matrices (like orthogonal matrices), are *normal* (see section 7.4).

7.4 Normal Operators

A normal operator is a $T \in \mathcal{L}(V)$ such that

$$T^*T = TT^*$$

I.e., a normal operator commutes with its adjoint.

T is normal if and only if $|Tv| = |T^*v|$ for all v.

For normal T, T and T^* have the same eigenvectors:

$$Tv = \lambda v \Rightarrow T^*v = \overline{\lambda}v$$

Lastly, normal operators have **orthogonal eigenvectors**. Proof: let α, β be eigenvalues of u, v respectively. u, v are eigenvectors of T, so we know $T^*v = \overline{\beta}v$. Then

$$\begin{aligned}(\alpha - \beta)\langle u, v\rangle &= \langle \alpha u, v\rangle - \langle u, \overline{\beta}v\rangle \\ &= \langle Tu, v\rangle - \langle u, T^*v\rangle \\ &= 0\end{aligned}$$

since $\alpha \neq \beta$, $(\alpha - \beta)\langle u, v\rangle = 0$ means $\langle u, v\rangle$ must equal zero.

7.5 Complex Spectral Theorem

The following are equivalent:

- T is normal; $T^*T = TT^*$.
- V has an orthonormal basis consisting of eigenvectors of T.
- T has a diagonal matrix with respect to some orthogonal basis of V
- As such, there is a decomposition $A = UTU^*$, where U is a unitary matrix (complex equivalent of orthonormal matrix).

This is effectively the same as our spectral theorem as outlined in section 6.1, only it applies to any normal operator.

Chapter 8

Linear Transformations

8.1 Introduction to Linear Transformations

We now turn to the heart of linear algebra, an idea we've been using from the beginning without referring to it by name.

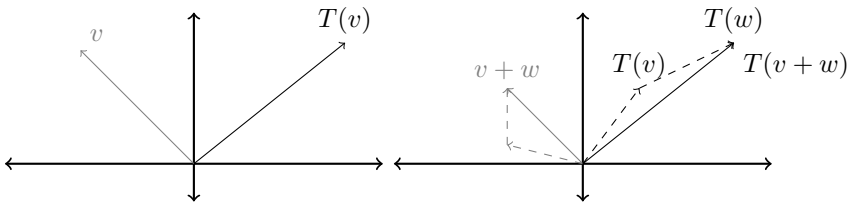

Figure 8.1: The linear transformation $T(v+w)$.

A linear transformation (or linear map) is a function that takes in a vector and outputs another vector. It obeys *linearity*; that is,

$$T(v+w) = T(v) + T(w), \;\; T(cv) = cT(v)$$

In linear, algebra we typically denote vector spaces V and W over a field $\mathbb{F}$ (i.e., either $\mathbb{R}$ or $\mathbb{C}$). A linear transformation is a function

$$T : V \to W$$

I.e., it takes a vector from the vector space V and produces a vector in the vector space W.

An example is $T : \mathbb{R}^2 \to \mathbb{R}^2$. This linear transformation can be, for example, a projection, or a rotation by $45°$. The norm of a vector, however, cannot be a linear transformation. Using $T(v) = |v|$, where $T : \mathbb{R}^2 \to \mathbb{R}$, we quickly notice that $T(-v) \neq -|v|$.

We denote the **set of all linear maps** from V to W

$$\mathcal{L}(V, W)$$

Example The identity map, or transformation, is $I \in \mathcal{L}(V, V)$ such that $Iv = v$.

8.1.1 Operators

A linear transformation T can be a map from a vector space to itself. That is,

$$T : V \to V$$

The set of all operators on V is denoted $\mathcal{L}(V, V)$, or, more commonly, $\mathcal{L}(V)$.

8.2 Change of Basis

When working with a certain basis, the matrix $\begin{bmatrix}1 & 0\end{bmatrix}^T$ *really* means "one of the first vector in the basis." Our basis could be any set of vectors. If we want to change to a different basis, we have to multiply by a matrix.

Say our input basis is $v_1 = (3, 7)$, $v_2 = (2, 5)$. The matrix $\begin{bmatrix}1 & 0\end{bmatrix}^T$ means one of v_1. To convert this to the output standard basis, we must express v_1 in the basis of $(1, 0)$, $(0, 1)$. I.e., we want to get *one* of v_1, and produce $w_1 = (3, 7)$:

$$\text{Standard basis} \leftarrow \begin{bmatrix}3 & 2\\ 7 & 5\end{bmatrix} \leftarrow v_1, v_2$$

So now, we change from $\begin{bmatrix}1 & 0\end{bmatrix}^T$ representing one v_1, to $\begin{bmatrix}1 & 0\end{bmatrix}^T$ representing one $(1, 0)$. You need 3 of $(1, 0)$ and 7 of $(0, 1)$ to represent $(3, 7)$.

If we want to go the other way around we're taking $v_1 = (1, 0)$, $v_2 = (0, 1)$ as an input and the basis $w_1 = (3, 7)$, $w_2 = (2, 5)$ as an output. In other words, if we're passing the output vector $(3, 7)$, as expressed in the standard basis, we must produce $\begin{bmatrix}1 & 0\end{bmatrix}^T$.

$$w_1, w_2 \leftarrow \begin{bmatrix}3 & 2\\ 7 & 5\end{bmatrix}^{-1} \leftarrow \text{ Standard basis.}$$

EXAMPLE Say we are working in function space. We have the following basis:

$$\begin{array}{ll} w_1 = \frac{1}{2}(x^2 + x) & w_1(-1) = 0,\ w_1(0) = 0,\ w_1(1) = 1 \\ w_2 \ = -x^2 + 1 & w_2(-1) = 0,\ w_2(0) = 1,\ w_2(1) = 0 \\ w_3 = \frac{1}{2}(x^2 - x) & w_3(-1) = 1,\ w_3(0) = 0,\ w_3(1) = 0 \end{array}$$

We want to find an $f(x)$ such that $f(1) = 4$, $f(-1) = 6$, and $f(0) = 5$. This is just some transformation that yields

$$4w_1 + 5w_2 + 6w_3$$

Which, in matrix form, corresponds to a

$$f(x) = \begin{bmatrix}4 & & \\ & 5 & \\ & & 6\end{bmatrix} \begin{bmatrix}w_1\\ w_2\\ w_3\end{bmatrix}$$

For which we obtain

$$\begin{aligned} 4w_1 + 5w_2 + 6w_3 &= 4\left(\frac{1}{2}(x^2 + x)\right) + 5(-x^2 + 1) + 6\left(\frac{1}{2}(x^2 - x)\right) \\ &= 2x^2 + 2x - 5x^2 + 5 + 3x^2 - 3x \\ &= -x + 5 \end{aligned}$$

8.3 The Matrix of a Linear Transformation

We'd like to find the coordinates of a linear map $T : V \to W$. Our intuition tells us we should start with a basis for V, so that we may determine what T does to *any* vector in V. We also must choose a basis for W—an output basis.

We choose a basis for V: $v_1, v_2, \dots v_n$. This our input basis. Now we choose an output basis for W: $w_1, w_2, \dots w_m$. $T(v_1)$ will yield

$$T(v_1) = c_1 w_1 + c_2 w_2 + \cdots + c_m 2_m$$

i.e., some combination of the output vectors, as expected. Given that v_1 is the first vector of our input basis (for example, for a standard basis, it is $\langle 1, 0\rangle$), we know that the coefficients c_i are the entries of the *first column* of the matrix of this linear transformation! Assuming that, of course, our matrix takes in the proper basis.

In order to use the standard basis, we must perform the usual change of basis, as described in section 8.2.

Example Suppose a linear transformation T takes $(1, 1)$ to $(2, 2)$ and $(2, 0)$ to $(0, 0)$. Our input basis is clearly $v_1 = (1, 1)$ and $v_2 = (2, 0)$. Thus,

$$T(v_1) = 1w_1, \;\; T(v_2) = 1w_2$$

Now say we want to find $T((4, 2))$. This vector is in $\mathbb{R}^2$, for which v_1, v_2 is a basis. Let us express it as a combination of the basis: $(4, 2) = 2(1, 1) + 1(2, 0) = 2v_1 + v_2$. Using linearity, we know this is

$$T(2v_1 + v_2) = 2T(v_1) + T(v_2) = 2w_1 + w_2$$

which is just $2(2, 2) + (0, 0) = (4, 4)$.

Example Say we want to express the above linear transformation in a matrix that takes in the standard basis $\langle 1, 0\rangle$, $\langle 0, 1\rangle$. Using the given input and output bases,

$$\text{Out: } w_1, w_2 \leftarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \leftarrow \text{In: } v_1, v_2$$

which is not very useful. We must convert from the standard basis to v_1 and v_2. Do this by finding the inverse of the matrix from v_1, v_2 to the standard basis:

$$\text{Out: } v_1, v_2 \leftarrow \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}^{-1} \leftarrow \text{ In: standard basis}$$

and the matrix from the output basis to the standard basis again,

$$\text{Out: standard basis} \leftarrow \begin{bmatrix} 2 & 0 \\ 2 & 0 \end{bmatrix} \leftarrow \text{ In: } w_1, w_2$$

At last we obtain

$$\text{Out: standard basis} \leftarrow \begin{bmatrix} 2 & 0 \\ 2 & 0 \end{bmatrix} \leftarrow w_1, w_2 \leftarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \leftarrow v_1, v_2 \leftarrow \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}^{-1} \leftarrow \text{In: standard basis}$$

$$\begin{bmatrix} 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}^{-1} \Rightarrow \begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix}$$

As expected, multiplying this matrix by $(1, 1)$ yields $(2, 2)$, and multiplying by $(2, 0)$ yields $(0, 0)$.

$$\begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 \\ 2 \end{bmatrix} \qquad \begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

EXAMPLE Find the matrix for the differential operator $D \in \mathcal{P}_2(\mathbb{R})$ with inner product $\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)\,\mathrm{d}x$. Our orthonormal basis is $\sqrt{1/2}, \sqrt{3/2}x, \sqrt{45/8}(x^2 - 1/3)$. Express the linear transformation in terms of the input and output bases:

$$\begin{aligned} D\sqrt{\frac{1}{2}} &= 0 = 0e_1 + 0e_2 + 0e_3 \\ D\sqrt{\frac{3}{2}}x &= \sqrt{\frac{3}{2}} = \sqrt{3}\left(\sqrt{\frac{1}{2}}\right) = \sqrt{3}e_1 \\ D\sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right) &= 2\sqrt{\frac{45}{8}}x = \sqrt{15}\left(\sqrt{\frac{3}{2}}x\right) = \sqrt{15}e_2 \end{aligned}$$

Our matrix $\mathcal{M}(D)$ is then

$$\mathcal{M}(D) = \begin{bmatrix} 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{15} \\ 0 & 0 & 0 \end{bmatrix}$$

8.4 Inner Product Spaces

Inner product spaces are vector spaces "equipped" with an inner product operation. By equipped, we just means that by defining an inner product space, we are defining a vector space with an explicitly defined or obvious inner product.

The inner product associated with the inner product R^2 is, for example, is the Euclidean inner product (dot product) $x^T y$. The inner product associated with $\mathbb{C}^2$ is $x^H y$.

For the vector space $\mathcal{P}_m(v)$, we often use the inner product

$$\langle p, q \rangle = \int_{-a}^{a} p(x)q(x)\,\mathrm{d}x$$

8.4.1 Inner Products

An inner product is a generalization of the dot product. It is notated $\langle u, v \rangle$. It is a function that takes each ordered pair (u, v) of $u, v \in V$ to a number $\langle u, v \rangle \in \mathbb{F}$ and has the following properties:

- **Positivity:** $\langle u, v\rangle \geq 0$ and real.
- **Definiteness:** $\langle v, v\rangle = 0$ only for $v = 0$.
- **Additivity in the first slot:** $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$.
- **Homogeneity in the first slot:** $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$ for $\lambda \in \mathbb{F}$.
- **Conjugate symmetry:** $\langle u, v\rangle = \overline{\langle v, u\rangle}$. For real vector spaces, $\langle u, v\rangle = \langle v, u\rangle$.

Following this definition, we obtain the following properties:

- For each fixed $u \in V$, the function from V to $\langle v, u\rangle$ is a linear map. It is $V \to \mathbb{F}$, so it is, in particular, a linear functional.
- $\langle 0, u\rangle = \langle u, 0\rangle = 0$ for all $u \in V$.
- $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$.
- $\langle u, \lambda v\rangle = \overline{\lambda}\langle u, v\rangle$.

The Euclidean inner product gives us a formula for distance. The distance between two vectors x and y is the norm of their difference:

$$d(x, y) = |x - y| = \sqrt{\langle x - y, x - y\rangle} = \sqrt{(y_1 - x_1)^2 - (y_2 - x_2)^2}$$

We define orthogonality: two vectors $v, u \in V$ are orthogonal if

$$\langle u, v\rangle = 0$$

Zero is orthogonal to every vector in V, and it is the only vector orthogonal to itself.

8.5 Adjoints

The adjoint of a matrix is nothing but the conjugate transpose of a matrix. The notation A^H is interchangeable with the notation A^*.

In the language of linear transformations, we say that the **adjoint** of T is the function $T^* : W \to V$ such that

$$\langle Tv, W\rangle = \langle v, T^*w\rangle$$

for every $v \in V$ and $w \in W$.

The adjoint is a linear map:

$$\begin{aligned}\langle v, T^*(w_1 + w_2)\rangle &= \langle Tv, w_1 + w_2\rangle \\ &= \langle Tv, w_1\rangle + \langle Tv, w_2\rangle \\ &= \langle v, T^*w_1\rangle + \langle v, T^*w_2\rangle \\ &= \langle v, T^*w_1 + T^*w_2\rangle\end{aligned}$$

$$\begin{aligned}\langle v, T^*(\lambda w)\rangle = \langle Tv, \lambda w\rangle &= \overline{\lambda}\langle Tv, w\rangle \\ &= \overline{\lambda}\langle v, T^* w\rangle \\ &= \langle v, \lambda T^* w\rangle\end{aligned}$$

Notice, that, since the matrix of the adjoint is the conjugate transpose of the matrix of the transformation,

$$\begin{aligned}N(T^*) &= (\operatorname{range} T)^\perp \\ \operatorname{range} T^* &= (N(T))^\perp \\ N(T) &= (\operatorname{range} T^*) \\ \operatorname{range} T &= (N(T^*))^\perp\end{aligned}$$

IMPORTANT: The matrix of a linear transformation must be with respect to an **orthonormal basis** for the matrix of its adjoint to be $(\mathcal{M}(T))^H$. If the matrix of a linear transformation is *not* with respect to an orthonormal basis, its conjugate transpose will *not* be the matrix of the adjoint.

8.6 Isometries

An operator $S \in \mathcal{L}(V)$ is called an **isometry** if $|Sv| = |v|$ for all $v \in V$. I.e., an operator is an isometry if it preserves the length of the input vector.

Unitary and orthogonal matrices are isometries.

The following are equivalent:

- S is an isometry.
- $\langle Su, Sv\rangle = \langle u, v\rangle$ for all $u, v \in V$.
- $Se_1 \dots Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \dots e_n$ in V.
- There exists an orthonormal basis $e_1, \dots e_n$ of V such that $Se_1, \dots Se_n$ is orthonormal.
- $S^*S = I$
- $SS^* = I$
- S^* is an isometry.
- S is invertible and $S^{-1} = S^*$.
- There exists an orthonormal basis of V consisting of eigenvectors of S whose corresponding eigenvalues all have $|\lambda| = 1$

Unitary matrices are special cases of isometries. In infinite-dimensional vector spaces, there may be non-unitary isometries.

Chapter 9

Duality

9.1 Linear Functionals

A **linear functional** on a vector space V is a linear map from V to $\mathbb{F}$. I.e., any $T : V \to \mathbb{F}$ is a linear functional. Linear functionals are typically denoted φ.

EXAMPLE Define $\varphi : \mathbb{R}^2 \to \mathbb{R}$ by $\varphi(x, y, z) = 4x - 5y + 2z$. Then φ is a linear functional on $\mathbb{R}^2$.

9.2 Dual Spaces

The **dual space** of V, denoted V', is the vector space of all linear functionals on V. I.e., $V' = \mathcal{L}(V, \mathbb{F})$. $\dim V' = \dim V$, per $\dim \mathcal{L}(V, W) = \dim V \times \dim W$.

If $v_1, \ldots v_n$, the **dual basis** is a list $\varphi_1, \ldots \varphi_n$ of elements of V', where each φ_i is the linear functional on V such that

$$\varphi_i(v_j) = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

The dual basis is then a basis for the dual space V'.

EXAMPLE The dual basis for the standard basis $e_1, \ldots e_n$ of $\mathbb{F}^n$ is list of linear functions $\varphi_1, \ldots \varphi_n$ where φ_i selects the i-th coordinate of a vector.

9.2.1 Dual Map

If $T \in \mathcal{V}, \mathcal{W}$, then the dual map of T is the linear map $T' \in \mathcal{W}', \mathcal{V}'$ defined by $T'(\varphi) = \varphi \circ T$ for $\varphi \in W'$. $\circ$ denotes function composition. T' thus takes in a linear functional on W, and produces a new linear functional which applies $T : V \to W$ and then takes in a vector in W. We are effectively converting a functional that takes in vectors in W into a functional that takes in vectors in W.

Evidently, $T' \in V'$, as $\varphi \circ T \in \mathcal{V}(\mathbb{F})$.

Properties of dual maps are as follows:

- $(S + T)' = S' + T'$ for all $S, T \in \mathcal{L}(V, W)$, and dual maps $S', T' \in \mathcal{L}(W', V')$.
- $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbb{F}, T \in \mathcal{L}(V, W)$.

The **null space** of T', using section 9.2.2, is

$$N(T') = (\text{range}\, T)^0$$

The **range** of T is

$$\text{range}\, T' = ()N(T)^0$$

Furthermore, if T is surjective, T' is injective; if T is injective, T' is surjective.

9.2.2 Annihilators

For $U \subset V$, the **annihilator** of U, denoted U^0, is

$$U^0 = \{\varphi \in V' : \varphi(u) = 0 \text{ for all } u \in U\}$$

In other words, the annihilator of U is the set of all linear functionals of V that produce zero when applied to all vectors in u. Evidently $U^0 \in V'$.

EXAMPLE Let $e_1, \dots e_5$ be the standard basis for $\mathbb{R}^5$, and $varphi_1, \dots \varphi_5$ be the dual basis of $(\mathbb{R}^5)'$. Suppose $U = \text{span}(e_1, e_2)$. φ_i is the linear functional that selects the i-th component of a vector. Then $U^0 = \text{span}(\varphi_3, \varphi_4, \varphi_5)$.

9.2.3 Matrix of a Dual Map

The matrix of T' is the transpose of the matrix of T:

Suppose $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(T) = (\mathcal{M}(T))^T$

9.3 Riesz Representation Theorem

If we fix $u \in V$, then the map that sends v to $\langle v, u\rangle$ is a linear functional.

Specifically, by the **Riesz representation theorem**:

Suppose V is finite dimensional and φ is a linear functional on V. Then there is an unique vector $u \in V$ such that $\varphi(v) = \langle v, u\rangle$ for every $v \in V$.

Proof: There exists a vector $u \in V$ such that $\varphi(v) = \langle v, u\rangle$ for $v \in V$. Let $e_1, \dots e_n$ be an orthonormal basis for V:

$$\begin{aligned}\varphi(v) &= \varphi(\langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n) \\ &= \langle v, e_1\rangle \varphi(e_1) + \cdots + \langle v, e_n\rangle \varphi(e_n) \\ &= \langle v, \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n\rangle\end{aligned}$$

Thus, setting

$$u = \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n \tag{9.1}$$

gives us a u such that $\varphi(v) = \langle v, u\rangle$. We can prove that only one vector $u \in V$ has this behavior. The proof is omitted.

Example Find $u \in \mathcal{P}_2\mathbb{R}$ such that

$$\int_{-1}^{1} p(t)\cos(\pi t)\,\mathrm{d}t = \langle p, u\rangle = \int_{-1}^{1} p(t)u(t\,\mathrm{d}t)$$

Applying (9.1) and using an orthonormal basis $\sqrt{1/2}, \sqrt{3/2}t, \sqrt{45/8}(t^2 - 1/3)$,

$$\begin{aligned} u(t) &= \int_{-1}^{1} \sqrt{\frac{1}{2}}\cos(\pi t)\,\mathrm{d}t \cdot \sqrt{\frac{1}{2}} + \int_{-1}^{1} \sqrt{\frac{3}{2}}t\cos(\pi t)\,\mathrm{d}t \cdot \sqrt{\frac{3}{2}}t \\ &\quad + \int_{-1}^{1} \sqrt{\frac{45}{8}}\left(t^2 - \frac{1}{3}\right)\cos(\pi t)\,\mathrm{d}t \cdot \sqrt{\frac{45}{8}}\left(t^2 - \frac{1}{3}\right) \\ u(t) &= -\frac{45}{2\pi^2}\left(t^2 - \frac{1}{3}\right) \end{aligned}$$