Math 235 Notes

Syed Mustafa Raza Rizvi

August 10, 2020

These are my 2nd year Linear Algebra 2 notes at the University of Waterloo (MATH 235). They are pretty similar to the content you may see in the course notes by D. Wolczuk.
You will find that these aren't very useful as notes, in the sense that they are not significantly shorter than the content in the course notes, they're really just a way for me to type down the content I am learning and absorb it. Also, I won't be including the proofs, it's best to read the course notes for that.
(**Note:** This document does not reflect the complete content of the course as I discontinued typsetting the lecture content in the last two weeks of the course)

Thanks to Professor Dan Wolczuk for providing me with the macros to typeset this LaTeX document.
If the university or staff feel that I should take down this document, please feel free to contact me on github (https://github.com/meowstafa)

7. Fundamental Subspaces

Let A be an $m \times n$ matrix. The **four fundamental subspaces** of A are — Definition Fudamental Subspaces

1. $\text{Col}(A) = \{A\vec{x} \in \mathbb{R}^m | \vec{x} \in \mathbb{R}^n\}$, called the **column space**

2. $\text{Row}(A)\{A^T\vec{x} \in \mathbb{R}^n | \vec{x} \in \mathbb{R}^m\}$, called the **row space**

3. $\text{Null}(A) = \{\vec{x} \in \mathbb{R}^n | A\vec{x} = \vec{0}\}$, called the **null space**

4. $\text{Null}(A^T) = \{\vec{x} \in \mathbb{R}^m | A^T\vec{x} = \vec{0}\}$, called the **left nullspace**

Theorem 7.1.1 If A is an $m \times n$ matrix, then $\text{Col}(A)$ and $\text{Null}(A^T)$ are subspaces of $\mathbb{R}^m$, and $\text{Row}(A)$ and $\text{Null}(A)$ are subspaces of $\mathbb{R}^n$

Theorem 7.1.2 If A is an $m \times n$ matrix, then the colums of A which correspond to leading ones in the RREF of A form a basis for $\text{Col}(A)$. Moreover,

$$\dim \text{Col}(A) = \text{rank } A$$

Lemma 7.1.3 If R is an $m \times n$ matrix and E is an $n \times n$ invertible matrix, then

$$\{RE\vec{x} | \vec{x} \in \mathbb{R}^n\} = \{R\vec{y} | \vec{y} \in \mathbb{R}^n\}$$

Theorem 7.1.4 If A is an $m \times n$ matrix, then the non-zero rows in the reduced row echelon form of A form a basis for $\text{Row}(A)$. Hence,

$$\dim \text{Row}(A) = \text{rank } A$$

Corollary 7.1.5 For any $m \times n$ matrix A we have $\text{rank } A = \text{rank } A^T$

Dimension Theorem If A is an $m \times n$ matrix, then

$$\text{rank } A + \dim \text{Null}(A) = n$$

8. Linear Mappings

Some review

A mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ is called **linear** if

$$L(s\vec{x} + t\vec{y}) = sL(\vec{x}) + tL(\vec{y})$$

for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $s, t \in \mathbb{R}$
Two linear mappings $L : \mathbb{R}^n \to \mathbb{R}^m$ and $M : \mathbb{R}^n \to \mathbb{R}^m$ are said to be **equal** if $L(\vec{v}) = M(\vec{v}), \forall \vec{v} \in \mathbb{R}^n$

Definition
Linear Mapping

The **range** of a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$\text{Range}(L) = \{L(\vec{x}) | \vec{x} \in \mathbb{R}^n\}$$

Definition
Range

The **kernel** of a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$\text{Ker}(L) = \{\vec{x} \in \mathbb{R}^n | L(\vec{x}) = \vec{0}\}$$

Definition
Kernel

The **standard matrix** of a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$[L] = [L(\vec{e}_1) \ \ldots \ L(\vec{e}_n)]$$

It satisfies

$$L(\vec{x}) = [L]\vec{x}$$

for all $\vec{x} \in \mathbb{R}^n$

Definition
Standard Matrix

If $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping, then $\text{Range}(L) = \text{Col}([L])$

Theorem 7.2.1

If $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping, then $\text{Ker}(L) = \text{Null}([L])$

Theorem 7.2.2

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Then,

$$\dim(\text{Range}(L)) + \dim(\text{Ker}(L)) = \dim(\mathbb{R}^n)$$

Theorem 7.2.3

8.1 General Linear Mappings

Linear Mappings $L : \mathbb{V} \to \mathbb{W}$

We will extend our definition of a linear mapping to the case where the domain and codomain are general vector spaces instead of just $\mathbb{R}^n$

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces. A mapping $L : \mathbb{V} \to \mathbb{W}$ is called **linear** if

Definition
Linear Mapping

$$L(s\mathbf{x} + t\mathbf{y}) = sL(\mathbf{x}) + tL(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ and $s, t \in \mathbb{R}$

Two linear mappings $L : \mathbb{V} \to \mathbb{W}$ and $M : \mathbb{V} \to \mathbb{W}$ are said to be **equal** if $L(\mathbf{v}) = M(\mathbf{v}), \forall \mathbf{v} \in \mathbb{V}$

Note: A linear mapping $L : \mathbb{V} \to \mathbb{V}$ is called a **linear operator**

Let $L : \mathbb{V} \to \mathbb{W}$ and $M : \mathbb{V} \to \mathbb{W}$ be linear mappings. We define $L + M : \mathbb{V} \to \mathbb{W}$ by

Definition
Addition
Scalar Multiplication

$$(L + M)(\mathbf{v}) = L(\mathbf{v}) + M(\mathbf{v})$$

and for any $t \in \mathbb{R}$ we define $tL : \mathbb{V} \to \mathbb{W}$ by

$$(tL)(\mathbf{v}) = tL(\mathbf{v})$$

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces. The set $\mathbb{L}$ of all linear mappings $L : \mathbb{V} \to \mathbb{W}$ with standard addition and scalar multiplication of mappings is a vector space

Theorem 8.1.1

Let $L : \mathbb{V} \to \mathbb{W}$ and $M : \mathbb{W} \to \mathbb{U}$ be linear mappings. We define $M \circ L : \mathbb{V} \to \mathbb{U}$ by

Definition
Composition

$$(M \circ L)(\mathbf{v}) = M(L(\mathbf{v})),\ \forall \mathbf{v} \in \mathbb{V}$$

If $L : \mathbb{V} \to \mathbb{W}$ and $M : \mathbb{W} \to \mathbb{U}$ are linear mappings, then $M \circ L : \mathbb{V} \to \mathbb{U}$ is a linear mapping

Theorem 8.1.2

Let $L : \mathbb{V} \to \mathbb{W}$ and $M : \mathbb{V} \to \mathbb{W}$ be linear mappings.
If $(M \circ L)(\mathbf{v}) = \mathbf{v}, \forall \mathbf{v} \in \mathbb{V}$ and $(L \circ M)(\mathbf{w}) = \mathbf{w}, \forall \mathbf{w} \in \mathbb{W}$, then L and M are said to be **invertible**.
We write $M = L^{-1}$ and $L = M^{-1}$

Definition
Invertible Mapping

8.2 Rank-Nullity Theorem

Let's extend the definitions of the range and kernel to general linear mappings.

Definition Range Kernel

For a linear mapping $L : \mathbb{V} \to \mathbb{W}$ the **kernel** of L is defined to be

$$\text{Ker}(L) = \{\mathbf{v} \in \mathbb{V} | L(\mathbf{v}) = \mathbf{0}_{\mathbb{W}}\}$$

and the **range** of L is defined to be

$$\text{Range}(L) = \{L(\mathbf{v}) \in \mathbb{W} | \mathbf{v} \in \mathbb{V}\}$$

Theorem 8.2.1

If $\mathbb{V}$ and $\mathbb{W}$ are vector spaces and $L : \mathbb{V} \to \mathbb{W}$ is a linear mapping, then

$$L(\mathbf{0}_{\mathbb{V}}) = \mathbf{0}_{\mathbb{W}}$$

Theorem 8.2.2

If $L : \mathbb{V} \to \mathbb{W}$ is a linear mapping, then $\text{Ker}(L)$ is a subspace of $\mathbb{V}$ and $\text{Range}(L)$ is a subspace of $\mathbb{W}$

Definition Rank Nullity

Let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. We define the **rank** of L to be

$$\text{rank}(L) = \dim(\text{Range}(L))$$

We define the **nullity** of L to be

$$\text{nullity}(L) = \dim(\text{Ker}(L))$$

Rank-Nullity Theorem

Let $\mathbb{V}$ be an n-dimensional vector space and let $\mathbb{W}$ be a vector space. If $L : \mathbb{V} \to \mathbb{W}$ is linear, then

$$\text{rank}(L) + \text{nullity}(L) = n$$

Remark: The proof for the Rank-Nullity Theorem is identical to the proof for the dimension theorem, only that this time it is a generalisation for general vector spaces. You could see the Rank-Nullity theorem as an analog of the Dimension theorem for general linear mappings.

8.3 Matrix of a Linear Mapping

We will now show that every linear mapping $l : \mathbb{V} \to \mathbb{W}$ can also be represented as a matrix mapping. However, we must be careful when dealing with general vector spaces as our domain and codomain. For example, it is certainly impossible to represent a linear mapping that maps polynomials to matrices, since we cannot multiply a matrix by a polynomial.
Thus, if we are going to convert vectors in $\mathbb{V}$ to vectors in $\mathbb{R}^n$ in order to define matrix representations of general linear mappings. Recall the coordinate vector of $\mathbf{x} \in \mathbb{V}$ with respect to a basis $\mathcal{B}$ is a vector in $\mathbb{R}^n$. In particular, if $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathbf{x} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$ then the coordinate vector of $\mathbf{x}$ with respect to $\mathcal{B}$ is defined to be

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Using coordinates, we can write a matrix mapping representaion of a linear mapping $L : \mathbb{V} \to \mathbb{W}$. We want to find a matrix A such that

$$[L(\mathbf{x})]_{\mathcal{C}} = A[\mathbf{x}]_{\mathcal{B}}$$

for every $\mathbf{x} \in \mathbb{V}$, where $\mathcal{B}$ is a basis for $\mathbb{V}$ and $\mathcal{C}$ is a basis for $\mathbb{W}$
Considering $[L(\mathbf{x})]_{\mathcal{C}}$ and using the properties of linear mappings and coordinates, we get

$$\begin{aligned} [L(\mathbf{x})]_{\mathcal{C}} = [L(b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n)]_{\mathcal{C}} &= b_1[L(\mathbf{v}_1)]_{\mathcal{C}} + \cdots + b_n[L(\mathbf{v}_n)]_{\mathcal{C}} \\ &= \begin{bmatrix} [L(\mathbf{v}_1)]_{\mathcal{C}} & \ldots & [L(\mathbf{v}_n)]_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \end{aligned}$$

This, we have the matrix $\begin{bmatrix} [L(\mathbf{v}_1)]_{\mathcal{C}} & \ldots & [L(\mathbf{v}_n)]_{\mathcal{C}} \end{bmatrix}$ being matrix-vector multiplied by the vector $[\mathbf{x}]_{\mathcal{B}}$ as desired.

Definition Matrix of a Linear Mapping

Suppose $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for a vector space $\mathbb{V}$ and $\mathcal{C}$ is any basis for a finite dimensional vector space $\mathbb{W}$. For a linear mapping $L : \mathbb{V} \to \mathbb{W}$, the **matirx of L with respect to basis $\mathcal{B}$ and $\mathcal{C}$** is defined by

$${}_{\mathcal{C}}[L]_{\mathcal{B}} = \begin{bmatrix} [L(\mathbf{v}_1)]_{\mathcal{C}} & \ldots & [L(\mathbf{v}_n)]_{\mathcal{C}} \end{bmatrix}$$

and satisfies

$$[L(\mathbf{x})]_{\mathcal{C}} = {}_{\mathcal{C}}[L]_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

for all $\mathbf{x} \in \mathbb{V}$

In the special case of a linear operator L acting on a finite dimensional vector space

$\mathbb{V}$ with basis $\mathcal{B}$, we often wish to find the matrix ${}_{\mathcal{B}}[L]_{\mathcal{B}}$

Definition
Matrix of a Linear Operator

Suppose $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is any basis for an n-dimensional vector space $\mathbb{V}$ and let $L : \mathbb{V} \to \mathbb{V}$ be a linear operator. The **$\mathcal{B}$-matrix of L** (or the matrix of L with respect to the basis $\mathcal{B}$) is defined by

$$[L]_{\mathcal{B}} = \begin{bmatrix} [L(\mathbf{v}_1)]_{\mathcal{B}} & \ldots & [L(\mathbf{v}_n)]_{\mathcal{B}} \end{bmatrix}$$

and satisfies

$$[L(\mathbf{x})]_{\mathcal{B}} = [L]_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

for all $\mathbf{x} \in \mathbb{V}$

(Makes more sense if you refer to the course notes) In example 4 we found that the matrix of L with respect to basis $\mathcal{B}$ is diagonal since the vectors in $\mathcal{B}$ are eigenvectors of L. Such a basis is called a **geometrically natural basis**

8.4 Isomorphisms

The ten vector space axioms define a ”structure’ for the set based on the operations of addition and scalar multiplication. Since all vector spaces satisfy the same ten properties, we expect that all n-dimensional vector spaces should have the same structure, and they do as seen by the work done with coordinate vectors. Whatever basis we use for an n-dimensional vector space $\mathbb{V}$, we have a nice way of relating vectors in $\mathbb{V}$ to vectors in $\mathbb{R}^n$. Moreover, we see that the operations of addition and scalar multiplication are preserved with respect to this basis.
No matter which vector space we are using and which basis for the vector space, any linear linear combination of vectors is really just performed on the coordinates of the vectors with respect to the defined basis.

Let's now look at how to use general linear mappings to mathematically prove these observations.

Let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. L is called **one-to-one** (**injective**) if for every $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ such that $L(\mathbf{u}) = L(\mathbf{v})$, we must have $\mathbf{u} = \mathbf{v}$

Definition
One-To-One
Onto

L is called **onto** (**surjective**) if for every $\mathbf{w} \in \mathbb{W}$, there exists $\mathbf{v} \in \mathbb{V}$ such that $L(\mathbf{v}) = \mathbf{w}$

We observe that for a linear mapping $L : \mathbb{V} \to \mathbb{W}$ being onto means that Range$(L) = \mathbb{W}$
And L being one-to-one means that for each $\mathbf{w} \in$ Range(L) there exists exactly one $\mathbf{v} \in \mathbb{V}$ such that $L(\mathbf{v}) = \mathbf{w}$. Another way to say it would be, the images of distinct vectors are distinct. We now establish a relationship between a mapping being one-to-one and its kernel

Lemma 8.4.1

Let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. L is one-to-one (injective) if and only if Ker(L)= $\{\mathbf{0}_{\mathbb{V}}\}$

For two vector spaces $\mathbb{V}$ and $\mathbb{W}$ to have the same structure we need each vector $\mathbf{v} \in \mathbb{V}$ to be identified with a unique vector $\mathbf{w} \in \mathbb{W}$ such that linear combinations are preserved. So to do that we need to find a mapping L from $\mathbb{V}$ to $\mathbb{W}$ which is both one-to-one and onto, and for linear combinations to be preserved we need the linear mapping to be linear.

Definition
Isomorphism
Isomorphic

A vector space $\mathbb{V}$ is said to be **isomorphic** to a vector space $\mathbb{W}$ if there exists a linear mapping $L : \mathbb{V} \to \mathbb{W}$ which is one-to-one and onto. L is called an **isomorphism** from $\mathbb{V}$ to $\mathbb{W}$

Looking at examples of isomorphisms will show that it is mapping basis vectors to basis vectors. And since it is that basis vectors are mapping to basis vectors, the dimensions of the two vector spaces should be the same

Let $\mathbb{V}$ and $\mathbb{W}$ be finite dimensional vector spaces. $\mathbb{V}$ is isomorphic to $\mathbb{W}$ if and only if $\dim \mathbb{V} = \dim \mathbb{W}$ Theorem 8.4.2

If $\mathbb{V}$ and $\mathbb{W}$ are both n-dimensional vector spaces and $L : \mathbb{V} \to \mathbb{W}$ is linear, then L is one-to-one if and only if L is onto Theorem 8.4.3

Let $\mathbb{V}$ and $\mathbb{W}$ be isomorphic vector spaces and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{V}$. A linear mapping $L : \mathbb{V} \to \mathbb{W}$ is an isomorphism if and only if $\{L(\mathbf{v}_1), \ldots, L(\mathbf{v}_n)\}$ is a basis for $\mathbb{W}$ Theorem 8.4.4

Linear Extensions

Suppose you know that the linear transformation $T : P_2(\mathbb{R}) \to P_1(\mathbb{R})$
Where, $T(1) = -3 + 2x, T(x) = 2 + 3x, T(x^2) = x^2$
In order to compute $ax^2 + bx + c$ you would rewrite $T(ax^2 + bx + c) = aT(x^2) + bT(x) + cT(1)$
And so, by assuming linearity, you **determined** the image of $ax^2 + bx + c$ **by the action of T on a basis** $\{1, x, x^2\}$

Let $S, T : \mathbb{V} \to \mathbb{W}$ be linear transformations and let $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_n\}$ be a basis of $\mathbb{V}$ Lemma 8.4.X
If $S(\vec{b}_i) = T(\vec{b}_i)$ for each i, then $S = T$
This can be proven by showing that $S(\mathbf{v}) = T(\mathbf{v})$ for every $\mathbf{v} \in \mathbb{V}$

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces, and let $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_n\}$ be a basis of $\mathbb{V}$ Theorem 8.4.Y
Furthermore, let $f : \mathcal{B} \to \mathbb{W}$ be a function
Then there exists a **unique** linear transformation $T : \mathbb{V} \to \mathbb{W}$ such that

$$T(\vec{b}_i) = f(\vec{b}_i)$$

for each i. Moreover, the general formula for an image of T is

$$T(c_1\vec{b}_1 + \cdots + c_n\vec{b}_n) = c_1 f(\vec{b}_1) + \cdots + c_n f(\vec{b}_n)$$

After proving that a T exists, uniqueness follows immediately from Lemma 8.4.X

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces, and let $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_n\}$ be a basis of $\mathbb{V}$ Definition Linear Extension
Furthermore, let $f : \mathcal{B} \to \mathbb{W}$ be a function.

Then the **linear extension** of f is the linear transformation $T : \mathbb{V} \to \mathbb{W}$ where

$$T(c_1\vec{b}_1 + \cdots + c_n\vec{b}_n) = c_1 f(\vec{b}_1) + \cdots + c_n f(\vec{b}_n)$$

Every linear transformation is the linear extension of its action on a basis

9. Inner Products

9.1 Inner Product Spaces

Definition Inner Product, Inner Product Space

Let $\mathbb{V}$ be a vector space. An **inner product** on $\mathbb{V}$ is a function $\langle , \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ that has the following properties: for every $\mathbf{v}, \mathbf{u}, \mathbf{w} \in \mathbb{V}$ and $s, t \in \mathbb{R}$ we have

I1 $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ iff $\mathbf{v} = \mathbf{0}$ (**Positive Definite**)

I2 $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ (**Symmetric**)

I3 $\langle s\mathbf{v} + t\mathbf{u}, \mathbf{w} \rangle = s\langle \mathbf{v}, \mathbf{w} \rangle + t\langle \mathbf{u}, \mathbf{w} \rangle$ (**Left linear**)

A vector space $\mathbb{V}$ with an inner product $\langle , \rangle$ on $\mathbb{V}$ is called an **inner product space**

Remark: Since an inner product is left linear and symmetric, then it is also **right linear**:

$$\langle \mathbf{w}, s\mathbf{v} + t\mathbf{u} \rangle = s\langle \mathbf{w}, \mathbf{v} \rangle + t\langle \mathbf{w}, \mathbf{u} \rangle$$

Thus, we say that an inner product is **bilinear**

The **dot product** is an inner product on $\mathbb{R}^n$, called the **standard inner product** on $\mathbb{R}^n$

$\langle A, B \rangle = tr(B^T A)$ on $M_{m\times n}(\mathbb{R})$ is called the **standard inner product** on $M_{m\times n}(\mathbb{R})$

Observe that calculating $\langle A, B \rangle$ corresponds exactly to finding the dot product on vectors in $\mathbb{R}^{mn}$ under the obvious isomorphism (one that matches the trace with the dot product expression). Hence, when using this inner product you do not actually have to compute $B^T A$ (and can get away with using the isomorphism and calculating the dot product)

The inner product $\langle f(x), g(x) \rangle = \int_{-\pi}^{\pi} f(x)g(x)dx$ on the vector space $C[-\pi, \pi]$ of continuous functions defined on the closed interval from $-\pi$ to π. It is extremely important in applied math, physics and engineering. This is the foundation for Fourier Series.

Theorem 9.1.1

If $\mathbb{V}$ is an inner product space with inner product $\langle , \rangle$, then for any $v \in \mathbb{V}$ we have

$$\langle \mathbf{v}, \mathbf{0} \rangle = 0$$

Remark: Whenever one considers an inner product space, they must define which inner product space they are using. For $\mathbb{R}^n$ or $M_{m\times n}(\mathbb{R})$ it's generally the standard inner product, this is the one we assume is being used if no other inner product is defined.

9.2 Orthogonality and Length

Length

Let $\mathbb{V}$ be an inner product space with inner product $\langle , \rangle$. For any $\mathbf{v} \in \mathbb{V}$ we define the **length** (or **norm**) of $\mathbf{v}$ by

Definition
Length

$$||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

The length of a vector is well defined as the inner product must be positive definite. The length of any vector in an inner product space depends on the inner product being used.

Let $\mathbb{V}$ be an inner product space with inner product $\langle , \rangle$. For any $\mathbf{x}, \vec{y} \in \mathbb{V}$ and $t \in \mathbb{R}$ we have

Theorem
9.2.1

(1) $||\mathbf{x}|| \geq 0$, and $||\mathbf{x}|| = 0$ iff $\mathbf{x} = \mathbf{0}$

(2) $||t\mathbf{v}|| = |t|||\mathbf{v}||$

(3) $\langle \mathbf{x}, \vec{y} \rangle \leq ||\mathbf{x}||||\vec{y}||$ Cauchy-Schwarz-Bunyakovski Inequality

(4) $||\mathbf{x} + \vec{y}|| \leq ||\mathbf{x}|| + ||\vec{y}||$ Triangle Inequality

Let $\mathbb{V}$ be an inner product space with inner product $\langle , \rangle$. If $\mathbf{v} \in \mathbb{V}$ is a vector such that $||\mathbf{v}|| = 1$, then $\mathbf{v}$ is called a **unit vector**

Definition
Unit vector

We often find a unit vector, $\hat{\mathbf{v}}$, in the direction of a certain vector $\mathbf{v} \in \mathbb{V}$, this is called **normalizing** the vector. By theorem 9.2.1 we see

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{||\mathbf{v}||}$$

Orthogonality

Let $\mathbb{V}$ be an inner product space with inner product $\langle , \rangle$. If $\mathbf{x}, \vec{y} \in \mathbb{V}$ such that

Definition
Orthogonal
vectors

$$\langle \mathbf{x}, \vec{y} \rangle = 0$$

then $\mathbf{x}$ and $\vec{y}$ are said to be **orthogonal**

Just like length, whether two vectors are orthogonal or not is dependent on its definition of the inner product.

If $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a set of vectors in an inner product space $\mathbb{V}$ with inner product $\langle , \rangle$ such that $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for all $i \neq j$, then $\mathcal{S}$ is called an **orthogonal set**

Definition
Orthogonal
Set

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal set in an inner product space $\mathbb{V}$, then

$$||\mathbf{v}_1 + \cdots + \mathbf{v}_k||^2 = ||\mathbf{v}_1||^2 + \cdots + ||\mathbf{v}_k||^2$$

Theorem 9.2.2

If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal set in an inner product space $\mathbb{V}$ with inner product $\langle , \rangle$ such that $\mathbf{v}_i \neq 0$ for all $1 \leq i \leq k$, then S is linearly independent

Theorem 9.2.3

If $\mathcal{B}$ is an orthogonal set in an inner product space $\mathbb{V}$ that is a basis for $\mathbb{V}$, then $\mathcal{B}$ is called an **orthogonal basis** for $\mathbb{V}$

Definition Orthogonal Basis

If $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for an inner product space $\mathbb{V}$ with linear product $\langle , \rangle$ and $\mathbf{v} \in \mathbb{V}$, then the coefficient of $\mathbf{v}_i$ when $\mathbf{v}$ is written as a linear combination of the vectors in $\mathcal{S}$ is $\frac{\langle \mathbf{v}, \mathbf{v}_i \rangle}{||\mathbf{v}_i||^2}$. In particular,

$$\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{v}_1 \rangle}{||\mathbf{v}_1||^2}\mathbf{v}_1 + \cdots + \frac{\langle \mathbf{v}, \mathbf{v}_n \rangle}{||\mathbf{v}_n||^2}\mathbf{v}_n$$

Theorem 9.2.4

Orthonormal Bases

Since the formula for the coordinates would be simpler if the vectors in the orthogonal basis were unit vectors. So we will **normalize** vectors so that they are of unit length. $\hat{\mathbf{v}} = \frac{\mathbf{v}}{||\mathbf{v}||}$

If $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal set in the inner product space $\mathbb{V}$ such that $||\mathbf{v}_i|| = 1$ for $1 \leq i \leq k$, then $\mathcal{S}$ is called an **orthonormal set**

Definition Orthonormal Set

A basis for an inner product space $\mathbb{V}$ which is an orthonormal set is called an **orthonormal basis** of $\mathbb{V}$

Definition Orthonormal Basis

If $\mathbf{v}$ is any vector in an inner product space $\mathbb{V}$ with inner product $\langle , \rangle$ and $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for $\mathbb{V}$, then

$$\mathbf{v} = \langle \mathbf{v}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \cdots + \langle \mathbf{v}, \mathbf{v}_n \rangle \mathbf{v}_n$$

Corollary 9.2.5

Orthogonal Matrices

For an $n \times n$ matrix P, the following are equivalent

Theorem 9.2.6

1. The columns of P form an orthonormal basis for $\mathbb{R}^n$
2. $P^T = P^{-1}$
3. The rows of P form an orthonormal basis for $\mathbb{R}^n$

If the columns of an $n \times n$ matrix P form an *orthonormal* basis for $\mathbb{R}^n$, then P is called an **orthogonal matrix**

Definition Orthogonal Matrix

If P and Q are $n \times n$ orthogonal matrices and $\mathbf{x}, \vec{y} \in \mathbb{R}^n$, then

Theorem 9.2.7

1. $(P\mathbf{x}) \cdot (P\vec{y}) = \mathbf{x} \cdot \vec{y}$
2. $||P\mathbf{x}|| = ||\mathbf{x}||$
3. $\det P = \pm 1$
4. All real eigenvalues of P are 1 or -1
5. PQ is an orthogonal matrix

9.3 The Gram-Schmidt Procedure

(**Gram-Schmidt Orthogonalization Theorem**)
Let $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be a basis for an inner product space $\mathbb{W}$. If we define $\mathbf{v}_1, \ldots, \mathbf{v}_n$ successively as follows:

Theorem 9.3.1

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{w}_1 \\
\mathbf{v}_2 &= \mathbf{w}_2 - \frac{\langle \mathbf{w}_2, \mathbf{v}_1 \rangle}{||\mathbf{v}_1||^2}\mathbf{v}_1 \\
\mathbf{v}_i &= \mathbf{w}_i - \frac{\langle \mathbf{w}_i, \mathbf{v}_1 \rangle}{||\mathbf{v}_1||^2}\mathbf{v}_1 - \frac{\langle \mathbf{w}_i, \mathbf{v}_2 \rangle}{||\mathbf{v}_2||^2}\mathbf{v}_2 - \cdots - \frac{\langle \mathbf{w}_i, \mathbf{v}_{i-1} \rangle}{||\mathbf{v}_{i-1}||^2}\mathbf{v}_{i-1}
\end{aligned}$$

for $3 \leq k \leq n$, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for $\text{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ for $1 \leq k \leq n$

Note: the theorem implies that *Every finite dimensional inner product space has an orthogonal basis.*
The process for finding an orthogonal basis for an inner product space in the theorem is called the **Gram-Schmidt procedure**
The order in which one processes the vectors while applying the procedure changes what the resultant basis looks like.

(QR-Decomposition) Theorem 9.3.2

Let $A = [\vec{a}_1 \ \dots \ \vec{a}_n] \in M_{m \times n}(\mathbb{R})$ where $\dim \operatorname{Col}(A) = n$. Let $\vec{q}_1, \dots, \vec{q}_n$ denote the vectors that result from applying the Gram-Schmidt procedure to the columns of A (in order) and then normalizing. If we define

$$Q = [\vec{q}_1 \ \cdots \ \vec{q}_n]$$

and

$$R = \begin{bmatrix} \vec{a}_1 \cdot \vec{q}_1 & \vec{a}_2 \cdot \vec{q}_1 & \dots & \vec{a}_n \cdot \vec{q}_1 \\ 0 & \vec{a}_2 \cdot \vec{q}_2 & \dots & \vec{a}_n \cdot \vec{q}_2 \\ 0 & 0 & \ddots & \vec{a}_n \cdot \vec{q}_{n-1} \\ 0 & \dots & 0 & \vec{a}_n \cdot \vec{q}_n \end{bmatrix}$$

then Q is orthogonal, R is invertible, and

$$A = QR$$

9.4 General Projections

As the title suggests, we will abstract the ideas of projections from our works that used the standard inner product in $\mathbb{R}^n$. Given a vector $\mathbf{x} \in \mathbb{R}^3$ and a plane $P \in \mathbb{R}^3$ which passes through the origin (a subspace), we wanted to write $\mathbf{x}$ as the sum of a vector in P and a vector orthogonal to every vector in P. Therefore, for a given subspace $\mathbb{W}$ of an inner product space $\mathbb{V}$ and any vector $\mathbf{v} \in \mathbb{V}$

$$\mathbf{v} = \text{proj}_{\mathbb{W}}(\mathbf{v}) + \text{perp}_{\mathbb{W}}(\mathbf{v})$$

Definition Orthogonal Complement

Let $\mathbb{W}$ be a subspace of an inner product space $\mathbb{V}$. The **orthogonal complement** of $\mathbb{W}$, $\mathbb{W}^{\perp}$, in $\mathbb{V}$ is defined by

$$\mathbb{W}^{\perp} = \{\mathbf{v} \in \mathbb{V} |\, \langle \mathbf{w}, \mathbf{v} \rangle = 0 \text{ for all } \mathbf{w} \in \mathbb{W}\}$$

Theorem 9.4.1

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a spanning set for a subspace $\mathbb{W}$ of an inner product space $\mathbb{V}$, and let $\mathbf{x} \in \mathbb{V}$. We have that $\mathbf{x} \in \mathbb{W}^{\perp}$ if and only if $\langle \mathbf{x}, \mathbf{v}_i \rangle = 0$ for $1 \leq i \leq k$

Theorem 9.4.2

If $\mathbb{W}$ is a subspace of an inner product space $\mathbb{V}$, then

1. $\mathbb{W}^{\perp}$ is a subspace of $\mathbb{V}$
2. If $\dim \mathbb{V} = n$, then $\dim \mathbb{W}^{\perp} = n - \dim \mathbb{W}$
3. If $\dim \mathbb{V} = n$, then $(\mathbb{W}^{\perp})^{\perp} = \mathbb{W}$
4. $\mathbb{W} \cap \mathbb{W}^{\perp} = \{\mathbf{0}\}$
5. If $\dim \mathbb{V} = n$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for $\mathbb{W}$, and $\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for $\mathbb{W}^{\perp}$, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for $\mathbb{V}$

Definition Projection, Perpendicular

Suppose $\mathbb{W}$ is a k-dimensional subspace of an inner product space $\mathbb{V}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for $\mathbb{W}$. For any $\mathbf{v} \in \mathbb{V}$ we define the **projection** of $\mathbf{v}$ onto $\mathbb{W}$ by

$$\text{proj}_{\mathbb{W}}(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{v}_1 \rangle}{||\mathbf{v}_1||^2}\mathbf{v}_1 + \cdots + \frac{\langle \mathbf{v}, \mathbf{v}_k \rangle}{||\mathbf{v}_k||^2}\mathbf{v}_k$$

and the **perpendicular** of the projection by

$$\text{perp}_{\mathbb{W}}(\mathbf{v}) = \mathbf{v} - \text{proj}_{\mathbb{W}}(\mathbf{v})$$

Notice that we need an *orthogonal basis* for $\mathbb{W}$ to calculate the projection. Hence, these are often called **orthogonal projections**.

To save ourselves some work, we defined the perpendicular of the projection in such a way that we do not need an orthogonal basis (or any basis) for $\mathbb{W}^\perp$.

Theorem 9.4.3 If $\mathbb{W}$ is a k-dimensional subspace of an inner product space $\mathbb{V}$, then for any $\mathbf{v} \in \mathbb{V}$ we have

$$\text{perp}_{\mathbb{W}}(\mathbf{v}) = \mathbf{v} - \text{proj}_{\mathbb{W}}(\mathbf{v}) \in \mathbb{W}^\perp$$

Theorem 9.4.4 If $\mathbb{W}$ is a k-dimensional subspace of an inner product space $\mathbb{V}$, then $\text{proj}_{\mathbb{W}}$ is a linear operator on $\mathbb{V}$ with kernel $\mathbb{W}^\perp$

Theorem 9.4.5 If $\mathbb{W}$ is a subspace of a finite dimensional inner product space $\mathbb{V}$, then for any $\mathbf{v} \in \mathbb{V}$ we have

$$\text{proj}_{\mathbb{W}^\perp}(\mathbf{v}) = \text{perp}_{\mathbb{W}}(\mathbf{v})$$

9.5 The Fundamental Theorem

Definition (Direct Sum) Let $\mathbb{V}$ be a vector space and $\mathbb{U}$ and $\mathbb{W}$ be subspaces of a vector space $\mathbb{V}$ such that $\mathbb{U} \cap \mathbb{W} = \{\mathbf{0}\}$. The **direct sum** of $\mathbb{U}$ and $\mathbb{W}$ is

$$\mathbb{U} \oplus \mathbb{W} = \{\mathbf{u} + \mathbf{w} \in \mathbb{V} | \mathbf{u} \in \mathbb{U}, \mathbf{w} \in \mathbb{W}\}$$

Theorem 9.5.1 If $\mathbb{U}$ and $\mathbb{W}$ are subspaces of a vector space $\mathbb{V}$ such that $\mathbb{U} \cap \mathbb{W} = \{\mathbf{0}\}$, then $\mathbb{U} \oplus \mathbb{W}$ is a subspace of $\mathbb{V}$. Moreover, if $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a basis for $\mathbb{U}$ and $\{\mathbf{w}_1, \dots, \mathbf{w}_l\}$ is a basis for $\mathbb{W}$, then $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_1, \dots, \mathbf{w}_l\}$ is a basis for $\mathbb{U} \oplus \mathbb{W}$

Corollary 9.5.2 If $\mathbb{U}$ and $\mathbb{W}$ are subspaces of a vector space $\mathbb{V}$ such that $\mathbb{U} \cap \mathbb{W} = \{\mathbf{0}\}$ and $\mathbf{v} \in \mathbb{U} \oplus \mathbb{W}$, then there exists unique $\mathbf{u} \in \mathbb{U}$ and $\mathbf{w} \in \mathbb{W}$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$

Theorem 9.5.3 If $\mathbb{V}$ is a finite dimensional inner product space and $\mathbb{W}$ is a subspace of $\mathbb{V}$, then

$$\mathbb{W} \oplus \mathbb{W}^\perp = \mathbb{V}$$

Theorem 9.5.4 **(The Fundamental Theorem of Linear Algebra)**

If A is an $m \times n$, then $\text{Col}(A)^\perp = \text{Null}(A^T)$ and $\text{Row}(A^T) = \text{Null}(A)$. In particular,

$$\mathbb{R}^n = \text{Row}(A) \oplus \text{Null}(A) \quad \text{and} \quad \mathbb{R}^m = \text{Col}(A) \oplus \text{Null}(A^T)$$

The Fundamental Theorem of Linear Algebra implies the Rank-Nullity Theorem. For example, if A is an $m \times n$ matrix, then we know that the rank and nullity of the linear mapping $L(\mathbf{x}) = A\mathbf{x}$ is rank L = rank A = dim Row A and nullity

$L = \dim$ (Null A). Since $\mathbb{R}^n = \text{Row}(A) \oplus \text{Null}(A)$ we get by Theorem 9.4.2 (2) that

$$n = \text{rank } L + \text{nullity } (L)$$

It can also be shown that the Fundamental Theorem of Linear Algebra implies a lot of our results about solving systems of linear equations.

9.6 The Method of Least Squares

In the sciences one often tries to find a correlation between two quantities by collecting data from repeated experimentation. Say, for example, a scientist is comparing quantities x and y which are known to satisfy a quadratic relation $y = a_0 + a_1 x + a_2 x^2$. The scientist would like to find appropriate values for a_0, a_1, and a_2. So the scientist performs several experiments involving these two quantities and gets y_i and x_i. And gets the equations in the three unknowns a_0, a_1, a_2

$$y_1 = a_0 + a_1 x_1 + a_2 x_1^2 \tag{1}$$

$$\vdots = \qquad \vdots \tag{2}$$

$$y_m = a_0 + a_1 x_m + a_2 x_m^2 \tag{3}$$

which has more euqations then unknowns. Such a system of equations is called an **overdetermined system**. Also, due to experimentation error, the system is very likely to be inconsistent. So, we need to find values of a_0, a_1, a_2 which best approximate the data collected.

To solve the problem, we rephrase it in terms of Linear Algebra. Let A be an $m \times n$ matrix with $m > n$ and let the system $A\vec{x} = \vec{b}$ be inconsistent. We want to find a vector $\vec{x}$ that minimizes the distance between $A\vec{x} = \vec{b}$, i.e. we need to minimize $||\vec{b} - A\vec{x}\,||$

We will return to the original problem soon

(Approximation Theorem) Theorem 9.6.1

Let $\mathbb{W}$ be a finite dimensional subspace fo an inner product space $\mathbb{V}$. If $\mathbf{v} \in \mathbb{V}$, then the vector closest to $\mathbf{v}$ in $\mathbb{W}$ is $\text{proj}_{\mathbb{W}}(\mathbf{v})$. That is,

$$||\mathbf{v} - \text{proj}_{\mathbb{W}}(\mathbf{v})|| < ||\mathbf{v} - \mathbf{w}||$$

for all $\mathbf{w} \in \mathbb{W}, \mathbf{w} \neq \text{proj}_{\mathbb{W}}(\mathbf{v})$

Notice that $A\vec{x}$ is in the column space of A. Thus, the Approximation Theorem tells us that we can minimize $||\vec{b} - A\vec{x}\,||$ by finding the projection of $\vec{b}$ onto the column space of A. Therefore, if we solve the consistent system

$$A\vec{x} = \text{proj}_{\text{Col}A}(\vec{b})$$

we will find the desired vector $\vec{x}$. A simpler form would be

$$\vec{b} - A\vec{x} = \vec{b} - \text{proj}_{\text{Col}A}(\vec{b}) = \text{perp}_{\text{Col}A}(\vec{b})$$

Thus, $\vec{b} - A\vec{x}$ is in the orthogonal complement of the column space of A which, by the Fundamental Theorem of Linear Algebra, means $\vec{b} - A\vec{x}$ is in the nullspace of A^T. Hence

$$A^T(\vec{b} - A\vec{x}) = \vec{0}$$

or equivalently

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

This is called the **normal system** and the individual equations are called the **normal equations**. This system will be consistent by construction. However, it need not have a unique solution. If it does have infinitely many solutions, then each of the solutions will minimize $||\vec{b} - A\vec{x}||$

This method of finding an approximate solution is called the **method of least squares**, because we are minimizing

$$||\vec{b} - A\mathbf{x}|| = \left\| \begin{bmatrix} \vec{v}_1 \\ \vdots \\ \vec{v}_m \end{bmatrix} \right\| = \sqrt{v_1^2 + \cdots + v_m^2}$$

which is the same as minimizing $v_1^2 + \cdots + v_m^2$

Returning to the problem of finding a curve of best fit for a set of data points. Let's say we have data points $(x_1, y_1), \ldots, (x_m, y_m)$ and we want to find the values of $a_0, \ldots, a_n$ such that $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ is the polynomial of best fit. That is, we want to find the values of $a_0, \ldots, a_n$ such that the values of y_i are approximated as well as possible by $p(x_i)$.

Let $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ and define $p(\vec{x}) = \begin{bmatrix} p(x_1) \\ \vdots \\ p(x_m) \end{bmatrix}$. To make this look like the method of least squares we write

$$p(\vec{x}) = \begin{bmatrix} p(x_1) \\ \vdots \\ p(x_m) \end{bmatrix} = \begin{bmatrix} a_0 + a_1(x_1) + \cdots + a_n(x_1)^n \\ \vdots \\ a_0 + a_1(x_m) + \cdots + a_n(x_m)^n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & \ldots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \ldots & x_m^n \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = X\vec{a}$$

Thus, we are trying to minimize $||X\vec{a} - \vec{y}||$ and we can use the method of least squares above.

Let m data points $(x_1, y_1), \ldots, (x_m, y_m)$ be given and write Theorem 9.6.2

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_1 & \ldots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \ldots & x_m^n \end{bmatrix}$$

If $\vec{a} = \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix}$ is any solution to the normal system

$$X^T X\vec{a} = X^T\vec{y}$$

then the polynomial

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

is the best fitting polynomail of degree n for the given data. Moreover, if at least $n+1$ of the numbers $x_1, \ldots, x_m$ are distinct, then the matrix $X^T X$ is invertible and this $\vec{a}$ is unique with

$$\vec{a} = (X^T X)^{-1} X^T\vec{y}$$

10. Applications of Orthogonal Matrices

10.1 Orthogonal Similarity

In chapter 6, we said that two matrices A and B are similar if there exists an invertible matrix P such that $P^{-1}AP = B$. If $L : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator and $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ is a basis for $\mathbb{R}^n$, then $P = [\vec{v}_1 \ \ \ldots \ \ \vec{v}_n]$ we get

$$[L]_{\mathcal{B}} = P^{-1}[L]P$$

We then look for a basis $\mathcal{B}$ of $\mathbb{R}^n$ such that $[L]_{\mathcal{B}}$ was diagonal. We found such a basis is made up of the eigenvectors of $[L]$

We found in the last chapter that orthogonal matrices have certain properties. In this chapter we try to find an orthonormal basis $\mathcal{B}$ of eigenvectors such that $[L]_{\mathcal{B}}$ is diagonal. Since the corresponding matrix P will be orthogonal

$$[L]_{\mathcal{B}} = P^T[L]P$$

Two matrices A and B are said to be **orthogonally similar** if there exists an orthogonal matrix P such that

$$P^TAP = B$$

Definition Orthogonally Similar

Since the matrix P would be orthogonal, we'd have that $P^T = P^{-1}$, so we also have $B = P^{-1}AP$ and so we still have the properties of similar matricies, $\text{rank}(A) = \text{rank}(B)$, $\text{tr } A = \text{tr } B$, $\det A = \det B$

(Triangularization Theorem)

Theorem 10.1.1

If A is an $n \times n$ matrix with real eigenvalues, then A is orthogonally similar to an upper triangular matrix T

If A is orthogonally similar to an upper triangular matrix T, then A and T must share the same eigenvalues. Thus, since T is upper triangular, the eigenvalues must appear along the main diagonal of T

The proof gives us a method for finding an orthogonal matrix P so that $P^TAP = T$ is upper triangular.

10.2 Orthogonal Diagonalization

We now know that for any linear operator L on $\mathbb{R}^n$ with real eigenvalues we can find an orthonormal basis $\mathcal{B}$ such that the $\mathcal{B}$-matrix of L is upper triangular. But having a diagonal matrix would be even better, so we'll now look at which matrices are orthogonally similar to a diagonal matrix.

An $n \times n$ matrix A is said to be **orthogonally diagonalizable** if there exists an orthogonal matrix P and diagonal matrix D such that

Definition Orthogonally Diagonalizable

$$P^T AP = D$$

that is, if A is orthogonally similar to a diagonal matrix So we need to first determine which matrices are orthogonally diagonalizable.
We'll work backwords, so lets assume that A is orthogonally diagonalizable. Then by definition, there exists an orthogonal matrix P such that $P^T AP = D$ is diagonal (since that's the end result).We want to see what the properties of A are which make it diagonalizable. A and D are similar (also, orthogonally similar), so they must have the same properties. So, we can instead look at D to tell us the properties of A. Since D is diagonal, $D^T = D$.
Using the fact that P is orthogonal, we can rewrite $P^T AP = D$ as $A = PDP^T$. Taking the transpose gives

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$$

so we have that $A = A^T$

If A is orthogonally diagonalizable, then $A^T = A$

Theorem 10.2.1

This theorem develops into an **if and only if** theorem. (See the upcoming remark)

A matrix A such that $A^T = A$ is said to be **symmetric**

Definition Symmetric Matrix

So, to be symmetric the matrix should be square. Also, it is symmetric iff $a_{ij} = a_{ji}$ for all $1 \leq i, j \leq n$

If A is a symmetric matrix with real entries, then all of its eigenvalues are real

Lemma 10.2.2

(Principle Axis Theorem)
Every symmetric matrix A is orthogonally diagonalizable

Theorem 10.2.3

Remark: We have now proven that a matrix is orthogonally diagonalizable if and only if it is symmetric

A matrix A is symmetric if and only if $\vec{x} \cdot (A\vec{y}) = (A\vec{x}) \cdot \vec{y}$ for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ Theorem 10.2.4

If $\vec{v}_1, \vec{v}_2$ are eigenvectors of a symmetric matrix A corresponding to distinct eigenvalues λ_1, λ_2, then $\vec{v}_1$ is orthogonal to $\vec{v}_2$ Theorem 10.2.5

Consequently, if a symmetric matrix A has n distinct eigenvalues, the basis of eigenvectors which daigonalizes A will naturally be orthogonal. Hence, to orthogonally diagonalize such a matrix A, we just need to normalize these eigenvectors to form an orthonormal basis for $\mathbb{R}^n$ of eigenvectors of A

10.3 Quadratic Forms

Let A be an $n \times n$ matrix. A function $Q : \mathbb{R}^n \to \mathbb{R}$ of the form Definition Quadratic Form

$$Q(\vec{x}) = \vec{x}^T A \vec{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

is called a **quadratic form**

We see that quadratic forms $Q(\vec{x})$ have no cross terms if and only if its corresponding symmetric matrix is diaogonal. We call such a quadratic from a **diagonal quadratic form**.

Classifying Quadratic Forms

Let $Q(\vec{x})$ be a quadratic form. We say that: Definition Classifications of Quadratic Forms

1. $Q(\vec{x})$ is **positive definite** if $Q(\vec{x}) > 0$ for all $\vec{x} \neq \vec{0}$,
2. $Q(\vec{x})$ is **negative definite** if $Q(\vec{x}) < 0$ for all $\vec{x} \neq \vec{0}$,
3. $Q(\vec{x})$ is **indefinite** if $Q(\vec{x}) > 0$ for some $\vec{x}$ and $Q(\vec{x}) < 0$ for some $\vec{x}$,
4. $Q(\vec{x})$ is **positive semidefinite** if $Q(\vec{x}) \geq 0$ for all $\vec{x}$,
5. $Q(\vec{x})$ is **negative semidefinite** if $Q(\vec{x}) \leq 0$ for all $\vec{x}$.

Let A be the symmetric matrix corresponding to a quadratic form $Q(\vec{x}) = \vec{x}^T A \vec{x}$. If P is an orthogonal matrix that diagonalizes A, then $Q(\vec{x})$ can be expressed as Theorem 10.3.1

$$\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

where $\vec{y} = P^T \vec{x}$ and where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A correspondong to the columns of P

Remark: The eigenvectors we used to make up P are called the **principal axes** of A. We will see a geometric interpretation of this in the section for sketching quadratic froms.

If A is a symmetric matrix, then the quadratic form $Q(\vec{x}) = \vec{x}^T A\vec{x}$ is

Theorem 10.3.2

1. positive definite iff the eigenvalues of A are all positive
2. negative definite iff the eigenvalues of A are all negative
3. indefinite iff some of the eigenvalues of A are positive and some are negative
4. positive semidefinite iff all of the eigenvalues are non-negative
5. negative semidefinite iff all of the eigenvalues are non-positive

10.4 Sketching Quadratic Forms

Refer to course notes

10.5 Optimizing Quadratic Forms

It is often useful in applications to find the maximum and/or minimum of a quadratic form subject to a constraint. It is usually possible to perform change of variables so that the constraint is of the form $||\vec{x}|| = 1$. Here we exploit the fact that

$$1 = ||\vec{x}||^2 = x_1^2 + \cdots + x_n^2$$

(Refer to examples in the course notes)

Let $Q(\vec{x})$ be a quadratic form on $\mathbb{R}^n$ with corresponding symmetric matrix A. The maximum value and minimum value of $Q(\vec{x})$ subject to the constraint $||\vec{x}|| = 1$ are the greatest and least eigenvalues of A respectively. Moreover, the values occur when $\vec{x}$ is taken to be a unit eigenvector corresponding to the eigenvalue

Theorem 10.5.1

10.6 Singular Value Decomposition

We are not always lucky to have a symmetric matrix, or even a square matrix. Therefore, we develop methods for finding the maximum and minimum of $||A\vec{x}||$ for an $m \times n$ matrix A subject to the constraint $||\vec{x}|| = 1$.
This derivation will provide us with a very important matrix factorization known as

the singular value decomposition.

If we have an $m \times n$ matrix A, and if we wish to maximize $||A\vec{x}||$ where $||\vec{x}|| = 1$, then if we consider

$$||A\vec{x}||^2 = (A\vec{x})^T(A\vec{x}) = \vec{x}^T A^T A\vec{x}$$

Since A^TA is symmetric, this becomes a quadratic form, and using Theorem 10.5.1, to maximize $||A\vec{x}||$ we just need to find the square root of the largest eigenvalue of A^TA

Since we take the square root of the eigenvalues, it is worth knowing that the eigenvalues of A^TA are always non-negative

If A is an $m\times n$ matrix and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A^TA with corresponding unit eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$, then $\lambda_1, \ldots, \lambda_n$ are all non-negative. In particular, — Theorem 10.6.1

$$||A\vec{v}_i|| = \sqrt{\lambda_i}\,, \quad 1 \le i \le n$$

The **singular values** $\sigma_1, \ldots, \sigma_n$ of an $m \times n$ matrix A are the square roots of the eigenvalues of A^TA arranged so that $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ — Definition Singular Values

It is easy to show that the number of non-zero eigenvalues of A^TA is the rank of A^TA. So we wish to know if there is any relationship between the number of non-zero singular values of an $m \times n$ matrix A and the rank of A

If A is an $m \times n$ matrix, then $\text{Null}(A^TA) = \text{Null}(A)$ — Lemma 10.6.2

If A is an $m \times n$ matrix, then $\text{rank}(A^TA) = \text{rank}(A)$ — Lemma 10.6.3

If A is an $m \times n$ matrix and $\text{rank}(A) = r$, then A has r non-zero singular values — Corollary 10.6.4

For $m \neq n$, we cannot have $A\vec{v} = \sigma\vec{v}$ since the dimensions do not match. So to mimic the eigenvector behavior, we define singular vectors. We pick suitable non-zero vectors $\vec{v} \in \mathbb{R}^n$ and $\vec{u} \in \mathbb{R}^m$ such that $A\vec{v} = \sigma\vec{u}$

By definition, for any non-zero singular value of σ of A there is a vector $\vec{v} \neq \vec{0}$ such that $A^TA\vec{v} = \sigma^2\vec{v}$. Thus if we have $A\vec{v} = \sigma\vec{u}$ then we get $A^TA\vec{v} = A^T(\sigma\vec{u})$ so $\sigma^2\vec{v} = \sigma A^T\vec{u}$. Dividing the last equation by σ, we see that we must also have $A^T\vec{u} = \sigma\vec{v}$. Additionally, if $\vec{v}$ is a unit eigenvector of A^TA, then we get that $\vec{u} = \frac{1}{\sigma}A\vec{v}$ is also a unit vector, by Theorem 10.6.1

Therefore, for a non-zero singular value of A, we require unit vectors $\vec{v}$ and $\vec{u}$ such that

$$A\vec{v} = \sigma\vec{u} \quad \text{and} \quad A^T\vec{u} = \sigma\vec{v}$$

This derivation does not work for $\sigma = 0$, in this case, we only need one of the conditions to be satisfied.

Definition Singular Vectors

Let A be an $m \times n$ matrix. If $\vec{v} \in \mathbb{R}^n$ and $\vec{u} \in \mathbb{R}^m$ are unit vectors and $\sigma \neq 0$ is a singular value of A such that

$$A\vec{v} = \sigma\vec{u} \quad \text{and} \quad A^T\vec{u} = \sigma\vec{v}$$

then we say that $\vec{u}$ is a **left singular vector** of A corresponding to σ and $\vec{v}$ is a **right singular vector** of A corresponding to σ.
For $\sigma = 0$, if $\vec{u}$ is such a unit vector such that $A^T\vec{u} = \vec{0}$, then $\vec{u}$ is a **left singular vector** of A corresponding to $\sigma = 0$. If $\vec{v}$ is such a unit vector such that $A\vec{v} = \vec{0}$, then $\vec{0}$ is a **right singular vector** of A corresponding to $\sigma = 0$

The derivation earlier proves the following useful theorem

Theorem 10.6.5

Let A be an $m \times n$ matrix. If $\vec{v} \in \mathbb{R}^n$ is a unit eigenvector of A^TA corresponding to a non-zero singular value σ of A, then $\vec{u} = \frac{1}{\sigma}A\vec{v}$ is a left singular vector of A corresponding to σ

Theorem 10.6.6

Let A be an $m \times n$ matrix. A vector $\vec{v} \in \mathbb{R}^n$ is a right singular vector of A if and only if $\vec{v}$ is a unit eigenvector of A^TA. A vector $\vec{u} \in \mathbb{R}^m$ is a left singular vector of A if and only if $\vec{u}$ is a unit eigenvector of AA^T