Math 245

James Yu

November 3, 2017

Contents

Week 1

September 8 - September 13

1.1 Inner Product

Definition 1.1.1 (Inner Product Space)**.** An *inner product space* (over $\mathbb{C}$) is a *vector space* V and a function:

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$$

satisfying:

i. $\langle \vec{x} + \vec{y}, \vec{z} \rangle = \langle \vec{x}, \vec{z} \rangle + \langle \vec{y}, \vec{z} \rangle$ for all $\vec{x}, \vec{y}, \vec{z} \in V$

ii. $\langle c\vec{x}, \vec{y} \rangle = c \langle \vec{x}, \vec{y} \rangle$ for all $\vec{x}, \vec{y} \in V$ and $c \in \mathbb{C}$

iii. $\langle \vec{x}, \vec{y} \rangle = \overline{\langle \vec{y}, \vec{x} \rangle}$ for all $\vec{x}, \vec{y} \in V$

iv. $\langle \vec{x}, \vec{x}, \in \rangle \mathbb{R}_{>0}$ if $\vec{x} \neq \vec{0}$, $\langle \vec{x}, \vec{x} \rangle = 0$ otherwise

Example 1.1.1.1 (Standard Complex Inner Product)**.**

$$V = \mathbb{C}^n \qquad \langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = x_1\overline{y_1} + \ldots + x_n\overline{y_n}$$

Properties *i*, *ii*, and *iii* clearly hold. For *iv*, for any $\vec{x} = (a_1 + b_1 i, \ldots, a_n + b_n i)$

$$\langle \vec{x}, \vec{x} \rangle = (a_1 + b_1 i)\overline{(a_1 + b_1 i)} + \ldots + (a_n + b_n i)\overline{(a_n + b_n i)} \tag{1.1}$$
$$= a_1^2 + b_1^2 + \ldots + a_n^2 + b_n^2 \tag{1.2}$$

This is the *standard complex inner product.* If we replace $\mathbb{C}^n$ with $\mathbb{R}^n$ then we get the *standard real inner product* (dot product).

Example 1.1.1.2 (L^2 Inner Product)**.**

$$V = C([0,1]) \qquad \langle f, g \rangle = \int_0^1 f\overline{g}$$

This is called the L^2 inner product on V

Example 1.1.1.3 (Frobenius Inner Product)**.**

$$V = M_n(\mathbb{C}) \qquad\qquad \langle A, B\rangle = \operatorname{tr}\left(A\overline{B^{\intercal}}\right)$$

This is called the *Frobenius inner product* on V. It satisfies *iv* because, for $A = (a_{ij})$, $B = (b_{ij})$ we have:

$$\operatorname{tr}\left(A\overline{B^{\intercal}}\right) = \sum a_{ij}\overline{b_i j}$$

which is equivalent to the *standard complex inner product.*

1.2 Cauchy-Bunyakovsky-Schwarz Inequality

Definition 1.2.1 (Length)**.** If $\vec{v}$ is a vector in an *inner product space*, the *length* of $\vec{v}$ is:

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

Theorem 1.2.2 (Cauchy-Schwarz)**.** *Let $\vec{x}, \vec{y} \in V$ be vectors in an* inner product space*, then:*

$$|\langle \vec{x}, \vec{y}\rangle| \leq \|\vec{x}\|\,\|\vec{y}\|$$

Proof. If $\vec{y} = \vec{0}$, this is trivial. Otherwise, for any $c \in \mathbb{C}$

$$\begin{aligned} 0 &\leq \|\vec{x} - c\vec{y}\|^2 && (1.3)\\ &\leq \|\vec{x}\|^2 - \bar{c}\langle \vec{x}, \vec{y}\rangle - c\langle \vec{y}, \vec{x}\rangle + c\bar{c}\|y\|^2 && (1.4)\end{aligned}$$

So, let $c = \dfrac{\langle \vec{x}, \vec{y}\rangle}{\|y\|^2}$:

$$\begin{aligned} 0 &\leq \|\vec{x}\|^2 - \frac{\overline{\langle \vec{x}, \vec{y}\rangle}}{\|y\|^2}\langle \vec{x}, \vec{y}\rangle - \frac{\langle \vec{x}, \vec{y}\rangle}{\|y\|^2}\langle \vec{y}, \vec{x}\rangle + \frac{\langle \vec{x}, \vec{y}\rangle\,\overline{\langle \vec{x}, \vec{y}\rangle}}{\|y\|^4}\|y\|^2 && (1.5)\\ &\leq \|\vec{x}\|^2 - \frac{|\langle \vec{x}, \vec{y}\rangle|^2}{\|y\|^2} - \frac{|\langle \vec{x}, \vec{y}\rangle|^2}{\|y\|^2} + \frac{|\langle \vec{x}, \vec{y}\rangle|^2}{\|y\|^2} && (1.6)\\ |\langle \vec{x}, \vec{y}\rangle|^2 &\leq \|\vec{x}\|^2\,\|\vec{y}\|^2 && (1.7)\\ |\langle \vec{x}, \vec{y}\rangle| &\leq \|\vec{x}\|\,\|\vec{y}\| && (1.8)\end{aligned}$$

$\square$

Remark. We can define the angle between $\vec{x}, \vec{y}$ as $\cos^{-1}\dfrac{|\langle \vec{x}, \vec{y}\rangle|}{\|\vec{x}\|\,\|\vec{y}\|}$.

1.3 Orthogonality

Definition 1.3.1 (Orthogonality)**.** Two vectors $\vec{x}, \vec{y}$ are *orthogonal* if and only if $\langle \vec{x}, \vec{y} \rangle = 0$.

Definition 1.3.2 (Unit vector)**.** A *unit vector* is a vector of length 1.

Definition 1.3.3 (Orthogonal Set)**.** An *orthogonal set* is a set S where $\langle \vec{x}, \vec{y} \rangle = 0$ for all $\vec{x}, \vec{y} \in S$, $\vec{x} \neq \vec{y}$.

Definition 1.3.4 (Orthonormal Set)**.** An *orthonormal set* is an *orthogonal set* in which each vector is a *unit vector*.

Example 1.3.4.1 (Standard Basis)**.** The standard basis in $\mathbb{R}^n$ is *orthonormal*

Theorem 1.3.5 (Orthonormal Coordinates)**.** *Let* $\{\vec{v}_1, \ldots, \vec{v}_n\}$ *be an* orthogonal basis *of an* inner product space V*. Then for any* $\vec{x} \in V$ *we have:*

$$\vec{x} = \sum_{i=1}^{n} \frac{\langle \vec{x}_i, \vec{v}_i \rangle}{\|\vec{v}_i\|^2} \vec{v}_i$$

Proof. Write $\vec{x} = a_1\vec{v}_1 + \ldots + a_n\vec{v}_n$. Then, for any i:

$$\begin{aligned}
\langle \vec{x}, \vec{v}_i \rangle &= \langle a_1\vec{v}_1 + \ldots + a_n\vec{v}_n, \vec{v}_i \rangle && (1.9)\\
&= a_1 \langle \vec{v}_1, \vec{v}_i \rangle + \ldots + a_n \langle \vec{v}_n, \vec{v}_i \rangle && (1.10)\\
&= a_i \langle \vec{v}_i, \vec{v}_i \rangle && (1.11)\\
a_i &= \frac{\langle \vec{x}, \vec{v}_i \rangle}{\|\vec{v}_i\|^2} && (1.12)
\end{aligned}$$

□

Remark. If $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is *orthonormal*, then $\vec{x} = \sum_{i=1}^{n} \langle \vec{x}_i, \vec{v}_i \rangle \vec{v}_i$

Remark. The $\vec{v}_i$ coordinate of $\vec{x}$ depends only $\vec{x}$ and $\vec{v}_i$. It does not depend on any other vectors in the basis.

Remark. In finite dimensions, *inner product spaces* always have *orthnormal bases.*

Theorem 1.3.6 (Orthogonal $\implies$ Linear Independence)**.** *Let* S *be an* orthogonal set *of non-zero vectors, then* S *is linearly independent.*

Proof. For any $\vec{v}_1, \ldots, \vec{v}_n \in S$, set $a_1\vec{v}_1 + \ldots + a_n\vec{v}_n = \vec{0}$. By similar construction as 1.3.5, we can show for any i that:

$$a_i \langle \vec{v}_i, \vec{v}_i \rangle = 0$$

Since $\vec{v}_i \neq 0$ by assumption, $a_i = 0$ for all i. □

1.4 Gram-Schmidt Procedure

Given a basis $\{\vec{w}_1, \ldots \vec{w}_n\}$ for a (finite dimensional) *inner product space* V, the Gram-Schmidt gives an *orthogonal basis* for V as follows:

Step ① Set $\vec{v}_1 = \vec{w}_1$

Step ② Set $\vec{v}_2 = \vec{w}_2 - \frac{\langle \vec{w}_2, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2} \vec{v}_1$

...

Step ⓘ Set $\vec{v}_i = \vec{w}_i - \sum_{j=1}^{i-1} \frac{\langle \vec{w}_2, \vec{v}_i \rangle}{\|\vec{v}_i\|^2} \vec{v}_i$

Claim 1.4.1. $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is an *orthogonal basis* of V

Proof. We first check that $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is *orthogonal*.

We proceed by induction on i. If $n = 1$, we are vacuously done.

Otherwise, assume that $\{\vec{v}_1, \ldots, \vec{v}_i\}$ is *orthogonal*. For any $1 \leq j \leq i$ we have:

$$\begin{aligned}
\langle \vec{v}_{i+1}, \vec{v}_j \rangle &= \left\langle \vec{w}_{i+1} - \sum_{k=1}^{i} \frac{\langle \vec{w}_{i+1}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2} \vec{v}_k, \vec{v}_j \right\rangle && (1.13)\\
&= \left\langle \vec{w}_{i+1} - \frac{\langle \vec{w}_{i+1}, \vec{v}_j \rangle}{\|\vec{v}_j\|^2} \vec{v}_j, \vec{v}_j \right\rangle && (1.14)\\
&= \langle \vec{w}_{i+1}, \vec{v}_j \rangle - \frac{\langle \vec{w}_{i+1}, \vec{v}_j \rangle}{\|\vec{v}_j\|^2} \|\vec{v}_j\|^2 && (1.15)\\
&= \langle \vec{w}_{i+1}, \vec{v}_j \rangle - \langle \vec{w}_{i+1}, \vec{v}_j \rangle && (1.16)\\
&= 0 && (1.17)
\end{aligned}$$

Furthermore, $\vec{v}_i \neq \vec{0}$. For $i = 1$ we have $\vec{v}_1 = \vec{w}_1 \neq \vec{0}$ by assumption. Otherwise, we have $\vec{v}_i = \vec{w}_i - \vec{x}$ for $x \in \text{span}\{\vec{w}_1, \ldots, \vec{w}_{i-1}\}$. Thus $\vec{v}_i$ is a nonzero linear combination of $\{\vec{w}_1, \ldots, \vec{w}_i\}$ and is therefore non-zero.

Thus $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is a set of n vectors in an n-dimensional space that are orthogonal and nonzero. By 1.3.6 they are linearly independent, and thus a basis of V. □

Remark. To obtain an *orthonormal basis* of V, simply divide each $\vec{v}_i$ by its length. This is called *normalizing*.

1.5 Orthogonal Complement

Definition 1.5.1 (Orthogonal Complement)**.** Let V be an *inner product space* and $W \subset V$ a subspace. The orthogonal complement of W is:

$$W^{\perp} = \{\vec{v} \in V : \langle \vec{v}, \vec{w} \rangle = 0 \text{ for all } \vec{w} \in W\}$$

Example 1.5.1.1 (Orthogonal complement of entire space is empty)**.**

$$W = V \qquad W^{\perp} = \{\vec{v} \in V : \langle \vec{v}, \vec{w} \rangle = 0 \text{ for all } \vec{w} \in V\} = \left\{\vec{0}\right\}$$
$$\text{because } \langle \vec{v}, \vec{v} \rangle = 0 \implies \vec{v} = \vec{0}$$

Example 1.5.1.2 (Orthogonal complement of empty subspace is the entire space)**.**

$$W = \left\{\vec{0}\right\} \qquad W^{\perp} = V$$

Example 1.5.1.3 (Orthogonal complelment of a line is a plane)**.**

$$V = \mathbb{R}^3 \qquad W = \{(0,0,z) : z \in \mathbb{R}\} \qquad W^{\perp} = \{(x,y,0) : x, y \in \mathbb{R}\}$$

Week 2

September 15 - September 20

2.1 Orthogonal Complement (continued)

Theorem 2.1.1. *Let V be a finite-dimensional inner product space, and $W \subset V$ be a subspace, then:*

$$V \simeq W \oplus W^{\perp}$$

via the transformation $T : W \oplus W^{\perp} \to V$ given by:

$$T(\vec{w}, \vec{w}') = \vec{w} + \vec{w}'$$

Proof. We prove the theorem by writing an inverse for T. Let $\{\vec{w}_1, \ldots, \vec{w}_k\}$ be an orthonormal basis of W and define:

$$\Psi : V \to W \oplus W^{\perp} \tag{2.1}$$

$$\Psi(\vec{v}) = \left(\sum_{i=1}^{k} \langle \vec{v}, \vec{w}_i \rangle \, \vec{w}_i \ , \ \vec{v} - \sum_{i=1}^{k} \langle \vec{v}, \vec{w}_i \rangle \, \vec{w}_i \right) \tag{2.2}$$

Ψ is well defined since the first entry is in W by being a linear combination of $\vec{w}_i$, and the right entry is in $W^{\perp}$ because it is orthogonal to each $\vec{w}_i$ in our basis. It clear that $T \circ \Psi = \mathrm{id}_V$, so it remains to be shown that $\Psi \circ T = \mathrm{id}_{W \oplus W^{\perp}}$:

$$\begin{aligned}
\Psi\left(T\left(\vec{w}, \vec{w}'\right)\right) &= \Psi(\vec{w} + \vec{w}') && (2.3)\\
&= \left(\sum_{i=1}^{k} \langle \vec{w} + \vec{w}', \vec{w}_i \rangle \, w_i \ , \ \vec{w} + \vec{w}' - \sum_{i=1}^{k} \langle \vec{w} + \vec{w}', \vec{w}_i \rangle \, w_i \right) && (2.4)\\
&= \left(\sum_{i=1}^{k} \langle \vec{w}, \vec{w}_i \rangle \, w_i \ , \ \vec{w} + \vec{w}' - \sum_{i=1}^{k} \langle \vec{w}, \vec{w}_i \rangle \, w_i \right) && (2.5)\\
&= (\vec{w}, \vec{w}') && (2.6)
\end{aligned}$$

Thus T and Ψ are inverses. Since T and Ψ are linear transformations, T is an isomorphism. $\square$

Corollary 2.1.1.1 (Extension of orthonormal basis)**.** *Let* $\{\vec{w}_1, \ldots, \vec{w}_k\}$ *be an orthonormal basis of a subspace* W*. One can extend this to an orthornomal basis of the entire space:*

$$\{\vec{w}_1, \ldots, \vec{w}_k, \vec{v}_{k+1}, \ldots, \vec{v}_n\}$$

where $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ *is an orthonormal basis of* $W^{\perp}$*.*

Corollary 2.1.1.2 (Dimension of orthogonal complement)**.**

$$\dim V = \dim W + \dim W^{\perp}$$

Corollary 2.1.1.3 (Duality of orthogonal complement)**.**

$$\left(W^{\perp}\right)^{\perp} = W$$

Corollary 2.1.1.4 (Intersection of subspace and orthogonal complement)**.**

$$W \cap W^{\perp} = (0)$$

Definition 2.1.2 (Orthogonal Projection onto a subspace)**.** Let $W \subset V$ be a subspace and $\vec{v} \in V$. Then for $\Psi : V \to W \oplus W^{\perp}$ as defined in 2.1.1, we define the *orthogonal projection of* $\vec{v}$ *onto* W to be the first coordinate $\Psi(\vec{v})$, denoted:

$$\operatorname{proj}_W(\vec{v})$$

Remark. If $\{\vec{w}_1, \ldots, \vec{w}_k\}$ is an orthonormal basis of W, then:

$$\operatorname{proj}_W(\vec{v}) = \sum_{i=1}^{k} \langle \vec{v}, \vec{w}_i \rangle \vec{w}_i$$

2.2 Adjoints

Definition 2.2.1 (Conjugate Transpose)**.** For any matrix B, we define B^* to be the *conjugate transpose* given by taking the conjugate of each entry in $B^{\intercal}$, that is:

$$B^* = \overline{B^{\intercal}}$$

Lemma 2.2.2 (Unique inner product form of a linear transformation)**.** *Let* $\mathcal{U} : V \to \mathbb{F}$ *be a linear transformation, then there exists some unique* $z \in V$ *such that:*

$$\mathcal{U}(\vec{v}) = \langle \vec{v}, \vec{z} \rangle \textit{ for all } \vec{v} \in V$$

Proof. Let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ be an orthonormal basis of V and define $\vec{z} \in V$ to be:

$$\vec{z} = \overline{\mathcal{U}(\vec{v}_1)}\vec{v}_1 + \ldots + \overline{\mathcal{U}(\vec{v}_n)}\vec{v}_n$$

Then we check that $\mathcal{U}(\vec{v}) = \langle \vec{v}, \vec{z} \rangle$ for all $\vec{v} \in V$:

$$\begin{aligned} \mathcal{U}(\vec{v}) &= \mathcal{U}\left(a_1\vec{v}_1 + \ldots + a_n\vec{v}_n\right) && (2.7) \\ &= a_1\mathcal{U}(\vec{v}_1) + \ldots + a_n\mathcal{U}(\vec{v}_n) && (2.8) \end{aligned}$$

$$\begin{aligned} \langle \vec{v}, \vec{z} \rangle &= \left\langle \vec{v}, \overline{\mathcal{U}(\vec{v}_1)}\vec{v}_1 + \ldots + \overline{\mathcal{U}(\vec{v}_n)}\vec{v}_n \right\rangle && (2.9) \\ &= \mathcal{U}(\vec{v}_1)\langle \vec{v}, \vec{v}_1 \rangle + \ldots + \mathcal{U}(\vec{v}_n)\langle \vec{v}, \vec{v}_n \rangle && (2.10) \\ &= a_1\mathcal{U}(\vec{v}_1) + \ldots + a_n\mathcal{U}(\vec{v}_n) && (2.11) \\ &= \mathcal{U}(\vec{v}) && (2.12) \end{aligned}$$

To show that $\vec{z}$ is unique, suppose that $\langle \vec{v}, \vec{z} \rangle = \langle \vec{v}, \vec{z}' \rangle$ for all $\vec{v} \in V$, then:

$$0 = \langle \vec{v}, \vec{z} \rangle - \langle \vec{v}, \vec{z}' \rangle = \langle \vec{v}, \vec{z} - \vec{z}' \rangle$$

Since this holds for all $\vec{v}$, we must have $\vec{z} - \vec{z}' = 0$ (indeed, $V^{\perp} = (0)$), we have $\vec{z}' = \vec{z}$ as required. □

Theorem 2.2.3 (Existence of unique adjoint)**.** *Let $T : V \to V$ be a linear transformation on an inner product space V. There exists a unique linear transformation $T^* : V \to V$ satisfying:*

$$\langle T\vec{x}, \vec{y} \rangle = \langle \vec{x}, T^*\vec{y} \rangle \text{ for all } \vec{x}, \vec{y} \in V$$

This T^* is called the adjoint *of T.*

Proof. For any $\vec{y} \in V$, define $g_{\vec{y}} : V \to \mathbb{F}$ (where $\mathbb{F}$ is $\mathbb{C}$ or $\mathbb{R}$), by:

$$g_{\vec{y}}(\vec{v}) = \langle T\vec{v}, \vec{y} \rangle$$

Then $g_{\vec{y}}$ is a linear transformation because:

$$\begin{aligned} g_{\vec{y}}\left(\vec{v} + \vec{w}\right) &= \langle T\left(\vec{v} + \vec{w}\right), \vec{y} \rangle && (2.13) \\ &= \langle T\vec{v} + T\vec{w}, \vec{y} \rangle && (2.14) \\ &= \langle T\vec{v}, \vec{y} \rangle + \langle T\vec{w}, \vec{y} \rangle && (2.15) \\ &= g_{\vec{y}}(\vec{v}) + g_{\vec{y}}(\vec{w}) && (2.16) \end{aligned}$$

$$\begin{aligned} g_{\vec{v}}\left(c\vec{v}\right) &= \langle T\left(c\vec{v}\right), \vec{y} \rangle && (2.17) \\ &= c\langle T\vec{v}, \vec{y} \rangle && (2.18) \\ &= cg_{\vec{y}}(\vec{v}) && (2.19) \end{aligned}$$

Then we can define $T^* : V \to V$ by the map from $\vec{y} \in V$ to the unique $\vec{z}$ generated by 2.2.2 for $g_{\vec{y}}$. Then, by definition of $\vec{z}$ we have:

$$\langle T\vec{x}, \vec{y} \rangle = g_{\vec{y}}(\vec{x}) = \langle \vec{x}, \vec{z} \rangle = \langle \vec{x}, T^*\vec{y} \rangle$$

By uniqueness of $\vec{z}$, this mapping T^* is unique. Thus it remains only to show that T^* is linear. For all $\vec{x}$, $\vec{y}$, $\vec{z} \in V$ and $c \in \mathbb{F}$:

$$\begin{aligned} \langle \vec{x}, T^*(c\vec{y}) \rangle &= \langle T\vec{x}, c\vec{y} \rangle && (2.20) \\ &= \bar{c} \langle T\vec{x}, \vec{y} \rangle && (2.21) \\ &= \bar{c} \langle \vec{x}, T^*\vec{y} \rangle && (2.22) \\ &= \langle \vec{x}, cT^*\vec{y} \rangle && (2.23) \end{aligned}$$

By the same orthogonal complement dimension argument as in 2.2.2, since this holds for all $\vec{x}$, $T^*(c\vec{y}) = cT^*\vec{y}$ as required. Similarly:

$$\begin{aligned} \langle \vec{x}, T^*(\vec{y} + \vec{z}) \rangle &= \langle T\vec{x}, \vec{y} + \vec{z} \rangle && (2.24) \\ &= \langle T\vec{x}, \vec{y} \rangle + \langle T\vec{x}, \vec{z} \rangle && (2.25) \\ &= \langle \vec{x}, T^*\vec{y} \rangle + \langle vecx, T^*\vec{z} \rangle && (2.26) \\ &= \langle \vec{x}, T^*\vec{y} + T^*\vec{z} \rangle && (2.27) \end{aligned}$$

Again, by the argument used in 2.2.2, since this holds for all $\vec{x}$, we have $T^*(\vec{y} + \vec{z}) = T^*\vec{y} + T^*\vec{z}$ as required. Thus T^* is unique and linear as required. □

Theorem 2.2.4 (Equivalence of conjugate transpose and adjoint)**.** *If B is an orthonormal basis of V, then:*

$$[T]_B^* = [T^*]_B$$

Proof. Let $B = \{\vec{v}_1, \ldots, \vec{v}_n\}$, $[T]_B = (a_{ij})$ and $[T^*]_B = (b_{ji})$. Then for any i, j:

$$b_{ij} = \langle T^*\vec{v}_j, \vec{v}_i \rangle = \langle \vec{v}_j, T\vec{v}_i \rangle = \overline{\langle T\vec{v}_i, \vec{v}_j \rangle} = \overline{a_{ji}}$$

□

2.3 Least Squares (example)

Say $\{(x_1, y_1), \ldots, (x_m, y_n)\}$ is a set of points in $\mathbb{R}^2$ and we want to find the line that best fits the data. More precisely, we want to find $a, b \in \mathbb{R}$ such that the line $y = ax + b$ minmizes the quantity:

$$E = \sum_{i=1}^{m} |y_i - (ax_i + b)|^2$$

Define the following:

$$A = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} a \\ b \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

Then we can rewrite the error E as:

$$E = \|A\vec{x} - \vec{y}\|^2$$

This is minimized when $A\vec{x} = \mathrm{proj}_{\mathrm{im}\,A}(\vec{y})$, so we just need to find $\vec{x}$ given $A\vec{x}$.

Remark (Author's Note)**.** In the following section we will take the adjoint of A even though $A : \mathbb{R}^2 \to \mathbb{R}^n$ and we only defined the adjoint for endofunctions. You can justify this to yourself by imagining the extensions of A and $\vec{x}$ given by:

$$A' = \begin{bmatrix} x_1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_m & 1 & 0 & \dots & 0 \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} a \\ b \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

And convincing yourself that we have not fundamentally changed the problem. In general though, there is something called the Hermitian adjoint which properly describes this property for $A : H_1 \to H_2$ where H_1 and H_2 are Hilbert spaces that have an appropriate inner product. I don't know anything about Hilbert spaces, this just happens to be a thing that exists.

We know that $\langle A\vec{x}, \vec{y}\rangle = \langle \vec{x}, A^*\vec{y}\rangle$. Thus, if $A^*A\vec{x} = \vec{0}$ we have:

$$\begin{aligned} A^*A\vec{x} = \vec{0} &\implies \langle A^*A\vec{x}, \vec{x}\rangle = 0 && (2.28) \\ &\implies \langle A\vec{x}, A\vec{x}\rangle = 0 && (2.29) \\ &\implies A\vec{x} = \vec{0} && (2.30) \end{aligned}$$

This tells us that if $\ker A = (\vec{0})$, then $\ker(A^*A) = (\vec{0})$ meaning A^*A is invertible. In any practical case $\ker A = (\vec{0})$ since, otherwise, that would mean all of our x_is are equal, so our line doesn't represent anything interesting. Thus, if $\vec{x}$ is our best fit line, then:

$$\begin{aligned} & A\vec{x} - \vec{y} \in (\operatorname{im} A)^\perp && (2.31) \\ \implies & \langle A\vec{z}, A\vec{x} - \vec{y}\rangle = 0 \text{ for all } \vec{z} \in \mathbb{R}^2 && (2.32) \\ \implies & \langle \vec{z}, A^*\left(A\vec{x} - \vec{y}\right)\rangle = 0 && (2.33) \\ \implies & A^*\left(A\vec{x} - \vec{y}\right) = 0 && (2.34) \\ \implies & A^*A\vec{x} = A^*\vec{y} && (2.35) \\ \implies & \vec{x} = \left(A^*A\right)^{-1} A^*\vec{y} && (2.36) \end{aligned}$$

Week 3

September 22 - September 27

3.1 Normal Operators

Definition 3.1.1. Let $T : V \to V$ be a linear transformation on an inner product space V. We say T is *normal* if:

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

Remark. If there exists an orthonormal basis B such that $[T]_B$ is diagonal, then $[T^*]_B = [T]_B^*$ is also diagonal thus:

$$[T]_B^*[T]_B = [T]_B[T]^*B \tag{3.1}$$

$$T^*T = TT^* \tag{3.2}$$

So T is *normal.*

Definition 3.1.2. Let $T : V \to V$ be a linear transformation on a vector space V, and let W be a subspace of V. We say W is T-invariant if, for all $\vec{w} \in W$, $T\vec{w} \in W$.

Lemma 3.1.3 (Schur)**.** *Let $T : V \to V$ be a linear transformation on an inner product space V. If the characteristic polynomial of T splits completely, then there is an orthonormal basis B of V such that $[T]_B$ is upper triangular.*

Proof. We induce on $\dim V$. The case $\dim V = 1$ is trivial since all 1×1 matrices are upper triangular. So we assume the lemma holds for all inner product spaces W with $\dim W < \dim V$. Since the characteristic polynomial splits completely, there is some eigenvector $\vec{v} \in V$ and corresponding eigenvalue λ satisfying:

$$T\vec{v} = \lambda\vec{v}$$

Thus, for any $\vec{x} \in V$:

$$0 = \langle (T - \lambda I)\, \vec{v}, \vec{x} \rangle \tag{3.3}$$

$$= \left\langle \vec{v}, \left(T^* - \overline{\lambda} I\right) \vec{x} \right\rangle \tag{3.4}$$

Which means that $\vec{v} \in \left(\operatorname{im}\left(T^* - \overline{\lambda} I\right)\right)^\perp$. Thus $\left(T^* - \overline{\lambda} I\right)$ is not surjective, so by rank-nullity theorem, there is some nonzero $\vec{z} \in \ker\left(T^* - \overline{\lambda} I\right)$, giving:

$$\left(T^* - \overline{\lambda} I\right) \vec{z} = 0 \tag{3.5}$$
$$T^* \vec{z} = \overline{\lambda} \vec{z} \tag{3.6}$$

Without loss of generality, assume that $\|\vec{z}\| = 1$, since the equality holds under scalar multiplication of $\vec{z}$. Let $W = \operatorname{span}\{\vec{z}\}$, then W is T^*-invariant. Then, for all $\vec{y} \in W^\perp$:

$$\langle T\vec{y}, c\vec{z}\rangle = \overline{c}\,\langle \vec{y}, T^*\vec{z}\rangle \tag{3.7}$$
$$= \overline{c}\lambda\,\langle \vec{y}, \vec{z}\rangle \tag{3.8}$$
$$= 0 \text{ by choice of } \vec{y} \tag{3.9}$$

Thus $W^\perp$ is T-invariant. This means $T\big|_{W^\perp} : W^\perp \to W^\perp$ is a linear transformation (whose characteristic polynomial splits completely, proof omitted in this class but this follows from the fact that T splits), and $\dim W^\perp = \dim V - 1$. Thus, by our inductive hypothesis there exists an orthonormal basis $\beta = \{\vec{v}_1, \ldots, \vec{v}_{n-1}\}$ of $W^\perp$ such that $\left[T\big|_{W^\perp}\right]_P$ is upper triangular. Thus:

$$[T^*]_B = \begin{bmatrix} \left[T^*\big|_{W^\perp}\right]_\beta & 0 \\ * & \overline{\lambda} \end{bmatrix} \tag{3.10}$$
$$[T]_B = \begin{bmatrix} \left[T\big|_{W^\perp}\right]_\beta & * \\ 0 & \lambda \end{bmatrix} \tag{3.11}$$

which is upper triangular. □

Theorem 3.1.4 (Orthonormal Diagonalizability of Complex Linear Transformations)**.** *If $T : V \to V$ is a* normal *linear transformation on a complex inner product space V, then there exists an orthonormal basis B such that $[T]_B$ is diagonal.*

Proof. Since all polynomials split over $\mathbb{C}$, by 3.1.3, there is an orthonormal basis:

$$B = \{\vec{v}_1, \ldots, \vec{v}_n\}$$

such that $[T_B]$ is upper triangular. We will show that $[T]_B$ is also diagonal. Let $[T]_B = (a_{ij})$, we will show that $a_{ij} = 0$ if $i \neq j$ by induction on j. If $j = 1$, this is immediate from upper triangularity, so if the claim holds for all $j' < j$. If $i < j$ then:

$$0 = \|T\vec{v}_i - \lambda\vec{v}_i\|^2 \text{ for } \lambda = a_{ii} \tag{3.12}$$
$$= \langle T\vec{v}_i - \lambda\vec{v}_i, T\vec{v}_i - \lambda\vec{v}_i\rangle \tag{3.13}$$
$$= \langle (T - \lambda I)\vec{v}_i, (T - \lambda I)\vec{v}_i\rangle \tag{3.14}$$
$$= \langle \vec{v}_i, (T - \lambda I)^*(T - \lambda I)\vec{v}_i\rangle \tag{3.15}$$
$$= \langle \vec{v}_i, (T - \lambda I)(T - \lambda I)^*\vec{v}_i\rangle \tag{3.16}$$
$$= \left\langle (T^* - \overline{\lambda} I)\vec{v}_i, (T^* - \overline{\lambda} I)\vec{v}_i\right\rangle \tag{3.17}$$
$$= \|T^*\vec{v}_i - \lambda\vec{v}_i\| \tag{3.18}$$

Thus $T^*\vec{v}_i = \overline{\lambda}\vec{v}_i$. Then consider:

$$T\vec{v}_j = a_{1j}\vec{v}_1 + \cdots + a_{jj}\vec{v}_j$$

By orthonormality of our basis, it follows that:

$$\begin{aligned} a_{ij} &= \langle T\vec{v}_j, \vec{v}_i \rangle && (3.19)\\ &= \langle \vec{v}_j, T^*\vec{v}_i \rangle && (3.20)\\ &= \left\langle \vec{v}_j, \overline{\lambda}\vec{v}_i \right\rangle && (3.21)\\ &= 0 && (3.22) \end{aligned}$$

As required, each entry a_{ij} with $i < j$ is 0, and entries with $i > j$ follow from upper triangularity. □

Corollary 3.1.4.1. *If $T : V \to V$ is a linear transformation on a complex inner product space V, then there exists an othonormal basis B such that $[T]_B$ is diagonal if and only if T is* normal*.*

Remark. 3.1.4.1 does not apply to real inner product spaces. Consider $V : \mathbb{R}^2$ and $T : \mathbb{R}^2 \to R^2$ given by the rotation matrix:

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

Then T^* describes the opposite rotation, thus $T^*T = TT^* = I$ so T is *normal*, however if $\theta \notin \pi\mathbb{Z}$, T has no real eigenvectors and is thus not diagonalizable.

3.2 Self-Adjoint Operators

Definition 3.2.1. A linear transformation T is *self-adjoint* (or Hermitian) if $T = T^*$.

Remark. If $T : V \to V$ is a linear transformation on a real inner product space V, and there exists an orthonormal basis B for which $[T]_B$ is diagonal, then:

$$[T]_B = [T]_B^*$$

So T is *self-adjoint*

Remark. If $T = T^*$, then $T^*T = TT^*$ so T is normal.

Theorem 3.2.2 (Orthonormal Diagonalizability of Real Linear Transformations)**.** *If $T : V \to V$ is a linear transformation on a real inner product space V, then T is* self-adjoint *if and only if there is an orthonormal basis B such that $[T]_B$ is diagonal.*

Proof. Note that the characteristic polynomial of T must split over $\mathbb{C}$, so consider any eigenvector $\vec{x}$ and eigenvalue $\lambda \in \mathbb{C}$ such that $T\vec{x} = \lambda\vec{x}$, then:

$$(T - \lambda I)\vec{x} = \vec{0} \implies (T^* - \overline{\lambda} I)\vec{x} = 0(\text{see proof of 3.1.4}) \tag{3.23}$$
$$\implies T^*\vec{x} = \overline{\lambda} x \tag{3.24}$$

So if T is *self-adjoint*:

$$\overline{\lambda}\vec{x} = T^*\vec{x} = T\vec{x} = \lambda\vec{x} \tag{3.25}$$
$$\overline{\lambda} = \lambda \tag{3.26}$$

Thus $\lambda \in \mathbb{R}$, so all eigenvalues of T are real. Thus the characteristic polynomial of T splits completely over $\mathbb{R}$, so invoking 3.1.3, there must exist an orthonormal basis B such that $[T]_B$ is upper triangular. However $[T]_B^* = [T^*]_B$ which must be lower triangular, so $[T]_B$ is both upper and lower triangular, meaning $[T]_B$ is diagonal. □

Corollary 3.2.2.1 (Orthonormal Diagonalizability of Symmetric Real Matrices)**.** *A real matrix is orthogonally diaganalizable if and only if it's symmetric.*

Proof. A real matrix that is *self-adjoint* is just a symmetric matrix, so this follows immediately from 3.2.2. □

3.3 Isometries

Definition 3.3.1. A linear transformation $T : V \to W$ from an inner product space V to an inner product space W is an *isometry* if $\langle T\vec{x}, T\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle$ for all $\vec{x}$, $\vec{y} \in V$.

Definition 3.3.2. An *isometry* T is *unitary* if T is surjective.

Definition 3.3.3. Let V, W be inner product spaces. If there exists a *unitary isometry* $T : V \to W$, we say V and W are *isometric*.

Remark. Every *isometry* T is injective because:

$$T\vec{x} = \vec{0} \implies \langle \vec{x}, \vec{x}\rangle = \langle T\vec{x}, T\vec{x}\rangle = 0 \tag{3.27}$$
$$\implies \vec{x} = 0 \tag{3.28}$$

Thus $\ker T = (0)$.

Remark (Author's Note)**.** Again, in this section, we will use the adjoint of T even if T is not an endomorphism. In finite dimensional vector spaces, this exists, and the conjugate transpose of the matrix representation still works, you'll just have to convince yourself.

Remark. For every *isometry* $T : V \to W$, $T^*T = I$ since, for all $\vec{x}, \vec{y} \in V$:

$$\langle T\vec{x}, T\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle \tag{3.29}$$
$$\langle \vec{x}, T^*T\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle \text{ for any } \vec{x} \tag{3.30}$$

Remark. If T is *unitary*, T is invertible so $TT^* = I = T^*T$, so T is also normal.

Lemma 3.3.4. *Let $\mathcal{U} : V \to V$ be a self-adjoint linear transformation, and $\langle \vec{x}, \mathcal{U}\vec{x} \rangle = 0$ for all $\vec{x} \in V$, then $\mathcal{U} = 0$.*

Proof. Suppose $\vec{x}$ is an eigenvector of $\mathcal{U}$ and λ be its corresponding eigenvalue, then:

$$\begin{aligned} 0 &= \langle \vec{x}, \mathcal{U}\vec{x} \rangle && (3.31) \\ &= \langle \vec{x}, \lambda\vec{x} \rangle && (3.32) \\ &= \overline{\lambda} \langle \vec{x}, \vec{x} \rangle && (3.33) \end{aligned}$$

But $\vec{x} \neq \vec{0}$ by choice of $\vec{x}$ being an eigenvector, so $\lambda = 0$. Since all eigenvalues of $\mathcal{U}$ are 0 and $\mathcal{U}$ is diagonalizable (since it is self-adjoint), $\mathcal{U} = 0$. □

Theorem 3.3.5. *Let $T : V \to V$ be a surjective linear transformation on a finite dimensional inner product space V, then the following are equivalent:*

i. $TT^* = T^*T = I$

ii. $\langle T\vec{x}, T\vec{y} \rangle = \langle \vec{x}, \vec{y} \rangle$ *for all* $\vec{x}, \vec{y} \in V$

iii. *If B is an orthonormal basis, then so is $T(B)$*

iv. *There exists an orthonormal basis B such that $T(B)$ is also orthonormal*

v. $\|T\vec{x}\| = \|\vec{x}\|$ *for all* $\vec{x} \in V$.

Proof. We will prove a ring of implications:

i. $\implies$ ii. For all $\vec{x}, \vec{y} \in V$:

$$\begin{aligned} \langle T\vec{x}, T\vec{y} \rangle &= \langle \vec{x}, T^*T\vec{y} \rangle && (3.34) \\ &= \langle \vec{x}, I\vec{y} \rangle && (3.35) \\ &= \langle \vec{x}, \vec{y} \rangle && (3.36) \end{aligned}$$

ii. $\implies$ iii. Let B be a basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$, then for any $\vec{v}_i$, $\vec{v}_j \in B$:

$$\langle T\vec{v}_i, T\vec{v}_j \rangle = \langle \vec{v}_i, \vec{v}_j \rangle = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Thus $T(B)$ is orthonormal. Recall from 1.3.6 that this is sufficient to show $T(B)$ is linearly independent and thus a basis.

iii. $\implies$ iv. Immediate from the fact that V is finite dimensional so an orthonormal basis exists.

iv. $\implies$ v. For any $\vec{x} \in V$ write:

$$\vec{x} = a_1\vec{v}_1 + \cdots + a_n\vec{v}_n$$

where $\{\vec{v}_1, \ldots, \vec{v}_n\}$ are a subset of the orthonormal basis B provided by the assumption. Then:

$$\begin{aligned}
\|T\vec{x}\|^2 &= \|T(a_1\vec{v}_1 + \cdots + a_n\vec{v}_n)\|^2 && (3.37)\\
&= \|a_1T\vec{v}_1 + \cdots + a_nT\vec{v}_n\|^2 && (3.38)\\
&= |a_1|^2 + \cdots + |a_n|^2 && (3.39)\\
&= \|\vec{x}\|^2 && (3.40)
\end{aligned}$$

By non-negativity of the norm, v. holds.

v. $\implies$ i. From our assumption, for all $\vec{x}$:

$$\begin{aligned}
\|\vec{x}\| &= \|T\vec{x}\| && (3.41)\\
&= \langle T\vec{x}, T\vec{x}\rangle && (3.42)\\
&= \langle \vec{x}, T^*T\vec{x}\rangle && (3.43)
\end{aligned}$$

We have $\langle \vec{x}, (T^*T - I)\vec{x}\rangle = 0$ for all $\vec{x}$. Note that $(T^*T - I)$ is self-adjoint because $(T^*T - I)^* = T^*T - I$. Thus by 3.3.4, $T^*T - I = 0$ so $T^*T = I$. Thus T^* is a left inverse of T, so since T is an endomorphism, $T^*T = I = TT^*$.

□

Corollary 3.3.5.1. *Let V, W be* isometric *finite dimensional inner product spaces, then* $\dim V = \dim W$.

Corollary 3.3.5.2. *If* $\dim V = \dim W$ *for finite dimensional inner product spaces V, W, then V, W are* isometric.

Proof. Since V and W are finite dimensional, they have orthonormal bases $\{\vec{v}_1, \ldots, \vec{v}_n\}$, $\{\vec{w}_1, \ldots, \vec{w}_n\}$. Then we can define a linear transformation $T : V \to W$ given by $T(\vec{v}_i) = \vec{w}_i$ for all i. By 3.3.5, T is an *isometry.* □

Corollary 3.3.5.3. *Any n-dimensional inner product space is* isometric *to $\mathbb{R}^n$ or $\mathbb{C}^n$ with the standard inner product.*

Corollary 3.3.5.4. *If $T : V \to W$ is* unitary*, then its eigenvalues all have absolute value 1.*

Proof. For all $\vec{x} \in T$:

$$\|\vec{x}\| = \|T\vec{x}\| = \|\lambda\vec{x}\| = |\lambda|\,\|\vec{x}\|$$

Thus for any eigenvalue λ, $|\lambda| = 1$. □

Week 4

September 29 - October 4

4.1 Orthogonal Matrices

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal linear transformation:

Definition 4.1.1. We say B is an *eigenbasis* for T if B is an orthonormal basis of eigenvectors of T.

Remark. If $n = 1$, T is one of the following:

$$[1] \qquad\qquad [-1]$$

Remark. If $n = 2$, and A is the matrix for T, A must be a real matrix satisfying:

$$AA^{\intercal} = A^{\intercal}A = I$$

and since $\{(1,0),(0,1)\}$ is an orthonormal basis we must have:

$$\left\| A \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\| = \left\| A \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\| = 1$$

thus A must be of the following form for some $\theta \in [0, 2\pi)$:

$$A = \begin{pmatrix} \cos\theta & \pm\sin\theta \\ \sin\theta & \mp\cos\theta \end{pmatrix}$$

Remark. If $n = 2$, by lifting T to being a unitary transformation $\mathbb{C}^n \to \mathbb{C}^n$, we can distinguish between rotations and reflections from the eigenvectors and eigenvalues of A. We know the eigenvalues must be complex numbers of length 1, so if they are real, they are ± 1. So let A be the matrix of T under an eigenbasis, it must be of the form:

$$\underset{\text{rotation}}{\begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}} \quad \text{or} \quad \underset{\text{reflection}}{\begin{bmatrix} \pm 1 & 0 \\ 0 & \mp 1 \end{bmatrix}}$$

Otherwise, if the eigenvalues are not real, they are of the form $\cos\theta + i\sin\theta$ for some $\theta \in [0, 2\pi)$ Matrices of the form:

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

have eigenvalues $\cos\theta \pm i\sin\theta$ and are rotations.

Remark. An orthogonal 2x2 matrix can be the composition of a rotation and a reflection.

Theorem 4.1.2. *Let A be a real, orthogonal, $n \times n$ matrix. Then A is block diagonal with blocks of size* 0 *or* 1.

Proof. Lift A to a $n \times n$ complex, unitary matrix. Then, since the entries are real

$$A\vec{x} = \lambda\vec{x} \text{ for } \vec{x} \neq 0 \implies A\vec{x} = \overline{\lambda\vec{x}}$$

Thus non-real eigenvalues come in conjugate pairs. Since A is unitary as a complex matrix, we can find an *eigenbasis* B of $\mathbb{C}^n$ for A. Then consider an arbitrary pair of eigenvalues $\vec{v}$ and $\vec{w} = \overline{\vec{v}}$. We want to find two real vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ such that $\text{span}\{\vec{x}, \vec{y}\} = \text{span}\{\vec{v}, \vec{w}\}$ over $\mathbb{C}$. So define

$$\vec{x} = \vec{v} + \vec{w} \qquad (= 2\Re(\vec{v})) \tag{4.1}$$

$$\vec{y} = i\vec{v} + i\vec{w} \qquad (= -2\Im(\vec{v})) \tag{4.2}$$

Clearly, by definition, $\vec{x}, \vec{y} \in \text{span}\{\vec{v}, \vec{w}\}$, and furthermore we have:

$$\vec{v} = \frac{1}{2i}(i\vec{x} + \vec{y}) \tag{4.3}$$

$$\vec{w} = \frac{1}{2i}(i\vec{x} - \vec{y}) \tag{4.4}$$

$$\tag{4.5}$$

Thus $\vec{w}, \vec{w} \in \text{span}\{\vec{v}, \vec{w}\}$. Applying Gram-Schmidt allows us to turn $\{\vec{x}, \vec{y}\}$ into a real orthonormal basis of $\text{span}\{\vec{v}, \vec{w}\}$. Doing this for every conjugate pair of non-real $\vec{v}_i \in B$ gives us a new, real orthonormal basis B' such that:

$$[A]_{b'} = \begin{pmatrix} (2\times 2) & 0 & \dots & 0 \\ 0 & (2\times 2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (1\times 1) \end{pmatrix}$$

where each block is also orthogonal matrix. □

Remark. This means that any orthogonal transformation T, when viewed under the right basis, is a collection of pairwise orthogonal rotations (2×2 blocks) together with some fixed and reflected lines (± 1 eigenvalues).

Example 4.1.2.1. In $\mathbb{R}^3$, an orthogonal matrix A may look like:

$$A = \begin{bmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} & 0 \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

4.2 Rigid Motions

Definition 4.2.1. A *rigid motion* is a function $f : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\|\vec{x} - \vec{y}\| = \|f(\vec{x}) - f(\vec{y})\| \text{ for all } \vec{x}, \vec{y} \in \mathbb{R}^n$$

That is, f preserves distances.

Example 4.2.1.1. A translation $\{\vec{x} \mapsto \vec{x} + \vec{a}\}$ is a *rigid motion*

Example 4.2.1.2. An orthogonal linear transformation is a *rigid motion*

Theorem 4.2.2. *Any* rigid motion $f : \mathbb{R}^n \to \mathbb{R}^n$ *can be written uniquely as*

$$f = g \circ T$$

where g is a translation and T is an orthogonal linear transformation

Proof. Define $T : \mathbb{R}^n \to \mathbb{R}^n$ by

$$T(\vec{x}) = f(\vec{x}) - f(\vec{0})$$

T is clearly a *rigid motion*, and $T(\vec{0}) = f(\vec{0}) - f(\vec{0}) = \vec{0}$. Also $f = g \circ T$ where g is the translation $g(\vec{x}) = \vec{x} + f(\vec{0})$. We will prove that T is linear. First observe that, for any $\vec{x} \in \mathbb{R}^n$

$$\|T\vec{x}\| = \left\|T\vec{x} - T\vec{0}\right\| = \left\|\vec{x} - \vec{0}\right\| = \|\vec{x}\|$$

Next we will show that, for any $\vec{x}, \vec{y} \in \mathbb{R}^n$

$$\langle T\vec{x}, T\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle$$

this is true because

$$\begin{aligned}
\|T\vec{x} - T\vec{y}\|^2 &= \langle T\vec{x} - T\vec{y}, T\vec{x} - T\vec{y}\rangle && (4.6)\\
&= \langle T\vec{x}, T\vec{x}\rangle - \langle T\vec{x}, T\vec{y}\rangle - \langle T\vec{y}, T\vec{x}\rangle + \langle T\vec{y}, T\vec{y}\rangle && (4.7)\\
&= \|T\vec{x}\|^2 + \|T\vec{y}\|^2 - 2\langle T\vec{x}, T\vec{y}\rangle \qquad = \|\vec{x}\|^2 + \|\vec{y}\|^2 - 2\langle T\vec{x}, T\vec{y}\rangle && (4.8)
\end{aligned}$$

but also

$$\begin{aligned}
\|T\vec{x} - T\vec{y}\|^2 &= \|\vec{x} - \vec{y}\|^2 && (4.9)\\
&= \|\vec{x}\|^2 + \|\vec{y}\|^2 - 2\langle \vec{x}, T\vec{y}\rangle && (4.10)
\end{aligned}$$

subtracting these two equations from each other we obtain

$$\langle T\vec{x}, T\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle$$

Using this fact we can show the two properties of linearity:

i. For any $a \in \mathbb{R}$, $\vec{x} \in \mathbb{R}^n$:

$$\begin{aligned} \|T(a\vec{x}) - aT(\vec{x})\|^2 &= \|T(a\vec{x})\|^2 + \|aT(\vec{x})\|^2 - 2\langle T(a\vec{x}), aT(\vec{x})\rangle && (4.11)\\ &= \|a\vec{x}\|^2 + a^2\|\vec{x}\|^2 - 2a\langle a\vec{x}, \vec{x}\rangle && (4.12)\\ &= 2a^2\|\vec{x}\|^2 - 2a^2\|\vec{x}\|^2 = 0 && (4.13) \end{aligned}$$

Thus, by positive definiteness of the norm $T(a\vec{x}) = aT(\vec{x})$.

ii. For any $\vec{x}, \vec{y} \in \mathbb{R}^n$:

$$\begin{aligned} \|T(\vec{x}+\vec{y}) - T(\vec{x}) - T(\vec{y})\|^2 &= \|T(\vec{x}+\vec{y})\|^2 + \|T(\vec{x})\|^2 + \|T(\vec{y})\|^2 && (4.14)\\ &\quad - 2\langle T(\vec{x}+\vec{y}), T(\vec{x})\rangle - 2\langle T(\vec{x}+\vec{y}), T(\vec{y})\rangle \\ &\quad + 2\langle T(\vec{x}), T(\vec{y})\rangle \\ &= \|\vec{x}+\vec{y}\|^2 + \|\vec{x}\|^2 + \|\vec{y}\|^2 && (4.15)\\ &\quad - 2\langle \vec{x}+\vec{y}, \vec{x}\rangle - 2\langle \vec{x}+\vec{y}, \vec{y}\rangle + 2\langle \vec{x}, \vec{y}\rangle \\ &= \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\langle \vec{x}, \vec{y}\rangle - \|\vec{x}+\vec{y}\|^2 = 0 && (4.16) \end{aligned}$$

Thus, by positive definiteness of the norm $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$.

Thus T is a linear *rigid motion*, so T is orthogonal. It remains only to be shown that $f = g \circ T$ is unique. Suppose $f = g' \circ T'$ for a translation g' and orthogonal transformation T', then:

$$\begin{aligned} f(\vec{0}) &= (g' \circ T')\vec{0} = g'(\vec{0}) && (4.17)\\ &= (g \circ T)\vec{0} = g(\vec{0}) && (4.18) \end{aligned}$$

Thus $g(\vec{x}) = g'(\vec{x}) = \vec{x} + f(\vec{0})$. But then

$$T'\vec{x} = \left(g^{-1} \circ f\right)\vec{x} = f(\vec{x}) - f(\vec{0}) = T\vec{x}$$

so $g' = g$ and $T' = T$ as required for uniqueness. $\square$

Week 5

October 6 - October 18

5.1 Quadratic Forms

Definition 5.1.1 (Forms)**.** A *form* or *homogenous polynomial* is a polynomial where every term has the same degree

Definition 5.1.2 (Quadratic Form)**.** A *quadratic form* is a *homogeneous polynomial* of degree 2

Remark. A *quadratic form* in 2 variables over $\mathbb{F}$ is a polynomial of the form:

$$ax^2 + bxy + cy^2 \qquad \text{for some } a, b, c \in \mathbb{F} \tag{5.1}$$

Remark. Any *quadratic form* over $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}\}$ can be expressed as a symmetric matrix:

$$\begin{pmatrix} x & y \end{pmatrix} \begin{bmatrix} a & \frac{b}{2} \\ \frac{b}{2} & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \qquad \text{for some } a, b, c \in \mathbb{F} \tag{5.2}$$

Indeed, expanding this expression we see:

$$\begin{aligned} \begin{pmatrix} x & y \end{pmatrix} \begin{bmatrix} a & \frac{b}{2} \\ \frac{b}{2} & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} ax + \frac{b}{2}y & \frac{b}{2}x + cy \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} && (5.3) \\ &= ax^2 + \frac{b}{2}yx + \frac{b}{2}xy + cy^2 && (5.4) \\ &= ax^2 + bxy + cy^2 && (5.5) \end{aligned}$$

Definition 5.1.3 (Plane Curve)**.** A *plane curve* for a *quadratic* form is a set in $\mathbb{R}^2$ given by

$$ax^2 + bxy + cy^2 = d \qquad \text{for some } a, b, c, d \in \mathbb{R} \tag{5.6}$$

Remark (Author's Note)**.** In class we defined a quadratic form to be exactly that set, but really a quadratic form refers to the polynomial itself, the set of solutions is a plane curve.

Remark. Consider the plane curve for *quadratic form* over $\mathbb{R}$ with an associated symmetric matrix A, and let P be an orthogonal change of basis matrix. Then under our new basis

$$\begin{align}
\begin{pmatrix} x \\ y \end{pmatrix} \mapsto P \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} x & y \end{pmatrix} \mapsto \quad & \left[P \begin{pmatrix} x \\ y \end{pmatrix} \right]^{\intercal} \tag{5.7} \\
&= \begin{pmatrix} x & y \end{pmatrix} P^{\intercal} \tag{5.8} \\
&= \begin{pmatrix} x & y \end{pmatrix} P^{*} \qquad \text{since } P \in \mathbb{R}^{2\times 2} \tag{5.9} \\
&= \begin{pmatrix} x & y \end{pmatrix} P^{-1} \qquad \text{since } PP^* = I = P^*P \tag{5.10}
\end{align}$$

so, under our new basis we see

$$\begin{align}
\begin{pmatrix} x & y \end{pmatrix} A \begin{pmatrix} x \\ y \end{pmatrix} &\mapsto \begin{pmatrix} x & y \end{pmatrix} P^{-1}AP \begin{pmatrix} x \\ y \end{pmatrix} \tag{5.11} \\
\left(P^{-1}AP\right)^{\intercal} &= P^{\intercal}A^{\intercal}\left(P^{-1}\right)^{\intercal} \tag{5.12} \\
&= P^{-1}A(P^{\intercal})^{\intercal} \tag{5.13} \\
&= P^{-1}AP \tag{5.14}
\end{align}$$

that is P changes our symmetric matrix A into a similar matrix.

Since A is symmetric and is a matrix over $\mathbb{R}$, it is self-adjoint, so we can choose an orthonormal basis B of $\mathbb{R}^2$ so that $[A]_B$ is diagonal.

Then if P represents the change of coordinates from B to the standard basis, $P^{-1}AP$ is diagonal, so the plane curve for its associated quadratic form is of the form

$$Ax^2 + Cy^2 = D \qquad \text{for some } A, C, D \in \mathbb{R} \tag{5.15}$$

and these are readily understood to be be:

- an **ellipse** if A, C, D all have the same sign
- a **hyperbola** if A, C have different signs
- **degenerate** if $ACD = 0$ or if A, C have the same sign and D the opposite

Furthermore, we can choose P to be a rotation $P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$ where $\vec{v}_1$, $\vec{v}_w$ are the eigenvectors of A. If P is not a rotation, then the determinant is negative, so consider the matrix $P' = \begin{bmatrix} \vec{v}_1 & -\vec{v}_2 \end{bmatrix}$ which also orthogonally diagonalizes A where $\det P' = -\det P > 0$, making P' a rotation.

Example 5.1.3.1. Consider the plane curve for a quadratic form given by:

$$3x^2 + 2xy - y^2 = 14$$

then the quadratic form has an associated symmetric matrix A given by:

$$A = \begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix} \tag{5.16}$$

to diagonalize, we determine the eigenvalues of A

$$\det(A - \lambda I) = \det \begin{vmatrix} 3-\lambda & 1 \\ 1 & -1-\lambda \end{vmatrix} \tag{5.17}$$

$$= \lambda^2 - 2\lambda - 4 \tag{5.18}$$

$$\lambda = \frac{2 \pm \sqrt{4+16}}{2} = 1 \pm \sqrt{5} \tag{5.19}$$

and using these eigenvalues we determine the eigenvectors of A

$$\left(A - (1+\sqrt{5})I\right)\vec{v}_1 = 0 \qquad \left(A - (1-\sqrt{5})I\right)\vec{v}_2 = 0 \tag{5.20}$$

$$\begin{bmatrix} 2-\sqrt{5} & 1 \\ 1 & -2-\sqrt{5} \end{bmatrix}\vec{v}_1 = 0 \qquad \begin{bmatrix} 2+\sqrt{5} & 1 \\ 1 & -2+\sqrt{5} \end{bmatrix}\vec{v}_1 = 0 \tag{5.21}$$

row reduction (or just inspection in this case) yields eigenvectors

$$\vec{v}_1 = \begin{pmatrix} 1 \\ -2+\sqrt{5} \end{pmatrix} \qquad \vec{v}_2 = \begin{pmatrix} 1 \\ -2-\sqrt{5} \end{pmatrix} \tag{5.22}$$

We gave up in class some time around here because it turns out normalizing these vectors is gross. But pretty much you normalize, take $P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$, then $P^{-1}AP$ becomes diagonal where $P^{-1} = P^{\intercal}$.

5.2 Projections

Definition 5.2.1 (Projection)**.** A *projection* (not to be confused with an *orthogonal projection*) is any linear transformation $T : V \to V$ satisfying

$$T = T^2$$

We call T the projection of V onto $\operatorname{im} T$ along $\ker T$.

Example 5.2.1.1 (Non-Orthogonal Projection)**.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(x, y) = (x - y, 0)$. Then T is not an *orthogonal projection*, but it is a *projection*. Indeed:

$$T^2(x, y) = T(x - y, 0) = (x - y, 0) = T(x, y)$$

Remark (Restriction of T onto $\operatorname{im} T$)**.** If $T : V \to V$ is a projection then $T\big|_{\operatorname{im} T} = id\big|_{\operatorname{im} T}$. Indeed for any $\vec{v} \in \operatorname{im} T$, we must have $\vec{v} = T\vec{w}$ for some $\vec{w} \in V$ and

$$T\vec{v} = T(T\vec{w}) = T^2\vec{w} = T\vec{w} = \vec{v}$$

Theorem 5.2.2 (Orthogonal Projections are Projections)**.** *Let W be a subspace of V, then $proj_W(\cdot) : V \to V$ is a* projection *(under the natural injection into V, technically $proj_W(\cdot) : V \to W$).*

Proof. Recall that we have $W \oplus W^{\perp} \simeq V$ under a bijective map $(w, w') \mapsto w + w'$. Thus for any $\vec{v} \in W$ consider the corresponding $(\vec{w}, \vec{w}') \in W \oplus W^{\perp}$

$$\begin{aligned} \mathrm{proj}_W\left(\mathrm{proj}_W\left(\vec{v}\right)\right) &= \mathrm{proj}_W\left(\mathrm{proj}_W\left(\vec{w} + \vec{w}'\right)\right) && (5.23)\\ &= \mathrm{proj}_W\left(\vec{w}\right) && (5.24)\\ &= \mathrm{proj}_W\left(\vec{w} + \vec{w}'\right) && (5.25)\\ &= \mathrm{proj}_W\left(\vec{v}\right) && (5.26) \end{aligned}$$

so $\mathrm{proj}_W\left(\cdot\right)^2 = \mathrm{proj}_W\left(\cdot\right)$ □

Theorem 5.2.3. *A* projection $T : V \to V$ *is an* orthogonal projection *if and only if* $(\operatorname{im} T)^{\perp} = \ker T$

Proof. $\implies$ Suppose T is an orthogonal projection $T = \mathrm{proj}_W\left(\cdot\right) : V \to V$. Then $\ker T = W^{\perp}$ and $\operatorname{im} T = W$ by definition, so $\ker T = (\operatorname{im} T)^{\perp}$

$\impliedby$ Let $W = \operatorname{im} T$, then $\ker T = W^{\perp}$. Then T and $\mathrm{proj}_W\left(\cdot\right)$ agree on both W and $W^{\perp}$, so $T = \mathrm{proj}_W\left(\cdot\right)$.

□

Theorem 5.2.4. *A* projection $T : V \to V$ *is an* orthogonal projection *if and only if* $T = T^*$

Proof. $\implies$ Let $\{\vec{v}_1, \ldots, \vec{v}_r\}$ be an orthonormal basis of $\operatorname{im} T$, then we can extend it to a basis $B = \{\vec{v}_1, \ldots, \vec{v}_n\}$ of V. Then:

$$T\vec{v}_i = \begin{cases} \vec{v}_i & 1 \leq i \leq r \\ 0 & r < i \leq n \end{cases} \tag{5.27}$$

which gives us a matrix for $[T]_B$ of the form:

$$[T]_B = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{5.28}$$

thus $[T]_B^* = [T]_B$, so since B is orthonormal, $T = T^*$

$\impliedby$ Suppose $T = T^*$ and T is a projection.

Then T is self-adjoint so there exists some basis B such that $[T]_B$ is diagonal. The diagonal entries of $[T]_B$ are the eigenvalues of T.

So let λ be an eigenvalue of T and $\vec{v}$ be a $\lambda-$eigenvector. Since T is a projection:

$$\lambda\vec{v} = T\vec{v} = T^2\vec{v} = \lambda^2\vec{v} \tag{5.29}$$

so $\lambda \in \{0, 1\}$. Since B is an eigenbasis of T, by reordering B we may assume that $B = \{\vec{v}_1, \ldots, \vec{v}_n\}$ such that

$$\vec{v}_1, \ldots, \vec{v}_r \in \operatorname{im} T \qquad \text{since } T\vec{v}_i = \vec{v}_i \tag{5.30}$$

$$\vec{v}_{r+1}, \ldots, \vec{v}_n \in \ker T \qquad \text{since } T\vec{v}_i = 0 \tag{5.31}$$

Thus T and $\operatorname{proj}_{\operatorname{span}\{\vec{v}_1,\ldots,\vec{v}_r\}}(\cdot)$ agree on B, which is a basis for V so

$$T = \operatorname{proj}_{\operatorname{span}\{\vec{v}_1,\ldots,\vec{v}_r\}}(\cdot) \tag{5.32}$$

□

5.3 Spectral Theorem

Theorem 5.3.1 (Spectral Theorem)**.** *Let $T : V \to V$ be a linear transformation with an orthonormal B of V such that $[T]_B$ is diagonal.*

Then let $\lambda_1, \ldots, \lambda_k$ be the eigenvalues of T, defining

$$W_i = \lambda_i\text{-eigenspace of } T$$

Then define $T_i : V \to W_i$ to be the orthogonal projection of V onto W_i, then:

(1) $V \simeq W_1 \oplus \cdots \oplus W_K$ as an inner product space

(2) $T_i \circ T_j = \delta_{i,j} T_i$ where $\delta_{i,j}$ is the Kronecker delta:

$$\delta_{i,j} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

(3) $T_1 + \cdots + T_k = I$

(4) $T = \lambda_1 T_1 + \cdots \lambda_k T_k$

Proof. We will prove these properties in a coordinate free style, but intuitively they hold because we can reorder B so that the λ_1-eigenvectors come first, followed by λ_2-eigenvectors, and so on, giving:

$$[T]_B = \begin{bmatrix} \lambda_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \lambda_1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \lambda_k \end{bmatrix} \tag{5.33}$$

Also observe before proceeding that for any W_i, W_j with $i \neq j$ we have $W_i \subset W_j^\perp$ since W_i and W_j draw their bases from disjoint subsets of an orthonormal basis.

(1) Define $\phi : W_1 \oplus \cdots \oplus W_k \to V$ by

$$\phi(\vec{w}_1, \ldots, \vec{w}_k) = \vec{w}_1 + \cdots + \vec{w}_k \tag{5.34}$$

Then we have $\ker \phi = (0)$ since eigenvectors with different eigenvalues are linearly independent, and $\operatorname{im} \phi = V$ because V admits a basis of eigenvectors of T. Since ϕ is injective, surjective, and linear, ϕ is an isomorphism.

To check that ϕ respects inner products, consider for arbitrary points $(\vec{w}_1, \ldots, \vec{w}_k)$ and $(\vec{w}'_1, \ldots, \vec{w}'_k) \in W_1 \oplus \cdots \oplus W_k$

$$\begin{aligned}
\langle (\vec{w}_1, \ldots, \vec{w}_k), (\vec{w}'_1, \ldots, \vec{w}'_k) \rangle &= \langle \vec{w}_1, \vec{w}'_1 \rangle + \cdots + \langle \vec{w}_k, \vec{w}'_k \rangle && (5.35)\\
&= \sum_{i=1}^{k} \langle \vec{w}_i, \vec{w}'_i \rangle && (5.36)\\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} \left\langle \vec{w}_i, \vec{w}'_j \right\rangle \text{ since } \langle \vec{w}_i, \vec{w}_j \rangle = 0 \text{ if } i \neq j && (5.37)\\
&= \langle \vec{w}_1 + \cdots + \vec{w}_k, \vec{w}'_1 + \cdots + \vec{w}'_k \rangle && (5.38)\\
&= \langle \phi(\vec{w}_1, \ldots, \vec{w}_k), \phi(\vec{w}'_1, \ldots, \vec{w}'_k) \rangle && (5.39)
\end{aligned}$$

(2) If $i \neq j$ then $W_i \subset W_j^\perp = \ker T_j$ and $W_i = \operatorname{im} T_i$, so $T_j \circ T_i = 0$

Otherwise, if $i = j$ then $T_j \circ T_i = T_i^2 = T$ by definition of a projection

(3) Let $B = \{\vec{v}_1, \ldots, \vec{v}_n\}$. Then for any $\vec{v} \in V$ we have:

$$\vec{v} = a_1 \vec{v}_1 + \cdots + a_n \vec{v}_n \tag{5.40}$$

$$T_i \vec{v} = \sum_{j : \vec{v}_j \in W_j} a_j \vec{v}_j \tag{5.41}$$

$$\sum_{i=1}^{k} T_i \vec{v} = a_1 \vec{v}_1 + \cdots + a_n \vec{v}_n = \vec{v} \tag{5.42}$$

(4) For any λ_i-eigenvector, $\vec{v}_j \in B$, since $\vec{v}_j = T_i \vec{v}_j$ and $T_\ell \vec{v}_j = (T_\ell \circ T_i) \vec{v}_j = \delta_{\ell,j} \vec{v}_j$

$$\begin{aligned}
T \vec{v}_j &= \lambda_i \vec{v}_j && (5.43)\\
&= \lambda_i T_i \vec{v}_j && (5.44)\\
&= \lambda_1 T_1 \vec{v}_j + \cdots + \lambda_k T_k \vec{v}_j && (5.45)
\end{aligned}$$

So T and $\lambda_1 T_1 + \cdots \lambda_k T_k$ agree on the basis B, and are thus equal

$\square$

Remark (Motivation for the Spectral Theorem)**.** Consider the vector space $V = C^\infty(0,1) \subset \{f : [0,1] \to \mathbb{R}\}$ of infinitely differentiable functions such that $f^{(n)}(0) = f^{(n)}(1)$ for every derivative. We will wave our hands here and ignore the fact that V is not finite-dimensional and we haven't proven our theorems in infinite dimensions.

Consider the L_2 inner product $\langle f, g\rangle = \int_0^1 fg$ and recall that $T : V \to V$ given by $Tf = f'$ is well defined since differentiation is a linear operator. Then for any $f, g \in V$:

$$\langle Tf, g\rangle = \int_0^1 f'g = fg\big|_0^1 - \int_0^1 fg' = \int_0^1 fg' = \langle g, -Tg\rangle$$

Thus $T^* = -T$, so $(T^2)^* = (T^*)^2 = T^2$ meaning $T^2 = \{f \mapsto f''\}$ is self-adjoint. Thus T is orthogonally diagonalizable by an eigenbasis. In particular, the eigenvectors of T^2 are of the form:

$$f(x) = \sin(\lambda x) \text{ for some } \lambda \in \mathbb{R}$$

Decomposing a function into a combination of sin functions is called *spectral analysis.*

Lemma 5.3.2 (Determinant of Vandermonde Matrix)**.** *Let A be a Vandermonde matrix, that is A is of the form:*

$$A = \begin{bmatrix} 1 & \alpha_1 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \dots & \alpha_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \dots & \alpha_m^{n-1} \end{bmatrix}$$

then the determinant of A is given by:

$$\det A = \pm \prod_{i<j} (\alpha_i - \alpha_j)$$

No proof given in this course because it's neither easy nor particularly relevant. Not too hard to find one online if you care.

Theorem 5.3.3 (Lagrange Interpolation)**.** *Let $a_1, \dots, a_n$ be distinct elements of a field $\mathbb{F}$, and let $b_1, \dots, b_n$ be any elements of $\mathbb{F}$ (not necessarily distinct).*

Then there is a polynomial $g(x)$ of degree at most $n-1$ satisfying $g(a_i) = b_i$ for all $1 \le i \le n$

Proof. Let V be the vector space of polynomials in x with coefficients in $\mathbb{F}$ of degree at most $n-1$.

Let $p(x) = c_0 + c_1 x + \dots + c_n x^{n-1}$ be a general element of V, then $p(a_i) = b_i$ if and only if $c_0 + c_1 a_i + \dots + c_{n-1} a_i^{n-1} = b_i$, which is a linear equation in c_i.

Thus the system $\{g(a_i) = b_i\}$ is a linear system in the coefficients c_i:

$$\begin{cases} c_0 & + & c_1 a_1 & + & \cdots & + & c_{n-1} a_1^{n-1} & = & b_1 \\ c_0 & + & c_1 a_2 & + & \cdots & + & c_{n-1} a_2^{n-1} & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ c_0 & + & c_1 a_n & + & \cdots & + & c_{n-1} a_n^{n-1} & = & b_n \end{cases}$$

Note that the corresponding matrix A given by

$$A = \begin{bmatrix} 1 & a_1 & \cdots & a_1^{n-1} \\ 1 & a_2 & \cdots & a_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_n & \cdots & a_n^{n-1} \end{bmatrix}$$

is a Vandermonde matrix, so $\det A = \pm \prod_{i<j}(a_i - a_j)$, but a_i are all distinct.

Thus $\det A \neq 0$ so the system is consistent and has a solution $(c_0, \ldots, c_{n-1})$ with corresponding polynomial p satisfies $p(a_i) = b_i$ for all $1 \leq i \leq n$. $\square$

Theorem 5.3.4. *Let $T : V \to V$ be a complex linear transformation. Then T is normal if and only if $T^* = g(T)$ for some polynomial $g(x) \in \mathbb{C}[x]$*

Proof. $\Longrightarrow$ Assume $T^*T = TT^*$, then T is normal in a complex inner product space so T is orthogonally diagonalizable. So we can write the spectral decomposition of T:

$$T = \lambda_1 T_1 + \cdots + \lambda_k T_k \tag{5.46}$$

$$T^* = \overline{\lambda_1} T_1 + \cdots + \overline{\lambda_k} T_k \tag{5.47}$$

By Lagrange Interpolation there is a polynomial $g \in \mathbb{C}[x]$ satisfying $g(\lambda_i) = \overline{\lambda_i}$ for all i. Then

$$g(T) = g(\lambda_1 T_1 + \cdots + \lambda_k T_k) \tag{5.48}$$

$$= g(\lambda_1)T_1 + \cdots + g(\lambda_k)T_k \text{ since } T_iT_j = \delta_{i,j} \text{ and } T_i^n = T_i \tag{5.49}$$

$$= \overline{\lambda_1} T_1 + \cdots + \overline{\lambda_k} T_k = T^* \tag{5.50}$$

$\Longleftarrow$ This is immediate from the fact that T commutes with $g(T)$ for any $g \in \mathbb{C}[x]$. Indeed, for each term a_iT^i we have

$$a_iT^iT = a_iT^{i+1} = aTT^i = T(aT^i)$$

$\square$

Theorem 5.3.5. *Let $T : V \to V$ be a complex linear transformation. Then T is unitary if and only if T is normal and all the eigenvalues of T have length 1*

Proof. Recall that we have already shown earlier that if T is unitary it is normal with eigenvalues all having length 1.

Suppose then that T is normal and all eigenvalues have length 1. Since it is normal in a complex inner product space it is orthogonally diagonalizable. SO we can write the spectral decomposition of T:

$$T = \lambda_1 T_1 + \cdots + \lambda_k T_k \tag{5.51}$$

$$T^* = \overline{\lambda_1} T_1 + \cdots + \overline{\lambda_k} T_k \tag{5.52}$$

$$= \lambda_1^{-1} T_1 + \cdots + \lambda_k^{-1} T_k \tag{5.53}$$

$$= T^{-1} \tag{5.54}$$

Thus $TT^* = I = T^*T$ so T is unitary $\square$

Theorem 5.3.6. *Let T be a normal linear transformation. Then $T = T^*$ if and only if every root of the characteristic polynomial of T is real.*

Proof. Recall that we have already shown that if T is self-adjoint then every root of the characteristic polynomial of T is real.

Suppose then that every root of the characteristic polynomial of T is real. Then T is orthogonally diagonalizable so we can write its spectral decomposition:

$$T = \lambda_1 T_1 + \cdots + \lambda_k T_k \tag{5.55}$$

$$T^* = \overline{\lambda_1} T_1 + \cdots + \overline{\lambda_k} T_k \tag{5.56}$$

$$= \lambda_1 T_1 + \cdots + \lambda_k T_k = T \tag{5.57}$$

□

Remark. Let T be an orthogonally diagonalizable transformation with spectral decomposition

$$T = \lambda_1 T_1 + \cdots + \lambda_k T_k \tag{5.58}$$

then T_i is a polynomial in T. Indeed, choose $g_i(x)$ such that $g_i(\lambda_i) = \delta_{i,j}$, then

$$g_i(T) = g_i(\lambda_1 T_1 + \cdots + \lambda_k T_k) \tag{5.59}$$

$$= g_i(\lambda_1) T_1 + \cdots + g_i(\lambda_k) T_k \tag{5.60}$$

$$= T_i \tag{5.61}$$

Week 6

October 20 - October 25

6.1 Singular Values

Remark. Let $T : V \to W$ be a linear transformation from inner product spaces V to W. Then let B_V, B_W be orthonormal bases for V and W respectively, then we can define an adjoint $T^* : W \to V$ by

$$[T^*]_{B_W \to B_V} = [T]^*_{B_V \to B_W} \tag{6.1}$$

Lemma 6.1.1. *For any linear transformation $\mathcal{U} : V \to W$, any eigenvalues of $\mathcal{U}^*\mathcal{U}$ are non-negative real numbers.*

Proof. Let $\mathcal{U}^*\mathcal{U}\vec{v} = \lambda\vec{v}$, then:

$$\begin{align}
\langle \mathcal{U}^*\mathcal{U}\vec{v}, \vec{v}\rangle &= \lambda \|\vec{v}\|^2 \tag{6.2}\\
\langle \mathcal{U}^*\mathcal{U}\vec{v}, \vec{v}\rangle &= \langle \mathcal{U}v, \mathcal{U}\vec{v}\rangle \tag{6.3}\\
&= \|\mathcal{U}\vec{v}\|^2 \tag{6.4}
\end{align}$$

However note that $\mathcal{U}\vec{v} = \sigma\vec{v}$ where $|\sigma|^2 = \lambda$, thus

$$\lambda \|\vec{v}\| = \|\mathcal{U}\vec{v}\|^2 = |\sigma|^2 \|\vec{v}\|^2 = |\lambda| \|\vec{v}\| \tag{6.5}$$

Since this holds for any eigenvector $\vec{v}$, we must have $\lambda = |\lambda|$ so $\lambda \in \mathbb{R}_{\geq 0}$. □

Theorem 6.1.2. *Let $T : V \to W$ be a linear transformation from inner product spaces V to W. Then there exist orthonormal bases*

$$\begin{align}
B_V &= \{\vec{v}_1, \ldots, \vec{v}_n\} && \textit{for } V \tag{6.6}\\
B_W &= \{\vec{w}_1, \ldots, \vec{w}_m\} && \textit{for } W \tag{6.7}
\end{align}$$

and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$, $\sigma_i \in \mathbb{R}_{\geq 0}$ for $k = \min\{n, m\}$ such that

$$T(\vec{v}_i) = \begin{cases} \sigma_i \vec{w}_i & 1 \leq i \leq k \\ \vec{0} & i > k \end{cases} \tag{6.8}$$

Moreover, $\vec{v}_i$ is an eigenvector of T^*T with eigenvalues σ_i^2 (or 0 if $i > k$)

Proof. Note that $T^*T : V \to V$ and $(T^*T)^* = T^*T$. So T^*T is self-adjoint so there exists an orthonormal basis $B_V = \{\vec{v}_1, \dots, \vec{v}_n\}$ of eigenvectors. By 6.1.1 we know that the eigenvalues must be positive real numbers. So let us order them

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0 \tag{6.9}$$

where λ_i is the eigenvalue for $\vec{v}_i$. Then define $\sigma_i = \sqrt{\lambda_i}$ and define

$$\vec{w}_i = \frac{1}{\sigma_i} T\vec{v} \qquad \text{if } \sigma_i \neq 0 \tag{6.10}$$

then extend $\{\vec{w}_1, \dots, \vec{w}_r\}$ to an orthonormal basis

$$B_W = \{\vec{w}_1, \dots, \vec{w}_m\} \qquad \text{for } W \tag{6.11}$$

where r happens to be the rank of T. We can do this because $\{\vec{w}_1, \dots, \vec{w}_r\}$ is orthonormal. Indeed

$$\begin{aligned}
\langle \vec{w}_i, \vec{w}_j \rangle &= \left\langle \frac{1}{\sigma_i} T\vec{v}_i, \frac{1}{\sigma_j} T\vec{v}_j \right\rangle & (6.12)\\
&= \frac{1}{\sigma_i \sigma_j} \langle T\vec{v}_i, T\vec{v}_j \rangle & (6.13)\\
&= \frac{1}{\sigma_i \sigma_j} \langle \vec{v}_i, T^*T\vec{v}_i\vec{v}_j \rangle & (6.14)\\
&= \frac{\sigma_j^2}{\sigma_i \sigma_j} \langle \vec{v}_i, \vec{v}_j \rangle & (6.15)\\
&= \delta_{i,j} & (6.16)
\end{aligned}$$

□

Definition 6.1.3 (Singular Values)**.** We call these $\sigma_1, \dots, \sigma_k$ the *singular values* of T.

Remark (Uniqueness of Singular Values)**.** If $\{\vec{v}_1, \dots, \vec{v}_n\}$ and $\{\vec{w}_1, \dots, \vec{w}_m\}$ are orthonormal bases of V and W satisfying

$$T(\vec{v}_i) = \begin{cases} \sigma_i \vec{w}_i & 1 \leq i \leq k \\ 0 & i > k \end{cases} \tag{6.17}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0$, then each $\vec{v}_i$ must necessarily be an eigenvalue of T^*T with eigenvalue σ_i^2.

Proof. For all i, j we get

$$\begin{aligned}
\langle T^*\vec{w}_i, \vec{v}_j \rangle &= \langle \vec{w}_i, T\vec{v}_j \rangle & (6.18)\\
&= \begin{cases} \sigma_i & 1 \leq i \leq k, i = j \\ 0 & \textit{otherwise} \end{cases} & (6.19)
\end{aligned}$$

So if $i \leq k$ we get

$$\begin{aligned} T^*T\vec{v}_i &= T^*(\sigma_i\vec{v}_i) && (6.20)\\ &= \sigma_i T^*\vec{w}_i && (6.21)\\ &= \sigma_i^2\vec{v}_i && (6.22)\end{aligned}$$

And similarly if $k < i$ we must have $T^*T\vec{v}_i = 0$ □

Remark (Singular Values of Adjoint)**.** By construction, the *singular values* of T are the same as the *singular values* of T^*

6.2 Singular Value Decomposition

Definition 6.2.1 (Singular Value Decomposition)**.** Given $T : V \to W$, let $B_V = \{\vec{v}_1, \ldots, \vec{v}_n\}$ and $B_W = \{\vec{w}_1, \ldots, \vec{w}_m\}$ be orthonormal bases for V and W respectively such that

$$T(\vec{v}_i) = \begin{cases} \sigma_i\vec{w}_i & 1 \leq i \leq k \\ 0 & k < i \end{cases} \tag{6.23}$$

where $k = \min(m, n)$ and σ_i^2 are the eigenvalues of T^*T. Then we have

$$\Sigma = [T]_{B_V \to B_W} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{bmatrix} \tag{6.24}$$

where Σ is not necessarily square, but all other entries are 0. Then let V be the change of basis matrix from B_V to std, and similarly let U be the change of basis from B_W to std. If A is a matrix representing such a linear transformation in $\mathbb{F}^n \to \mathbb{F}^m$ (for $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$), we can write

$$A = U\Sigma V^* \tag{6.25}$$

This is called the *singular value decomposition* of A

Remark. Recall that B_V can be determined by the eigenvectors of A^*A so V is the matrix whose columns are the vectors in B_V, and B_W is the extension of $A(B_V)$ to an orthonormal basis of W, so similarly U is the matrix whose columns are the vectors in B_W.

Example 6.2.1.1. We will determine the *singular value decomposition* of A where

$$A = \begin{bmatrix} \frac{19}{3} & \frac{-14}{3} & \frac{-10}{3} \\ \frac{-8}{3} & \frac{-2}{3} & \frac{-20}{3} \end{bmatrix} \tag{6.26}$$

We begin by determining A^*A

$$A^*A = \begin{bmatrix} \frac{19}{3} & \frac{-8}{3} \\ \frac{-14}{3} & \frac{-2}{3} \\ \frac{-10}{3} & \frac{-20}{3} \end{bmatrix} \begin{bmatrix} \frac{19}{3} & \frac{-14}{3} & \frac{-10}{3} \\ \frac{-8}{3} & \frac{-2}{3} & \frac{-20}{3} \end{bmatrix} \tag{6.27}$$

$$= \begin{bmatrix} \frac{425}{9} & \frac{-250}{9} & \frac{-350}{9} \\ \frac{-250}{9} & \frac{200}{9} & \frac{100}{9} \\ \frac{-230}{9} & \frac{100}{9} & \frac{500}{9} \end{bmatrix} \tag{6.28}$$

$$= \frac{25}{9} \begin{bmatrix} 17 & -10 & -14 \\ -10 & 8 & 4 \\ -14 & 4 & 20 \end{bmatrix} \tag{6.29}$$

We then determine the eigenvalues by the characteristic polynomial (details omitted)

$$\det(A^*A - \lambda I) = -\lambda^3 + 45\lambda^2 - 324\lambda \tag{6.30}$$

$$= -\lambda(\lambda^2 - 45\lambda + 324) \tag{6.31}$$

$$= -\lambda(\lambda - 4)(\lambda - 36) \tag{6.32}$$

Thus the eigenvalues of A^*A are 0, 25, and 100, giving us singular values

$$\sigma_1 = \sqrt{100} = 10 \qquad \sigma_2 = \sqrt{25} = 5 \tag{6.33}$$

Solving $(A^*A - \lambda I)\vec{v} = \vec{0}$ by echelon row reduction (omitted) gives us

$$\text{100-eigenvalue of } \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix} \qquad \text{25-eigenvalue of } \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix} \qquad \text{0-eigenvalue of } \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} \tag{6.34}$$

so normalizing yields an orthonormal basis

$$B_V = \left\{ \begin{pmatrix} \frac{-2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}, \begin{pmatrix} \frac{1}{3} \\ \frac{-2}{3} \\ \frac{2}{3} \end{pmatrix}, \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{pmatrix} \right\} \qquad V = \begin{bmatrix} \frac{-2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{-2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \tag{6.35}$$

We compute B_W by taking $\frac{1}{\sigma_i} A\vec{v}_i$ for our non-zero eigenvectors (omitted) giving

$$B_W = \left\{ \begin{pmatrix} \frac{-4}{5} \\ \frac{3}{5} \end{pmatrix}, \begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \end{pmatrix} \right\} \qquad U = \begin{bmatrix} \frac{-4}{3} & \frac{3}{3} \\ \frac{3}{3} & \frac{4}{3} \end{bmatrix} \tag{6.36}$$

$$A = U\Sigma V^* = \begin{bmatrix} \frac{-4}{3} & \frac{3}{3} \\ \frac{3}{3} & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 10 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix} \begin{bmatrix} \frac{-2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{-2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \tag{6.37}$$

Example 6.2.1.2 (Singular values are not the same as eigenvalues)**.** Consider the non-diagonalizable matrix A given by

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{6.38}$$

Then 1 is the only eigenvalue of A, however the singular values are $\sqrt{\dfrac{3 \pm \sqrt{5}}{2}}$ giving

$$A = U \begin{bmatrix} \sqrt{\dfrac{3+\sqrt{5}}{2}} & 0 \\ 0 & \sqrt{\dfrac{3-\sqrt{5}}{2}} \end{bmatrix} V^* \tag{6.39}$$

for some unitary U and V. Thus if $\|\vec{v}\| = 1$ we have $\|A\vec{v}\| \leq \sqrt{\dfrac{3+\sqrt{5}}{2}}$ giving us a bound

$$\frac{\|A\vec{v}\|}{\|\vec{v}\|} \leq \sqrt{\frac{3+\sqrt{5}}{2}} \tag{6.40}$$

6.3 Pseudo-Inverses

Definition 6.3.1 (Pseudo-Inverse of a Linear Transformation)**.** Let $T : V \to W$ be any (not necessarily invertible) linear transformation where V and W are finite dimensional. Then we say $T^\dagger : W \to V$ is the pseudo-inverse of T given by the linear transformation

$$T^\dagger \vec{v} = \begin{cases} \vec{0} & \vec{v} \in (\operatorname{im} T)^\perp \\ \left(T\big|_{(\ker T)^\perp}\right)^{-1} \vec{v} & \vec{v} \in \operatorname{im} T \end{cases} \tag{6.41}$$

Example 6.3.1.1 (Regular Inverse)**.** If T is invertible then $\operatorname{im} T = W$ and $\ker T = (0)$ so $T^\dagger = T^{-1}$

Example 6.3.1.2 (Pseudo-Inverse of 0)**.** If $T = 0$ then $\operatorname{im} T = (0)$ so $T^\dagger = 0$

Remark (Pseudo-Inverse by Singular Value Decomposition)**.** Let A be a matrix and $A = U\Sigma V^*$ be the *singular value decomposition* of A, then

$$A^\dagger = V\Sigma^\dagger U^* \tag{6.42}$$

where $\Sigma^\dagger$ is the transpose of Σ with all non-zero entries replaced with their reciprocals.

Example 6.3.1.3. Let A be as defined in 6.2.1.1. Then

$$A^\dagger = \begin{bmatrix} \frac{-2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{-2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{10} & 0 \\ 0 & \frac{1}{5} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{-4}{3} & \frac{3}{3} \\ \frac{3}{3} & \frac{4}{3} \end{bmatrix} \tag{6.43}$$

Also note that $AA^\dagger = I$ but $A^\dagger A \neq I$.

Week 7

October 27 - November 1

7.1 Pseudo-Inverse (continued)

Theorem 7.1.1. *Let $T : V \to W$ be a linear transformation. Then $T^\dagger T$ is the orthogonal projection onto $(\ker T)^\perp$ and $TT^\dagger$ is the orthogonal projection onto $\operatorname{im} T$.*

Proof. If $\vec{v} \in (\ker T)^\perp$ then $T^\dagger T\vec{v} = \vec{v}$ by definition of $T^\dagger$, and if $\vec{v} \in \left((\ker T)^\perp\right)^\perp = \ker T$, then $T^\dagger T\vec{v} = \vec{0}$.

If $\vec{v} \in \operatorname{im} T$, then $T^\dagger v$ is the $\vec{x} \in (\ker T)^\perp$ with $T\vec{x} = \vec{v}$ so $TT^\dagger\vec{v} = T\vec{x} = \vec{v}$. If $\vec{v} \in (\operatorname{im} T)^\perp$, then $\vec{v} \in \ker T^\dagger$ so $TT^\dagger\vec{v} = \vec{0}$. □

Remark (Author's Remark)**.** Recall from Homework 5 that if V and W are finite-dimensional inner product spaces and $T_1 : V \to W$ and $T_2 : W \to V$ are linear transformations, then if $T_1T_2T_1 = T_1$ and $T_2T_1T_2 = T_2$ with T_1T_2 and T_2T_1 self-adjoint, then $T_2 = T_1^\dagger$.

Theorem 7.1.2. *Let A be a $n \times m$ matrix and $b \in \mathbb{R}^n$, then $\vec{z} = A^\dagger\vec{b}$ is the best solution to $A\vec{x} = \vec{b}$ in the following senses:*

(a) If $A\vec{x} = \vec{b}$ has a solution, then $\vec{z}$ is the solution with the smallest length:

$$A\vec{y} = \vec{b} \implies \|\vec{y}\| \geq \|\vec{z}\| \text{ with equality if and only if } \vec{y} = \vec{z} \tag{7.1}$$

(b) If $A\vec{x} = \vec{b}$ has no solution, then for all vectors $\vec{y}$,

$$\left\|A\vec{y} - \vec{b}\right\| \geq \left\|A\vec{z} - \vec{b}\right\| \text{ with equality if and only if } A\vec{y} = A\vec{z} \tag{7.2}$$

Moreover, by part (a), $A\vec{y} = A\vec{z} \implies \|\vec{y}\| \geq \|\vec{z}\|$ with equality if and only if $\vec{y} = \vec{z}$

Proof. (a) $A\vec{z} = AA^\dagger\vec{b} = \vec{b}$ since $\vec{b} \in \operatorname{im} A$ by assumption and $AA^\dagger$ is the orthogonal projection onto $\operatorname{im} A$. If $A\vec{y} = \vec{b}$ as well, then $A^\dagger A\vec{y} = A^\dagger\vec{b} = \vec{z}$, so $\vec{z}$ is the orthogonal projection of $\vec{y}$ onto $(\ker T)^\perp$. So $\|\vec{y}\| \geq \|\vec{z}\|$ with equality if and only if $\vec{y} = \vec{z}$.

(b) $A\vec{z} = AA^\dagger\vec{b}$ is the orthogonal projection of $\vec{b}$ onto $\operatorname{im} A$, so $A\vec{z}$ is the closest vector to $\vec{b}$ that lies in $\operatorname{im} A$.

□

7.2 Bilinear Forms

Definition 7.2.1 (Bilinear Form)**.** Let V be a vector space over a field F, then a *bilinear form* on V is a function $B : V \times V \to F$ satisfying:

(a) $B(A\vec{x} + b\vec{y}, \vec{z}) = aB(\vec{x}, \vec{z}) + bB(\vec{y}, \vec{z})$

(b) $B(\vec{x}, a\vec{y} + b\vec{z}) = aB(\vec{x}, \vec{y}) + bB(\vec{x}, \vec{z})$

Example 7.2.1.1. The standard dot product in $\mathbb{R}^n$ is a *bilinear form*

Example 7.2.1.2. Let $V = \mathbb{R}^2$, $B((a, b), (c, d)) = ac \pm bd$ are *bilinear forms*

Example 7.2.1.3. Let $V = \mathbb{R}^n$ and let $A \in M_n(\mathbb{R})$. Define $B(\vec{v}, \vec{w}) = \vec{v}^\intercal A\vec{w}$. This is a bilinear form by linearity of matrix multiplication.

Theorem 7.2.2. *Let V be a finite dimensional vector space over F and let B be a bilinear form on V. Then for any basis S for V there is a $n \times n$ matrix $A \in M_n(F)$, where $n = \dim V$, such that*

$$B(\vec{v}, \vec{w}) = [\vec{v}]_S{}^\intercal A[\vec{w}]_S \tag{7.3}$$

Proof. Choose any basis S for V and write $S = \{\vec{v}_1, \dots, \vec{v}_n\}$. Then define

$$a_{ij} = B(\vec{v}_i, \vec{v}_j) \qquad\qquad A = (a_{ij}) \tag{7.4}$$

Then for any $\vec{v}, \vec{w} \in S$, consider their coefficients with respect to S

$$\vec{v} = \alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n \qquad\qquad \vec{w} = \beta_1\vec{v}_1 + \cdots + \beta_n\vec{v}_n \tag{7.5}$$

This gives us

$$[\vec{v}]_S{}^\intercal A[\vec{w}]_S = \begin{pmatrix}\alpha_1 & \cdots & \alpha_n\end{pmatrix} A \begin{pmatrix}\beta_1 \\ \vdots \\ \beta_n\end{pmatrix} \tag{7.6}$$

$$= \begin{pmatrix}\sum_{i=1}^n \alpha_i a_{i1} & \cdots & \sum_{i=1}^n \alpha_n a_{in}\end{pmatrix} \begin{pmatrix}\beta_1 \\ \vdots \\ \beta_n\end{pmatrix} \tag{7.7}$$

$$= \sum_{j=1}^n \sum_{i=1}^n \alpha_i\beta_j a_{ij} \tag{7.8}$$

$$= \sum_{j=1}^n \sum_{i=1}^n \alpha_i\beta_j B(\vec{v}_i, \vec{v}_j) \tag{7.9}$$

$$= B\left(\sum_{i=1}^n \alpha_i\vec{v}_i\ ,\ \sum_{i=1}^n \beta_j\vec{v}_j\right) \tag{7.10}$$

$$= B(\vec{v}, \vec{w}) \tag{7.11}$$

$\square$

7.3 Congruence

Suppose for bases S and S' of V that

$$B(\vec{v},\vec{w}) = [\vec{v}]_S{}^{\intercal} A[\vec{w}]_S = [\vec{v}]_{S'}{}^{\intercal} A'[\vec{w}]_{S'} \tag{7.12}$$

Let $P_{S\to S'}$ be the change of basis matrix from S to S', then

$$\begin{aligned} [\vec{v}]_{S'}{}^{\intercal} A'[\vec{w}]_{S'} &= (P_{S\to S'}[\vec{v}]_S)^{\intercal} A' (P_{S\to S'}[\vec{w}]_S) && (7.13)\\ &= [\vec{v}]_S{}^{\intercal} P_{S\to S'}{}^{\intercal} A' P_{S\to S'}[\vec{w}]_S && (7.14)\end{aligned}$$

So $A = P_{S\to S'}{}^{\intercal} A' P_{S\to S'}$.

Definition 7.3.1. We say that matrices A and B are *congruent* if $A = P^{\intercal} B P$ for some invertible P.

Remark. If two matrices represent the same bilinear form in different coordinates, then they are *congruent*.

Conversely, if $A, A' \in M_n(F)$ are *congruent*, then there is a basis S for F^n such that

$$\vec{v}^{\intercal} A\vec{w} = [\vec{v}]_S{}^{\intercal} A'[\vec{w}]_S \qquad \text{for all } \vec{v},\vec{w} \in V \tag{7.15}$$

Definition 7.3.2. A bilinear form is *diagonalizable* if there is some basis S such that the corresponding matrix is diagonal.

Definition 7.3.3. A bilinear form B is *symmetric* if

$$B(\vec{v},\vec{w}) = B(\vec{w},\vec{v}) \qquad \text{for all } \vec{v},\vec{w} \in V \tag{7.16}$$

Note that B is symmetric if and only if its corresponding matrix is symmetric in any basis.

Remark. For the following proofs, we assume that we are working in a vector space over a field F with characteristic at least 2. That is, $2 \neq 0$.

Theorem 7.3.4. *Let B be a bilinear form on a vector space V. Then B is* diagonalizable *if and only if B is* symmetric.

Proof. If B is diagonalizable then for some basis the corresponding matrix is symmetric so B is symmetric. Suppose that B is symmetric, then we will show that B is also diagonalizable. We proceed by induction on $\dim V$.

Suppose that $\dim V = 1$, then B is diagonalizable since any 1×1 matrix is diagonalizable.

Suppose the claim holds for any vector space V' with $\dim V' < \dim V$. If $B = 0$ then B is diagonalizable. Otherwise, there must exist some $\vec{x},\vec{y} \in V$ with $B(\vec{x},\vec{y}) \neq 0$.

If $B(\vec{x},\vec{x}) = B(\vec{y},\vec{y}) = 0$ then we have

$$B(\vec{x}+\vec{y},\vec{x}+\vec{y}) = B(\vec{x},\vec{x}) + 2B(\vec{x},\vec{y}) + B(\vec{y},\vec{y}) = 2B(\vec{x},\vec{y}) \neq 0 \tag{7.17}$$

So there must exist some vector $\vec{v} \in V$ with $B(\vec{v}, \vec{v}) \neq 0$.

Define $L : V \to F$ by $L(\vec{y}) = B(\vec{v}, \vec{y})$. This is a linear transformation of rank 1 since $L(\vec{v}) \neq 0$. Thus $\ker L$ has dimension $\dim V - 1$.

So then there is a basis S for $\ker L$ such that the matrix of B with respect to S is diagonal. Let $S' = S \cup \{\vec{v}\}$. Then for all $\vec{a} \in S$ we have $B(\vec{v}, \vec{a}) = B(\vec{a}, \vec{v}) = L(\vec{a}) = 0$. Thus S' makes B diagonal. □

Remark. There is a 1-1 correspondence between quadratic forms and symmetric bilinear forms.

If B is a symmetric bilinear form we can define a corresponding quadratic form

$$Q(\vec{x}) = B(\vec{x}, \vec{x}) \tag{7.18}$$

And if Q is a quadratic form we can define the corresponding B by

$$B(\vec{x}, \vec{y}) = \frac{1}{2}\left(Q(\vec{x} + \vec{y}) - Q(\vec{x}) - Q(\vec{y})\right) \tag{7.19}$$

It is clear that this map is symmetric, but we can show that it is bilinear by doing some algebra which I'm not typesetting because it's tedious and it's late. The key idea is that

$$Q(\vec{y}) = \sum_{1 \le i,j \le n} a_{ij} y_i y_j \qquad \text{where } a_{ij} = a_{ji} \tag{7.20}$$

Expanding $B(a\vec{x} + b\vec{y}, \vec{z})$ using this definition of Q will give the desired bilinearity.

Remark. The symmetric bilinear form for a quadratic form $Q(x) = \sum_{i \le j} a_{ij} x_i x_j$ is given by:

$$B(\vec{x}, \vec{y}) = \vec{x}^{\mathsf{T}} \begin{bmatrix} a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{21} & a_{22} & \cdots & \frac{12}{a\ 2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{n1} & \frac{1}{2}a_{n2} & \cdots & a_{nn} \end{bmatrix} \vec{y} \tag{7.21}$$

Remark. A diagonal quadratic form is one where the corresponding matrix is diagonal, or equivalently, where all cross-terms in the polynomial are 0