Math 245

James Yu

September 28, 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 $c \in \mathbb{C}$

$$0 \leq \|\vec{x} - c\vec{y}\|^2 \tag{1.3}$$
$$\leq \|\vec{x}\|^2 - \bar{c}\,\langle \vec{x}, \vec{y}\rangle - c\,\langle \vec{y}, \vec{x}\rangle + c\bar{c}\,\|y\|^2 \tag{1.4}$$

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

$$0 \leq \|\vec{x}\|^2 - \frac{\overline{\langle \vec{x}, \vec{y}\rangle}}{\|y\|^2}\langle \vec{x}, \vec{y}\rangle - \frac{\langle \vec{x}, \vec{y}\rangle}{\|y\|^2}\langle \vec{y}, \vec{x}\rangle + \frac{\langle \vec{x}, \vec{y}\rangle\,\overline{\langle \vec{x}, \vec{y}\rangle}}{\|y\|^4}\|y\|^2 \tag{1.5}$$
$$\leq \|\vec{x}\|^2 - \frac{|\langle \vec{x}, \vec{y}\rangle|^2}{\|y\|^2} - \frac{|\langle \vec{x}, \vec{y}\rangle|^2}{\|y\|^2} + \frac{|\langle \vec{x}, \vec{y}\rangle|^2}{\|y\|^2} \tag{1.6}$$
$$|\langle \vec{x}, \vec{y}\rangle|^2 \leq \|\vec{x}\|^2\,\|\vec{y}\|^2 \tag{1.7}$$
$$|\langle \vec{x}, \vec{y}\rangle| \leq \|\vec{x}\|\,\|\vec{y}\| \tag{1.8}$$

□

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 $\{\vec{w}_1, \dots \vec{w}_n\}$ 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 $\vec{v}_2 = \vec{w}_2 - \dfrac{\langle \vec{w}_2, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2}\vec{v}_1$

...

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

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

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

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

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

$$\langle \vec{v}_{i+1}, \vec{v}_j \rangle = \left\langle \vec{w}_{i+1} - \sum_{k=1}^{i} \frac{\langle \vec{w}_{i+1}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2}\vec{v}_k, \vec{v}_j \right\rangle \tag{1.13}$$

$$= \left\langle \vec{w}_{i+1} - \frac{\langle \vec{w}_{i+1}, \vec{v}_j \rangle}{\|\vec{v}_j\|^2}\vec{v}_j, \vec{v}_j \right\rangle \tag{1.14}$$

$$= \langle \vec{w}_{i+1}, \vec{v}_j \rangle - \frac{\langle \vec{w}_{i+1}, \vec{v}_j \rangle}{\|\vec{v}_j\|^2}\|\vec{v}_j\|^2 \tag{1.15}$$

$$= \langle \vec{w}_{i+1}, \vec{v}_j \rangle - \langle \vec{w}_{i+1}, \vec{v}_j \rangle \tag{1.16}$$

$$= 0 \tag{1.17}$$

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 $x \in \operatorname{span}\{\vec{w}_1, \dots, \vec{w}_{i-1}\}$. Thus $\vec{v}_i$ is a nonzero linear combination of $\{\vec{w}_1, \dots, \vec{w}_i\}$ and is therefore non-zero.

Thus $\{\vec{v}_1, \dots, \vec{v}_n\}$ 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. $\square$

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)**.**

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

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

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

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 $\{\vec{w}_1, \dots, \vec{w}_k\}$ 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 (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 *projection of* $\vec{v}$ *onto* W to be the first coordinate $\Psi(\vec{v})$, denoted:

$$\text{proj}_W\left(\vec{v}\right)$$

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

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

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 $\vec{v} \in V$:

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

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

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:

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

$$g_{\vec{v}}\left(c\vec{v}\right) = \left\langle T\left(c\vec{v}\right), \vec{y} \right\rangle \tag{2.17}$$
$$= c\langle T\vec{v}, \vec{y} \rangle \tag{2.18}$$
$$= c g_{\vec{y}}(\vec{v}) \tag{2.19}$$

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 $\vec{y} \in W^\perp$:

$$\begin{aligned} \langle T\vec{y}, c\vec{z}\rangle &= \overline{c}\,\langle \vec{y}, T^*\vec{z}\rangle && (3.7)\\ &= \overline{c}\lambda\,\langle \vec{y}, \vec{z}\rangle && (3.8)\\ &= 0 \text{ by choice of } \vec{y} && (3.9)\end{aligned}$$

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:

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

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.

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 3.3.3. *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 3.3.4. *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 3.3.3, $T^*T - I = 0$ so $T^*T = I$. Thus T^* is a left inverse of T, so since T is an endomorphism, $T^*T = I = TT^*$.

$\square$