EECS 16B Designing Information Devices and Systems II
Summer 2020 Note 16

1 Overview

In this note, we will mainly be building Linear Algebra fundamentals, specificly on inner products and orthogonality. We will take a deeper look at how the meaning behind orthonormal bases and matrices with orthonormal columns.

Once we have built these fundamentals on inner products and orthonormal matrices, we will revisit the idea of representing matrices in different bases. With this in mind, we take a look at **Schur Decomposition** which states that any matrix A has a representation R which is upper triangular and finally we will conclude by looking at one of most important results of Linear Algebra called the **Spectral Theorem** which states that symmetric matrices are orthogonally diagonalizable.

2 Inner Products

You may have seen the **inner product** or the dot-product from EE16A or Math 54. However, we will recap the most important properties of the inner product.

2.1 Definition

The inner product $\langle \cdot, \cdot \rangle$ on a vector space V over $\mathbb{R}$ is a function that takes in two vectors and outputs a scalar, such that $\langle \cdot, \cdot \rangle$ is symmetric, linear, and positive-definite.

- Symmetry: $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$
- Scaling: $\langle c\vec{u}, \vec{v} \rangle = c\langle \vec{u}, \vec{v} \rangle$ and $\langle \vec{u}, c\vec{v} \rangle = c\langle \vec{u}, \vec{v} \rangle$
- Additivity: $\langle \vec{u}+\vec{v}, \vec{w} \rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle$ and $\langle \vec{u}, \vec{v}+\vec{w} \rangle = \langle \vec{u}, \vec{v} \rangle + \langle \vec{u}, \vec{w} \rangle$
- Positive-definite: $\langle \vec{u}, \vec{u} \rangle \geq 0$ with $\langle \vec{u}, \vec{u} \rangle = 0$ if and only if $\vec{u} = \vec{0}$

A vector space V equipped with an inner product $\langle \cdot, \cdot \rangle$ is called an **Inner Product Space**.

The **standard inner product** in $\mathbb{R}^n$ is often referred to as the **dot product**

$$\langle \vec{x}, \vec{y} \rangle = \vec{x} \cdot \vec{y} = \vec{x}^T \vec{y} = x_1 y_1 + \ldots + x_n y_n \tag{1}$$

However, the inner product itself is defined over an arbitrary vector space V. This means that other vector spaces may have different definitions of multiplication through the inner product.

Another example you will see later in this course is the complex-inner product of the vector space $\mathbb{C}^n$ over the field $\mathbb{C}$.

2.1.1 Vector Space of Continuous Functions

Another vector space that you may encounter in courses like EE120 or Physics 137A is the vector space of Continuous Functions $f : \mathbb{R} \to \mathbb{R}$ on the interval $[a,b]$. Here the "vectors" of this vector space are functions and we can define an inner product as

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt \tag{2}$$

2.1.2 Vector Space of Random Variables

Another example of an inner product that you may see in courses like EE126 is one over the vector space of Random Variables. If X and Y are random variables on a sample space Ω, then we can define the following inner product [1]

$$\langle X, Y \rangle = \mathbb{E}[XY] \tag{3}$$

We have shown all of these examples to emphasize the fact that an inner product is a generalized map that takes two vectors in a vector space and outputs a scalar. Whether the vector space is $V = \mathbb{R}^n$, or some other vector space should not matter and the four properties listed above will always hold.

2.2 Geometry of Inner Products

By defining an inner product, we can build some geometric intuition in our inner product space. Note that our vector space again need not be $\mathbb{R}^n$ and one of the amazing ideas behind an inner product space is that we are able to visualize vectors geometrically regardless of what the vectors in our space are.

An inner product allows us to define the **norm** of a vector. Geometrically, this is the length of the vector

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} \tag{4}$$

Another geometric viewpoint comes from the cosine definition of inner products.

$$\langle \vec{x}, \vec{y} \rangle = \|\vec{x}\| \|\vec{y}\| \cos(\theta) \tag{5}$$

where θ is the angle between the two vectors. Again note that these vectors need not be in $\mathbb{R}^n$.

We define two vectors to be orthogonal if their inner product is equal to zero.

$$\langle \vec{x}, \vec{y} \rangle = 0 \tag{6}$$

Note how the geometric definition of inner products tells us that $\theta = \frac{\pi}{2}$. This allows us to view orthogonal vectors as prependicular in an inner product space. [2]

Lastly, we note that a set of vectors U is orthonormal is they are all orthogonal and have norm equal to 1.

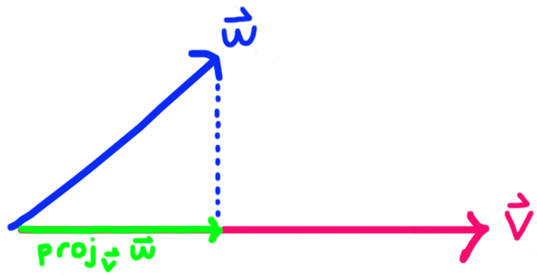
- For all vectors $\vec{u}_i$ and $\vec{u}_j$, $\langle \vec{u}_i, \vec{u}_j \rangle = 0$ for $i \neq j$
- All vectors $\vec{u}_i$ have norm $1 : \|\vec{u}_i\|^2 = \langle \vec{u}_i, \vec{u}_i \rangle = 1$

[1] If you haven't taken CS 70, don't worry about these examples. They are here just to show alternate inner products on vector spaces outside of $\mathbb{R}^n$.

[2] In the Inner Product Space of random variables, if two vectors with zero mean are orthogonal, the covariance is equal to 0. Under this geometric interpretation, uncorrelated random variables are perpendicular to each other.

2.3 Vector Projections

Another geometric insight that we can build comes from projecting a vector $\vec{w}$ onto another vector $\vec{v}$



We call this projected vector $\text{proj}_{\vec{v}}\,\vec{w}$ and through the cosine definition of inner products, we can show that it is equal to

$$\text{proj}_{\vec{v}}\,\vec{w} = \frac{\langle \vec{w}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle}\vec{v} = \frac{\langle \vec{w}, \vec{v} \rangle}{\|\vec{v}\|^2}\vec{v} \tag{7}$$

If we project onto a vector $\vec{v}$ that is normalized, $\|\vec{v}\| = 1$, then the projection will be

$$\text{proj}_{\vec{v}}\,\vec{w} = \langle \vec{w}, \vec{v} \rangle \vec{v} \tag{8}$$

Note how the residual vector $\vec{r} = \vec{w} - \text{proj}_{\vec{v}}\,\vec{w}$ is orthogonal to the vector $\vec{v}$.

3 Orthonormality

Recall from the previous section that we defined a set of vectors $U = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ to be orthonormal if

- For all vectors $\vec{u}_i$ and $\vec{u}_j$, $\langle \vec{u}_i, \vec{u}_j \rangle = 0$ for $i \neq j$
- All vectors $\vec{u}_i$ have norm $1 : \|\vec{u}_i\|^2 = \langle \vec{u}_i, \vec{u}_i \rangle = 1$

3.1 Orthonormal Matrices

A matrix U with orthonormal columns has the special property that

$$U^T U = \begin{bmatrix} - & \vec{u}_1^T & - \\ & \vdots & \\ - & \vec{u}_k^T & - \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{u}_1 & \ldots & \vec{u}_k \\ | & & | \end{bmatrix} = \begin{bmatrix} \vec{u}_1^T \vec{u}_1 & \vec{u}_1^T \vec{u}_2 & \vdots & \vec{u}_1^T \vec{u}_k \\ \vec{u}_2^T \vec{u}_1 & \vec{u}_2^T \vec{u}_2 & \ddots & \vec{u}_2^T \vec{u}_k \\ \vdots & \vdots & \ddots & \vdots \\ \vec{u}_k^T \vec{u}_1 & \vec{u}_k^T \vec{u}_2 & \ldots & \vec{u}_k^T \vec{u}_k \end{bmatrix} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} = I \tag{9}$$

It is quite unfortunate that in standard Linear Algebra literature, a square matrix with orthonormal columns is defined to be an **orthogonal matrix.** We will be calling these matrices **orthonormal matrices** to emphasize the fact that all of the columns are orthonormal.

Note: U does not have to be square. However, if U is square, then it must be that $U^T = U^{-1}$.

Warning: If U has $k < n$ columns, then $UU^T \neq I$. [3]

[3] The matrix U has rank $k < n$. Since $\text{Rank}(A) = \text{Rank}(AA^T)$, UU^T cannot be the identity since it only has rank $k < n$.

In addition, an orthonormal matrix does not change the norm of a vector.

$$||U\vec{x}||^2 = \vec{x}^T U^T U\vec{x} = \vec{x}^T\vec{x} = ||\vec{x}||^2 \tag{10}$$

In fact, we can view $U\vec{x}$ as a rotation of the vector $\vec{x}$.

3.2 Gram-Schmidt

Every n-dimensional vector space V has basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. However, does every vector space V have an orthonormal basis $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$?

In this section, we will show that it is indeed possible to pick a basis in which all of the vectors are orthonormal. To do this, we must ensure that

$$\text{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} = \text{Span}\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\} \tag{11}$$

The process of creating this orthonormal basis is called the **Gram-Schmidt Process**.

Gram-Schmidt is an algorithm that takes a basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ and generates an orthonormal set of vectors $\{\vec{u}_1, \ldots, \vec{u}_n\}$ that span the same space as the original set. We will walk through the algorithm step by step.

3.2.1 Base Case:

Let's start with the first vector $\vec{v}_1$. Since there is only one vector, the set $\{\vec{v}_1\}$ is already orthogonal. Therefore, all we have to do is normalize the vector

$$\vec{u}_1 = \frac{\vec{v}_1}{||\vec{v}_1||} \tag{12}$$

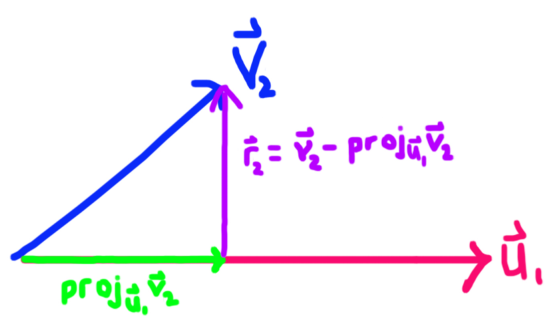
3.2.2 Projection Step:

Now moving onto the second vector $\vec{v}_2$, the vectors $\{\vec{u}_1, \vec{v}_2\}$ are not necssarily orthogonal. However, if we project $\vec{v}_2$ onto $\vec{u}_2$, then recall that the residual of the projection will be orthogonal to $\vec{u}_1$.

$$\text{proj}_{\vec{u}_1} \vec{v}_2 = \langle \vec{v}_2, \vec{u}_1 \rangle \vec{u}_1 \tag{13}$$

3.2.3 Residual Step:

The residual $\vec{r}_2 = \vec{v}_2 - \langle \vec{v}_2, \vec{u}_1 \rangle \vec{u}_1$ of the projection is orthogonal to $\vec{u}_1$.



3.2.4 Normalization Step:

The set $\{\vec{u}_1, \vec{r}_2\}$ is orthogonal, but not orthonormal. It remains to normalize $\vec{r}_2$. [4]

$$\vec{u}_2 = \frac{\vec{r}_2}{\|\vec{r}_2\|} \tag{14}$$

3.2.5 Projection onto a Subspace

For the later iterations when $k > 2$, we will be projecting the next vector $\vec{v}_{k+1}$ onto the subspace spanned by $U_i = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$. One way to do this is to realize that any vector in the subspace is defined through a unique set of coordinates with respect to its basis.

$$\vec{w}_{k+1} = \mathrm{proj}_{U_i} \vec{v}_{k+1} = \alpha_1 \vec{u}_1 + \ldots + \alpha_k \vec{u}_k \tag{15}$$

Since all of the u_i are orthonormal, we can show that

$$\langle \vec{w}_{k+1}, \vec{u}_i \rangle = \alpha_1 \langle \vec{u}_1, \vec{u}_i \rangle + \ldots + \alpha_k \langle \vec{u}_k, \vec{u}_i \rangle = \alpha_i \tag{16}$$

Therefore, the projection of a vector $\vec{v}_{k+1}$ onto an subspace spanned by an orthonormal set is

$$\vec{w}_{k+1} = \mathrm{proj}_{U_i} \vec{v}_{k+1} = \langle \vec{v}_{k+1}, \vec{u}_1 \rangle \vec{u}_1 + \ldots + \langle \vec{v}_{k+1}, \vec{u}_k \rangle \vec{u}_k \tag{17}$$

3.3 Algorithm:

The process of projecting and performing the residual step will continue on for n iterations. We provide a pseudocode algorithm of the Gram-Schmidt process.

---

procedure GRAM-SCHMIDT($v_1, v_2, \ldots, v_n$)
    $U = \{\}$ ▷ Initialize empty set
    for $i = 1$ to n **do**
        if $i = 1$ **then**
            $u_1 \leftarrow \frac{v_1}{\|v_1\|}$
        else
            $w_i \leftarrow \langle v_i, u_1 \rangle + \ldots + \langle v_i, u_{i-1} \rangle$ ▷ Projection Step
            $r_i \leftarrow v_i - w_i$ ▷ Residual Step
            $u_i = \frac{r_i}{\|r_i\|}$ ▷ Normalization Step
        end if
        Add u_i to U
    end for
    return U
end procedure

---

[4]Note: If $\vec{v}_2$ was orthogonal to $\vec{v}_1$ to begin with, the projection $\mathrm{proj}_{\vec{u}_1} \vec{v}_2 = \vec{0}$. In this case, no residual step is needed and we can normalize $\vec{v}_2$.

4 Schur Decomposition (Optional)

The Gram-Schmidt process tells us that we can pick an orthonormal basis $\{\vec{u}_1, \ldots, \vec{u}_n\}$ for an $n-$dimensional vector space V. This will now let us prove that any matrix A can in fact be decomposed into the following product

$$\boxed{A = URU^T} \tag{18}$$

where U is an orthonormal matrix and R is upper-triangular with the eigenvalues of A on its diagonal. This form is called **Schur Decomposition** and any matrix, even those that are non-diagonalizable can be put into this form. This form was alluded to multiple times in the previous notes and we will finally prove its existence.

4.1 Proof

To prove that any $n \times n$ matrix A has an upper-triangular representation in an orthonormal basis is quite a difficult task. We will break down the problem into multiple steps and will built it up using mathematical induction on the size of the matrix n.

4.1.1 Base Case:

A 1×1 matrix is a scalar. Therefore picking $u = 1$ and $r = a$ shall suffice.

4.1.2 Inductive Hypothesis:

Assume that an $(n-1) \times (n-1)$ matrix A_{n-1} can be put into upper-triangular form

$$A_{n-1} = U_{n-1} R_{n-1} U_{n-1}^T \tag{19}$$

4.1.3 Lemma:

Before we move onto the inductive step, we prove the existence of an eigenvector $\vec{v}$ and eigenvalue λ.

A matrix A must have at least one eigenvalue λ. This is because the characteristic polynomial

$$\chi_A(\lambda) = \det(A - \lambda I) \tag{20}$$

is an n^{th} degree polynomial. By the Fundamental Theorem of Algebra, this polynomial must have at least one root $\lambda \in \mathbb{C}$. As a result, the matrix A must have an eigenpair $(\lambda, \vec{v})$.

4.1.4 Inductive Step:

We now need to show that an $n \times n$ matrix A has a basis U in which $A = URU^T$. Let us first pick an orthonormal basis V for $\mathbb{R}^n$.

We start by adding a normalized eigenvector $\vec{v}$ and we will pick the remaining vectors using Gram-Schmidt.

$$V = \begin{bmatrix} \vec{v} & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix} \qquad W = \begin{bmatrix} \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix} \tag{21}$$

We define W to be an $n \times (n-1)$ matrix whose columns are all orthonormal to $\vec{v}$. Note that we have picked V to be a matrix with orthonormal columns meaning $V^{-1} = V^T$.

Now let us look at the matrix product $V^T AV$

$$V^T AV = \begin{bmatrix} \vec{v}^T \\ W^T \end{bmatrix} A \begin{bmatrix} \vec{v} & W \end{bmatrix} = \begin{bmatrix} \vec{v}^T \\ W^T \end{bmatrix} \begin{bmatrix} A\vec{v} & AW \end{bmatrix} \tag{22}$$

$$= \begin{bmatrix} \vec{v}^T A\vec{v} & \vec{v}^T AW \\ W^T A\vec{v} & W^T AW \end{bmatrix} = \begin{bmatrix} \vec{v}^T \lambda \vec{v} & \vec{v}^T AW \\ \lambda W^T \vec{v} & W^T AW \end{bmatrix} \tag{23}$$

$$= \begin{bmatrix} \lambda & \vec{v}^T AW \\ \vec{0} & W^T AW \end{bmatrix} \tag{24}$$

The last line follows from the fact that the columns of W are all orthogonal to the vector $\vec{v}$ by construction.

To conclude our proof, the matrix $W^T AW$ is an $(n-1) \times (n-1)$ matrix and we know from our inductive hypothesis that it has an upper-triangular representation.

$$W^T AW = U_{n-1} R_{n-1} U_{n-1} \tag{25}$$

To verify that A can be put in upper-triangular form, we claim that the basis $U = \begin{bmatrix} \vec{v} & WU_{n-1} \end{bmatrix}$. First we show that this basis U is indeed orthonormal

$$U^T U = \begin{bmatrix} \vec{v}^T \\ U_{n-1}^T W^T \end{bmatrix} \begin{bmatrix} \vec{v} & WU_{n-1} \end{bmatrix} = \begin{bmatrix} \vec{v}^T \vec{v} & \vec{v}^T WU_{n-1} \\ U_{n-1}^T W^T \vec{v} & U_{n-1}^T W^T WU_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0} & I_{(n-1)} \end{bmatrix} = I \tag{26}$$

Then we can multiply out $U^T AU$ to verify that it is indeed an upper-triangular R

$$U^T AU = \begin{bmatrix} \vec{v}^T \\ U_{n-1}^T W^T \end{bmatrix} A \begin{bmatrix} \vec{v} & WU_{n-1} \end{bmatrix} = \begin{bmatrix} \vec{v}^T \\ U_{n-1}^T W^T \end{bmatrix} \begin{bmatrix} A\vec{v} & AWU_{n-1} \end{bmatrix} \tag{27}$$

$$= \begin{bmatrix} \lambda & \vec{v}^T AWU_{n-1} \\ \lambda U_{n-1}^T W^T \vec{v} & U_{n-1}^T W^T AWU_{n-1} \end{bmatrix} = \begin{bmatrix} \lambda & \vec{q}^T \\ \vec{0} & R_{n-1} \end{bmatrix} \tag{28}$$

We leave $\vec{q}^T$ as a vector since its values do not affect the upper-triangular form of A and the last step comes from substituting our inductive hypothesis $W^T AW = U_{n-1} R_{n-1} U_{n-1}$.

4.2 Result

Now that we have proven the result that a matrix A has an upper-triangular representation in some orthonormal basis U, we can go back and fully understand where this fact was applied.

4.2.1 Solving Differential Equations

When trying to solve a vector differential equation, we had always assumed that A was a diagonalizable matrix. However, now that we have equipped ourselves with Schur Decomposition, we are able to change coordinates and solve our upper-triangular system

$$\frac{d}{dt}\vec{x}(t) = A\vec{x}(t) + B\vec{u}(t) \implies \frac{d}{dt}\vec{z}(t) = R\vec{z}(t) + U^T B\vec{u}(t) \tag{29}$$

We used this fact to analyze an RLC circuit that was critically damped and we also used this fact when discretizing a system where the matrix A was non-diagonalizable.

4.2.2 Stability

A very similar argument was used when proving the eigenvalue test for the stability of a linear system. Again when A was non-diagonalizable, we viewed A in its Schur Form R and saw how we could back-substitute to show that all of the eigenvalues must be stable.

5 Symmetric Matrices

A special class of matrices that we will be looking at in the upcoming notes is the matrix $A^T A$. This matrix belongs to a very special class of matrices that are **symmetric.**

In fact, symmetric matrices are so special, we can show that they are always diagonalizable. This result is so important, it is called the **Spectral Theorem** and is referred to as one of the greatest results in Linear Algebra.

5.1 Spectral Theorem

If a matrix A with real entries is symmetric, then the following statements are true

- All of the eigenvalues of A are real.
- All of the eigenevectors of A are orthogonal.
- The matrix A can be diagonalized by an orthonormal basis made up of eigenvectors.

5.1.1 Real Eigenvalues

To prove the first statement, we will show that $\lambda = \overline{\lambda}$ which implies λ is real. Let's start by writing out $A\vec{v} = \lambda\vec{v}$ and conjugating both sides

$$Av = \lambda v \implies \overline{A}\overline{v} = \overline{\lambda}\overline{v} \tag{30}$$

Since A is a real matrix, $\overline{A} = A$ and we now take the transpose of both sides

$$\overline{v}^T A^T = \overline{\lambda}\overline{v}^T \implies \overline{v}^T A = \overline{\lambda}\overline{v}^T \tag{31}$$

Then if we right-multiply by $\vec{v}$, it follows that

$$\overline{v}^T Av = \overline{\lambda}\overline{v}^T v \implies \lambda\overline{v}^T v = \overline{\lambda}\overline{v}^T v \tag{32}$$

This is only possible when $\lambda = \overline{\lambda}$ since $\overline{v}^T\vec{v} \neq 0$ and $\vec{v}$ is an eigenvector not equal to $\vec{0}$.

5.1.2 Orthogonal Eigenvectors

We shall prove that if A has distinct eigenvalues $\lambda_1 \neq \lambda_2$, then their eigenspaces are orthogonal.

$$\lambda_1 \langle \vec{v}_1, \vec{v}_2 \rangle = \langle \lambda_1 \vec{v}_1, \vec{v}_2 \rangle = \langle A\vec{v}_1, \vec{v}_2 \rangle \tag{33}$$
$$= \vec{v}_1^T A^T \vec{v}_2 = \vec{v}_1^T A \vec{v}_2 = \lambda_2 \vec{v}_1^T \vec{v}_2 \tag{34}$$
$$= \lambda_2 \langle \vec{v}_1, \vec{v}_2 \rangle \tag{35}$$

$\lambda_1 \neq \lambda_2$ by construction so it must be that $\langle \vec{v}_1, \vec{v}_2 \rangle = 0$ proving eigenvectors of distinct eigenvalue are orthogonal.

5.1.3 Diagonalizability

Lastly, we must show that A can be diagonalized by an orthonormal basis made up of eigenvectors. To do this, we can start with the Schur-Decomposition of A.

$$A = URU^T \tag{36}$$

Since A is symmetric, $A = A^T$ and it must be that

$$A = A^T = UR^T U^T \implies R = R^T \tag{37}$$

R is an upper-triangular matrix that is also symmetric since $R = R^T$. Therefore, we conclude that R must be a diagonal matrix.

While subtle, R being diagonal shows that the matrix U forms a basis for $\mathbb{R}^n$ made up of eigenvectors since

$$AU = \begin{bmatrix} A\vec{u}_1 & \dots & A\vec{u}_n \end{bmatrix} = UR = \begin{bmatrix} r_1\vec{u}_1 & \dots & r_n\vec{u}_n \end{bmatrix} \tag{38}$$

Alternatively, we can also prove this property by performing the same induction that we did to prove Schur decomposition, but if we've already proven something, why not use its result?

