EECS 16B Designing Information Devices and Systems II
Fall 2016

Linear Algebra Notes

Introduction

In this set of notes, we will derive the linear least squares equation, study the properties symmetric matrices like A^*A including the complex spectral theorem and conclude with the Singular Value Decomposition and its applications.

Linear Least Squares - Derivation

Consider a systems of equations of the form,

$$A\vec{x} = \vec{b} \tag{1}$$

Here, A is a linear function from $\mathbb{C}^n$ to $\mathbb{C}^m$ represented as a matrix, x is an n dimensional vector of unknowns and b is an m dimensional vector. Let A have rank k[1].

If b is in the range[2]space of A, there exists a solution x to (1) by definition. If A has a null space[3], x is not the unique solution to (1). For example, let y be in the null space of A. Then $x+y$ is also a solution[4]. We will address this case after we have introduced the singular value decomposition.

What about if b is not in range(A)? This means that there is no x that exactly satisfies (1). To overcome this, we try to get as close as we can. This is to say, we are trying to find x such that Ax is as close as possible, distance wise, to b. To do so, we will think about the range space of A and how to characterize the space that is not reachable by A.

Lets construct a basis for $\mathbb{C}^m$ around the range space of A. Let range(A) be spanned by the orthonormal basis vectors denoted by $\mathscr{A}$[5].

$$\mathscr{A} = \{a_1, a_2, \ldots, a_k\}$$

We can then extend this to a complete basis of $\mathbb{C}^m$ with $m-k$ more orthogonal vectors. Let $\mathscr{B}$ denote the basis for $\mathbb{C}^m$ built around the range space of A.

$$\mathscr{B} = \{a_1, a_2, \ldots, a_k, c_1, \ldots, c_{m-k}\}$$

[1]Recall that this means the dimension of the range space of A is k
[2]Recall that the range of A is the subspace spanned by the columns of A
[3]Recall that the null space of A is the subspace of vectors that A takes to the zero vector
[4]As you should verify
[5]We can pick any set basis vectors that span the range space of A and run Gram Schmidt to orthonormalize them

Let,

$$\mathscr{C} = \{c_1, c_2, \ldots, c_{m-k}\}$$

Note that anything in the span of $\mathscr{C}$ is never reached by A for any input. Particularly, by construction, everything in the span of $\mathscr{C}$ is orthogonal to the span of $\mathscr{A}$. This makes the span of $\mathscr{C}$ the **orthogonal compliment** of the span of $\mathscr{A}$. Since the span of range(A) is the span of $\mathscr{A}$, we can conclude the orthogonal compliment of range(A) to be the span of $\mathscr{C}$, which is denoted as range$(A)^{\perp}$.

Taking into account that $\mathbb{C}^m$ is spanned by $\mathscr{B}$ which is intern composed of $\mathscr{A}$ and $\mathscr{C}$, it should follow that,

$$\mathbb{C}^m = \text{range}(A) \oplus \text{range}(A)^{\perp}$$

The $\oplus$ is called a **direct sum**, and it denotes that, other than the zero vector, there is no common vector between range(A) and range$(A)^{\perp}$.

Why have we done all this? Well, there is a relationship between range$(A)^{\perp}$ and null(A^*)[6]. Let x be any element in the null space of A^* and let y be any element in the domain C^n. Observe that,

$$y^* A^* x = 0 \implies (Ay)^* x = 0$$

Since y and x were arbitrary, this means that any element in the null space of A^* is perpendicular to all elements of the range space. In other words,

$$\text{range}(A)^{\perp} = \text{null}(A^*)$$

This is super useful. We can conclude that,

$$\mathbb{C}^m = \text{range}(A) \oplus \text{null}(A^*)$$

How is this useful? Let us revisit (1). We want to find an x such that $\|Ax - b\|$ is as small as possible. We do this by making sure that there is nothing in $Ax - b$ that is still in the range space of A, which means that we have done the best we can. This means that $Ax - b$ must be in the orthogonal complement of A, which consequently means in must be in the null space of A^*. Using this fact, we get that,

$$A^*(Ax - b) = 0 \implies A^* A x = A^* b$$

This is the linear least squares equation which allows us to solve for a x such that Ax is the closest we can get to b.

Some Definitions

Let T be a square matrix from $\mathbb{C}^n$ to $\mathbb{C}^n$.

[6] $(\cdot)^*$ denotes the adjoint of A. For the purposes of this class, this is just the complex conjugate transpose

Symmetric Matrices

A matrix T is symmetric if $T = T^*$.

Positive (Semi-) Definite Matrices

A matrix T is a positive semi-definite matrix if it is symmetric and,

$$v^*Tv \geq 0 \text{ for all } v \in \mathbb{C}^n$$

Additionally, it is positive definite if,

$$v^*Tv = 0 \text{ if and only if } v = 0$$

Properties Of Symmetric Matrices

Eigenvalues are real

Let λ be an eigenvalue with let v being the corresponding eigenvector. Observe that,

$$v^*Tv = v^*\lambda v = \lambda v^*v$$

Similarly,

$$v^*Tv = (T^*v)^*v = (Tv)^*v = \lambda^* v^*v$$

Equating the two equations, we get,

$$\lambda v^*v = \lambda^* v^*v$$

Since v is an eigenvector, it is nonzero which implies that v^*v is non-zero. Dividing both sides by v^*v, we get,

$$\lambda = \lambda^*$$

Eigenvectors corresponding to distinct eigenvalues are orthogonal

Let λ_1 and λ_2 be two distinct eigenvalues corresponding to eigenvectors v_1 and v_2. Since λ_1 and λ_2 are distinct, $\lambda_1 - \lambda_2 \neq 0$. Using the fact that the eigenvalues are real and that T is symmetric, observe that,

$$\begin{aligned}
(\lambda_1 - \lambda_2)v_1^*v_2 &= \lambda_1 v_1^*v_2 - \lambda_2 v_1^*v_2 \\
&= (Tv_1)^*v_2 - v_1^*(Tv_2) \\
&= v_1^*Tv_2 - v_1^*Tv_2 \\
&= 0
\end{aligned}$$

Therefore,

$$(\lambda_1 - \lambda_2)v_1^* v_2 = 0 \implies v_1^* v_2 = 0$$

We use the fact that $\lambda_1 - \lambda_2 \neq 0$. Thus v_1 and v_2 are orthogonal to each other.

Complex Spectral Theorem

Statement

Let T be a symmetric matrix from $\mathbb{C}^n$ to $\mathbb{C}^n$. Then,

1. There exists n linearly independent eigenvectors of T that form a basis for $\mathbb{C}^n$. In other words, T is diagonalizable. Furthermore, the eigenvectors of T are orthonormal.
2. The eigenvalues of T are real.

Proof

We are going to use a proof by induction. We are going to try and recursively build up our proof by abstracting the dimension of the matrix. The rough structure is as follows.

- We will show that it works for a $[1 \times 1]$ matrix.
- We assume it works for a $[k \times k]$ matrix.
- Using the fact that it works for a $[k \times k]$ matrix, we will show that it works for a $[k+1 \times k+1]$ matrix.

Lets get started. Let T be a $[1 \times 1]$ matrix such that,

$$T = [t]$$

Note that T is symmetric and has an eigenvalue t with eigenvector 1. Furthermore, since we only have one basis vector, we can conclude that all the basis vectors are orthonormal as well. This is our base case from which we will build our proof from.

Lets generalize to an $[n \times n]$ matrix. Let the first eigenvalue be λ_1 and the eigenvector be u_1. (We are guaranteed at least one eigenvalue-eigenvector pair.)

We can extend u_1 to a complete basis. We can then run Gram Schmidt and normalize to obtain an orthonormal basis $\mathscr{U}$ for $\mathbb{C}^n$.

$$\mathscr{U} = \{u_1, u_2, \ldots, u_n\}$$

Define U to be a matrix whose columns consist of the basis vectors in $\mathscr{U}$.

$$U = \begin{bmatrix} | & \dots & | \\ u_1 & \dots & u_n \\ | & \dots & | \end{bmatrix}$$

Since the columns of U are orthonormal, it follows that,

$$U^{-1} = U^* \text{ and } U^*U = UU^* = I$$

Let $\widetilde{U}$ be defined as,

$$\widetilde{U} = \begin{bmatrix} | & \dots & | \\ u_2 & \dots & u_n \\ | & \dots & | \end{bmatrix}$$

Note that,

$$U = \left[u_1, \widetilde{U}\right]$$

Let S be the matrix representation of T with respect to the U basis. It follows that,

$$T = USU^{-1} \text{ or } T = USU^* \tag{2}$$

Now, since u_1 is an eigenvector, it follows that,

$$S = \left[\begin{array}{c:c} \lambda_0 & 0 \\ \hdashline 0 & Q \end{array}\right]$$

Here, Q is a $[n-1 \times n-1]$ matrix that is equal to,

$$Q = \widetilde{U}^* T \widetilde{U}$$

This can be verified by looking at (2) in the following manner.

$$T = \left[u_1, \widetilde{U}\right] \left[\begin{array}{c:c} \lambda_0 & 0 \\ \hdashline 0 & \widetilde{U}^* T \widetilde{U} \end{array}\right] \begin{bmatrix} u_1^* \\ \widetilde{U}^* \end{bmatrix} \tag{3}$$

Note that,

$$Q^* = \widetilde{U}^* T^* \widetilde{U} = \widetilde{U}^* T \widetilde{U} = Q$$

Therefore, Q is a symmetric $[n-1 \times n-1]$ matrix. Since this is of the lower dimension, we can use our inductive assumption. Let V be an $[n-1 \times n-1]$ matrix consisting of the orthonormal eigenvectors of Q

and Λ be the diagonal matrix consisting of the eigenvalues of Q. Using this in (3), we get,

$$T = \begin{bmatrix} u_1, \widetilde{U} \end{bmatrix} \left[\begin{array}{c:c} \lambda_0 & 0 \\ \hdashline 0 & V\Lambda V^* \end{array}\right] \begin{bmatrix} u_1^* \\ \widetilde{U}^* \end{bmatrix}$$

This is equivalent to,

$$T = \begin{bmatrix} u_1, \widetilde{U}V \end{bmatrix} \left[\begin{array}{c:c} \lambda_0 & 0 \\ \hdashline 0 & \Lambda \end{array}\right] \begin{bmatrix} u_1^* \\ (\widetilde{U}V)^* \end{bmatrix} \tag{4}$$

We are almost done! We Just need to verify that $\widetilde{U}V$ consists of orthonormal columns that are orthonormal to u_1 as well. Note that,

$$(\widetilde{U}V)^*\widetilde{U}V = V^*\cancelto{I}{\widetilde{U}^*\widetilde{U}}V = V^*V = I$$

We have used the fact that $\widetilde{U}$ and $\widetilde{V}$ consists of orthonormal vectors. Similarly, u_1 is perpendicular to any column in $\widetilde{U}V$ since,

$$u_1^*\widetilde{U}V = \underbrace{[0, 0, \ldots, 0]}_{[1\times n-1]}$$

Thus, we have successfully shown that any symmetric matrix T can be diagonalized and the corresponding eigenvectors are orthonormal to each other.

In the previous set of notes, we considered the eigenvalue decomposition of symmetric matrices. In this set of notes, we will use the complex spectral theorem to build a generalization of the eigenvalue decomposition based on inner products called the Singular Value Decomposition.

Singular Value Decomposition (SVD)

Statement

Let T be a matrix from $\mathbb{C}^m$ to $\mathbb{C}^n$. There exists a $[m \times m]$ matrix U consisting of orthonormal vectors, a $[n \times n]$ matrix V consisting of orthonormal vectors, and a $[m \times n]$ diagonal matrix S consisting of real values such that,

$$T = USV^*$$

Proof

Assuming $m > n$. The proof is identical in both cases.

Consider T^*T. By the complex spectral theorem, we know that T^*T is diagonalizable with orthonormal eigenvectors and the eigenvalues are real.

$$T^*T = V\Lambda V^*$$

Where,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Further more, T^*T is positive definite.

$$v^*T^*Tv = ||Tv||_2^2 > 0 \text{ for all } v \neq 0 \text{ and } ||Tv||_2^2 = 0 \text{ if and only if } v = 0$$

Consequently, all the eigenvalues are non-negative. Consider any particular eigenvector v_i (which is by definition non-zero). We know that,

$$v_i^*T^*Tv_i > 0$$

Now,

$$v_i^*(T^*Tv_i) = v_i^*\lambda_i v_i = \lambda_i v_i^* v_i \geq 0$$

Since $v_i^*v_i > 0, \lambda_i$ must be non-negative as well.

Now, define $s_i = \sqrt{\lambda_i}$. This is well defined since the eigenvalues are real and positive. Let S be,

$$S = \begin{bmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & s_n \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

We now simply need to construct U. Consider a particular vector u_i and define it whenever possible as follows.

$$u_i = \frac{1}{s_i}Tv_i$$

Note that,

$$u_i^*u_i = \frac{v_i^*T^*Tv_i}{s_i^2} = \frac{\lambda_i v_i^* v_i}{\lambda_i} = 1$$

Further more, since V consists of orthonormal columns the u_is constructed this way will be orthogonal. In case there are insufficient number of u_is that can be constructed from the v_is, we simply complete U by

making U rank m and orthonormalize the added columns.

Interpretation

Let the singular value decomposition of T, defined from $\mathbb{C}^m$ to $\mathbb{C}^n$, be,

$$T = USV^*$$

1. U is a basis for $\mathbb{C}^m$.
2. V is a basis for $\mathbb{C}^n$.
3. S tells you T scales the V basis vectors.
4. To think of the SVD as a sequence of operators, V^* decomposes the input vector into the V basis, S scales the vectors appropriately and U rotates the result into the U basis.

Dyadic form of a matrix

The SVD allows us to characterize the SVD in what is called the Dyadic form. A dyad, or the outer product, between two vectors u and v is denoted as $u \otimes v$ and can be written as follows.

$$u \otimes v = uv^*$$

We can write T in terms of its singular vectors in Dyadic form, which makes it easier to see the structure of T.

$$T = \sum_{i=1}^{\min(m,n)} s_i u_i v_i^*$$

We can use the dyadic form of a matrix to construct low-rank approximations of matrices.

Pseudoinverse

Now that we have the SVD, we will consider the linear least squares case of when there exists a null space. Typically speaking, we would like to find the minimum norm solution x that satisfies (1). Let A, a $[m \times n]$ matrix, have the following singular value decomposition.

$$A = s_1 u_1 v_1^* + s_2 u_2 v_2^* + \cdots + s_k u_k v_k^* \text{ [7]}$$

We construct the pseudoinverse, denoted as $A^\dagger$, as follows.

$$A^\dagger = \frac{1}{s_1} v_1 u_1^* + \frac{1}{s_2} v_2 u_2^* + \cdots + \frac{1}{s_k} v_k u_k^* \text{ [8]}$$

[7] What is the rank of A?

[8] What does $A^\dagger$ look like in matrix form?

We leave it to the reader to verify that $x = A^{\dagger}b$ would satisfy (1) and that x is the minimum norm solution.

Low Rank Approximations of Matrices

Consider a matrix $T : \mathbb{C}^m \to \mathbb{C}^n$ of rank p. Say we want to construct the best rank $k < p$ approximation of T, which we will denote to be $\widetilde{T}$. In order to do this, we first need to make a couple of definitions.

- **Inner Product.** The concept of an inner product is the generalization of the commonly use dot product. An inner product is a function that takes into two elements of a vector space as an ordered pair (u, v) and outputs a complex number[9] which is denoted as $\langle u, v\rangle$. An inner product is required to satisfy the follow conditions.

  1. Let v be any element of the vector space. Then

     $$\langle v, v\rangle \geq 0$$

     Further more,

     $$\langle v, v\rangle = 0 \text{ if and only if } v = 0$$

  2. Let u, v, w be arbitrary elements of the vector space and let α and β be complex numbers. Then,

     $$\langle \alpha u + \beta v, w\rangle = \alpha\langle u, w\rangle + \beta\langle v, w\rangle$$

  3. Let u, v be arbitrary elements of the vector space. Then,

     $$\langle u, v\rangle = \overline{\langle v, u\rangle}$$

  You should verify that, in the vector space $\mathbb{C}^n$, the regular dot product satisfies the above properties. That is to say,

  $$\langle u, v\rangle = u^* v$$

  Recall the u^* denotes the complex conjugate transpose. It is worthwhile to note that the idea of projections still makes sense with this definition and Gram-Schmidt is completely valid. As as example, consider the follow projection of u onto v.

  $$\text{proj}_v u = \frac{\langle u, v\rangle}{\langle v, v\rangle} v$$

- **Pythagoras Theorem.** If u and v are orthonormal vectors, then,

  $$\|u + v\|_2^2 = \|u\|_2^2 + \|v\|_2^2$$

- **Trace.** The trace is only defined on square matrices. The trace of a square matrix T is defined to be the sum of the eigenvalues. The trace satisfies the following properties.

  1. The trace of a square matrix A is equal to the sum of its diagonal elements.

[9] Strictly speaking, an inner product maps to an element of the field that the vector space is defined with

2. Let A and B be square matrices.

$$\text{trace}(A+B) = \text{trace}(A) + \text{trace}(B)$$

3. Let A be a square matrix and c be a complex number.

$$\text{trace}(cA) = c \cdot \text{trace}(A)$$

4. Let A be a $[m \times n]$ matrix and B be a $[n \times m]$ matrix. Then,

$$\text{trace}(AB) = \text{trace}(BA)$$

- **Trace as a Matrix Inner Product.** Using the trace, we can define an inner product of matrices. We leave it as an exercise to the reader to verify that the trace inner product satisfies the inner product conditions. Let A and B be $[n \times n]$ matrices. Then,

$$\langle A, B \rangle = \text{trace}(A^* B)$$

Lets get back to deriving the optimum k rank approximation of T. Let the Dyadic form of T be,

$$T = s_1 u_1 v_1^* + s_2 u_2 v_2^* + \cdots + s_p u_p v_p^*$$

It only goes until p because T has rank p. Recall that T is a $[m \times n]$ matrix.

You should verify that the set of all matrices from $\mathbb{C}^m$ to $\mathbb{C}^n$ is a vector space of dimension mn. This is equivalent to showing that all matrices of dimension $[m \times n]$ form a vector space.

Since we're working in a finite mn vector space, we can define a basis for this vector space. In particular, we will define a basis around the Dyadic from of T. Let,

$$b_1 = u_1 v_1^*, b_2 = u_2 v_2^*, \cdots, b_p = u_p v_p^*$$

Observe that, when $i \neq j$,

$$\begin{aligned}
\langle b_i, b_j \rangle &= \text{trace}(b_i b_j) \\
&= \text{trace}((u_i v_i^*)^* u_j v_j) \\
&= \text{trace}(v_i u_i^* u_j v_j) \\
&= \begin{cases} 0, & i \neq j \text{ because } u_i \perp u_j \\ 1, & i = j \text{ because } u_i, v_i \text{ are orthonormal} \end{cases}
\end{aligned}$$

In other words $b_1, \ldots b_p$ are orthonormal vectors of the vector space. We will now extend $\{b_1, b_2, \ldots, b_p\}$ to a complete orthonormal basis[10] of the vector space.

$$\mathscr{B} = \{b_1, b_2, \ldots, b_p, b_{p+1}, \ldots b_{mn}\}$$

[10] Gram-Schmidt allows us to do this. How?

Without loss of generality, assume,

$$s_1 \geq s_2 \geq s_3 \geq \cdots \geq s_p$$

T with respect to $\mathscr{B}$ is,

$$T = s_1 b_1 + s_2 b_2 + \ldots s_p b_p$$

Let $\widetilde{T}$ be written with respect to $\mathscr{B}$ as,

$$\widetilde{T} = \alpha_1 b_1 + \alpha_2 b_2 + \ldots \alpha_p b_p + \alpha_{p+1} b_{p+1} + \cdots + \alpha_{mn} b_{mn}$$

Since we want $\widetilde{T}$ to be rank k, we know that there can be at most k non-zero α_is. In order to find the best approximation, we want to minimize,

$$\left\| T - \widetilde{T} \right\|_2^2$$

By Pythagoras Theorem,

$$\left\| T - \widetilde{T} \right\|_2^2 = |s_1 - \alpha_1|^2 + |s_2 - \alpha_2|^2 + \cdots |s_p - \alpha_p|^2 + |\alpha_{p+1}|^2 + \cdots + |\alpha_{mn}|^2$$

Since $s_1 \geq s_2 \geq s_3 \geq \cdots \geq s_p$ it follows that the best we can do is to match the first k α_is to the corresponding singular values. That is to say,

$$\widetilde{T} = s_1 u_1 v_1^* + \cdots + s_k u_k v_k^*$$

Written by Siddharth Iyer.