Indian Institute of Technology Bombay
Department of Electrical Engineering

Handout 2 EE 703 Digital Message Transmission
Lecturenotes 1 Aug 01, 2017

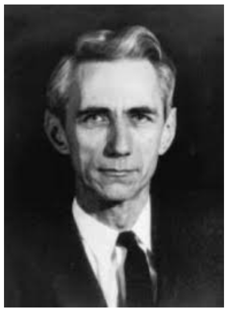
1 Introduction

You have *possibly* learned something about signals, done some implementation, and had some mathematics in your undergraduate days. This course essentially explains how they come together in equal proportions in modern digital communication systems.

While there is a lot to be written about the evolution of digital communication, we will make some short scribes here. Pioneering works from Maxwell, Hertz, Bell and others led to long distance communication becoming a reality, the synergy of theory and practice crucial to this evolution. It was Marconi's turn at the dawn of the past century to push communication out of their wired confinements. Unlike the analog telephones, Marconi's radio was about digital communication, for transmitting Morse codes through wireless telegraphic links. The less written part is that Marconi's effort in commercialization of communication techniques led to unprecedented developments, whose footprints still remain in the modern digital era. In particular, the development of diodes, triodes, transistors, ICs etc were all defining points in the race for communication superiority. History repeats itself!. Look around you, all sorts of communication gadgets springs to life, from tiny RF-ids to spohisticated smart phones, all striving to keep abreast of the latest communication demands. Furthermore, at the time of this writing, communication is driving the computational market too, from GPUs to cloud computers. It will be interesting at this point to read Chapter 1 of Wozencraft and Jacobs, it was written 50 years ago!

R. V. L. Hartley

H. Nyquist



C. E. Shannon

While Marconi is known as the *father* of radio, that attribute for digital communication unequivocally goes to Claude Shannon. Among the most influential figures of twentieth century, Shannon went several steps further from the initial stones laid by Hartley and Nyquist, the resulting theory had far reaching consequences. Unlike the other biggest development of twentieth century, which went *nuclear*, the communication revolution went the *ballistic* way. Above all, this was a giant leap in enabling communication theory to become a mathematical discipline, facilitating analysis and computation even before actual systems are practically rolled out. That we could be sure of communicating to the MARS shuttly, modulo mishaps, even before launching a test run is a testament to the resounding nature of communication theory. Following Shannon's foot-steps, in order to study or devise a communication technique for an environment, the following course of action is advocated.

- abstract the practical constraints to meaningful mathematical models for the signals and systems.
- solve the mathematical objective to obtain optimal communication schemes.
- translate and optimal schemes to practical design guidelines, which can be realized by existing apparatus.
- test the performance in operating environments, and refine the design and implementation.

Most communication systems that you see around have gone through such an evolutionary path, only to be named as a *revolutionary* technology later.

2 Some Notations

Let us list some notations that we will try to stick throughout the course. In particular, random variables will be denoted by capital letters, and their realizations will be written in small cases. For example, random variable Y takes values in the set $y \in \mathcal{Y}$. Here $\mathcal{Y}$ is the sample space of Y.

Unless otherwise specified, a vector will be considered as a column vector. Thus, a vector u is an $n \times 1$ vector, and u^T will be written as $\overrightarrow{u}$. In communication theory, often we have to dealy with complex numbers. So we will define a dot product of vectors in the complex field. For $n-$dimensional column vectors u and v, we define the dot product as

$$u \cdot v := u^\dagger v := \sum_{i=1}^{n} u_i^* v_i := (u^*)^T v := \langle u, v \rangle := \langle \overrightarrow{u}, \overrightarrow{v} \rangle,$$

where a^* is the complex conjugate of a, and $u^\dagger$ stands for u^{*T}, which is same as $\overrightarrow{u}^*$.

A collection of symbols $x_1, \cdots, x_n$ will be denoted as x^N. Generalizing this, Y^n denotes a random vector which can take the realization $y^n \in \mathcal{Y}^n$.

3 Digital Communication System

Let us demonstrate the first step in our design of a communication scheme. Fortunately, we can introduce a very generic model here, without even talking about any specific technology, see Figure 1.

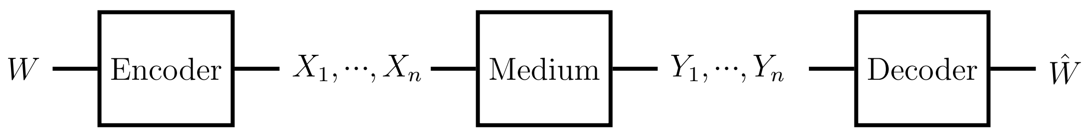


Figure 1: A Communication System

In the above figure, W is a random variable, which denotes a set of information symbols (usually bits), known as a message. The sample space for W can be taken as the index set $\{1, \cdots, M\}$, where M is some finite number[1]. It is usual to assume that the messages are uniformly distributed, i.e.

$$P(W = i) = \frac{1}{M}, 1 \le i \le M.$$

[1] The sigma field is not explicitly specified whenever it is understood by the context, the power set in this example

The encoder translates the message W to an n−dimensional vector $X_1, \cdots, X_n$ in an appropriate field. We will denote the vector $X_1, \cdots, X_n$ by X_1^n, or sometimes X^n, in this course. In particular, $X_i \in \mathcal{X}$ are symbols which are suitable to be transmitted over the given medium. Another way to visualize X^n is to consider them as samples of an underlying continuous time waveform sent over a medium, typically voltage/current waveforms in the baseband circuitry. We will opine more on this view later. A vector of values $Y^n := Y_1, \cdots, Y_n$ are received, and the decoder declares $\hat{W} \in \{\emptyset, 1, \cdots, M\}$ as the estimate of the transmitted message. This is a high level view of the communication systems considered in this course.

3.1 Encoder

The encoder can be described by a $M \times n$ matrix, where the rows are indexed by the messages in $\{1, \cdots, M\}$. This matrix is also known as a *codebook*, where each row is named as a *codeword*. Thus, each message is mapped to an n−dimensional codeword, this is demonstrated in Figure 2.

$W = 1$	$x_{11}, x_{12}, \cdots, x_{1n}$
$W = 2$	$x_{21}, x_{22}, \cdots, x_{2n}$
$\vdots$	
$W = M$	$x_{M1}, x_{M2}, \cdots, x_{Mn}$

Figure 2: Encoder Mapping

Often, the j^{th} codeword $x_{j1}, \cdots, x_{jn}$ will be conveniently written as $x^n(j)$.

3.2 Decoder

The observed output vector $Y^n \in \mathcal{Y}^n$ is also random. The randomness here is induced by two sources. Since W is random, so is the transmitted codeword $X^n(W)$, and the resulting output Y^n. Moreover, even when $M = 1$, the output can be random due to the randomness induced by the medium. The decoder $\hat{W}$ is a surjective mapping $\hat{W}(Y^n)$ from the space $\mathcal{Y}^n$ to $\{\emptyset, 1, \cdots, M\}$. The first element asserts the receiver's freedom to say 'I do not know', or declare an *erasure*. Nevertheless, in most cases we demand the decoder to declare one of the messages.

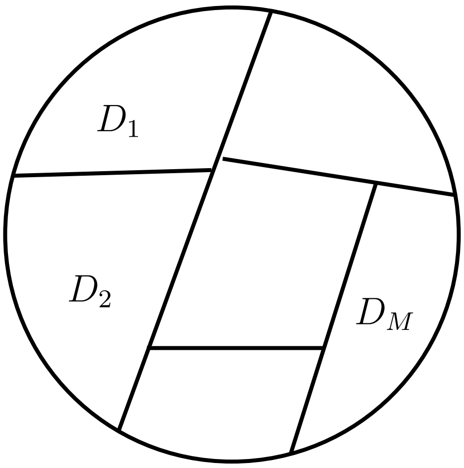


Figure 3: Decoder is a partition of $\mathcal{Y}^n$

In Figure 3, the region D_i represents the set of $Y^n \in \mathcal{Y}^n$ which are mapped to message i by the decoder, i.e.

$$D_i = \{y^n \in \mathcal{Y}^n | \hat{W}(y^n) = i\}.$$

More generally, a decoding rule is a partition of the space $\mathcal{Y}^n$ to M disjoint subsets.

We have left out the medium in the above description, modelling this appropriately is paramount to the success of our design. Down the line, we will treat this more rigorously. In fact, the system model for the medium can be very much application dependent. As an exercise, can you look around and identify three different communication systems that are used in our day to day life. Though the long haul optical cables and high speed LAN cables still adorn our offices and infrastructure, the ubiquitous spread of mobile devices has made digital communication almost synonymous with wireless communication.

3.3 Modelling the Medium

In spite of having many practical examples around us, we are going to introduce an abstract representation of a medium, that too of a peculiar type. In fact, the simple representation shown in Figure 4 may look almost trivial in the first sight.

$$X \longrightarrow \boxed{p(y|x)} \longrightarrow Y$$

Figure 4: A Memoryless Channel

The above representation is that of a memoryless channel, where the input symbol $x \in \mathcal{X}$ is mapped to one of the output symbols $y \in \mathcal{Y}$, with probability $p(y|x)$. Thus, the medium is specified by a collection of probability laws $p(y|x)$, one for each $x \in \mathcal{X}$. Notice that we did not insist $\mathcal{X}$ or $\mathcal{Y}$ being scalars, or even complex valued. This gives us enough flexibility as a generic model suitable for communication theoretic analysis. If you are not familiar with probability, don't worry, we can make the representation even *shorter*, and depict the communication medium literally as a pipe from X to Y, as in Figure 5.

Figure 5: A point to point link

Such a representation fits systems where the output takes the form $Y = f(X, Z)$, where Z is some randomness (often noise) independent of the transmissions. The function $f(\cdot,\cdot)$ captures the effect of the transmitted symbol at the receiver. In wireless and wireline systems, it often makes sense to assume that $f(\cdot,\cdot)$ is a linear function of the arguments.

Let us now demonstrate the components of a communication system using an example, albeit a futuristic one, i.e. something that is yet in the developing phase. The idea of *interference cancellation* is among the latest developments in digital communication.

4 Interference Cancellation

Interference is something we all worry about. While it is unclear how to model this, in one sentence, '*can we have several simultaneous communications over a shared medium*'. By appealing to physical laws, we know that simultaneous transmissions will cause superposition of the EM waves at the receiving terminal. Since we are yet to introduce the details of the medium, our immediate approach is to model the superposition of transmissions by an appropriate graphical representation, which can be visualized as an extension of Figure 5 to several communication links.

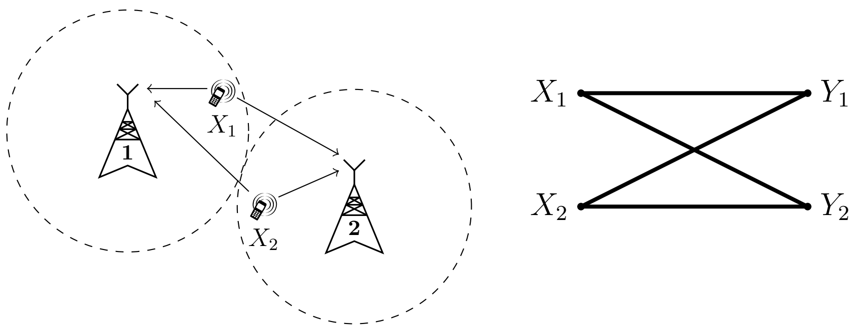


Figure 6: Cellular Users and Interference Graph

Imagine a cellular infrastructure with a frequency reuse factor of unity. Thus, the transmitter-receiver pairs in the neighboring cells may interfere with each other, as depicted by the graph in Figure 6. Transmitter $i \in \{1, 2\}$ emits the scalar symbols $X_i \in \mathbb{C}$. In our notation the transmitted vector for n consecutive transmission instants from user i is $X_{i1}, X_{i2}, \cdots, X_{in} := X_{i1}^n$. The receiver observes

$$Y_{1i} = f_1(X_{1i}, X_{2i}) \text{ and } Y_{2i} = f_2(X_{1i}, X_{2i}),\ 1 \le i \le n, \tag{1}$$

where the functions $f_1(\cdot,\cdot)$ and $f_2(\cdot,\cdot)$ model appropriate superpositions (linear). For example, we use the popular choice of $f_k(x, y) = \alpha_{1k}x + \alpha_{2k}y$, $k = 1, 2$ in our illustrations below. The coefficients α_{ik} are called the fading coefficients, usually assumed to be complex scalars.

Notice that we excluded any additive noise in (1), this is purely to illustrate the concept of interference management, we can later *add noise in our discussions.* Also let us somewhat naively assume that interference results in collision and data loss. How do we communicate in this situation? Sounds familiar!, imagine so many people discussing with their respective counterparts across a conference table, or a crowded party room. Such interference is also the subject of user management within a cell. In such situations, the good old TDMA (used in GSM), and the more recent CDMA come to our rescue. For instructive purposes, let us go through some abstract details.

4.0.1 TDMA

The essential idea here is to orthogonalize the communication resources in time. i.e. the users take turn in transmitting. Suppose $d_{ij}, j \geq 1$ is the data available at user i. In a simple model where the two users are given alternate transmission instants, user 1 transmits $X_{1j}, j \geq 1 = \{d_{11}, 0, d_{12}, 0, \cdots\}$, while user 2 sends $\{X_{2j,j\geq 1}\} = \{0, d_{21}, 0, d_{22}, \cdots\}$, where the zero symbol stands for *no transmission*. For simplicity, assume that the data symbols are real, though our discussion easily extends to complex values.

Observe that in order to send the data symbol d_{1i}, user 1 multiplies the symbol d_{1i} with the row vector $\vec{u}_1 = [1, 0]$, and schedules the resulting vector for the next two transmissions. In particular, we can write

$$\begin{bmatrix} X_{11} & X_{12} \end{bmatrix} = d_{11}\,\overrightarrow{u_1} \text{ and } \begin{bmatrix} X_{21}\, X_{22} \end{bmatrix} = d_{21}\,\overrightarrow{u_2},$$

where $u_2 = [0, 1]$. Thus, the vector $\overrightarrow{u_1}$ converts a single data symbols to a beam (vector) of values. Hence it is also known as a *beamformer*. Similarly, $\overrightarrow{u_2}$ is the beamformer for user 2. The received values at the users in two instants are

$$\begin{aligned} \begin{bmatrix} Y_{11}\, Y_{12} \end{bmatrix} &= \alpha_{11} d_{11} \overrightarrow{u_1} + \alpha_{21} d_{21} \overrightarrow{u_2} \\ \begin{bmatrix} Y_{21}\, Y_{22} \end{bmatrix} &= \alpha_{12} d_{11} \overrightarrow{u_1} + \alpha_{22} d_{21} \overrightarrow{u_2}. \end{aligned}$$

In order to get d_{11} back at user 1, we first combine the elements of Y_{11} and Y_{12}. This can be achieved by taking dot product with an appropriate weight vector $\vec{v} = [v_{11}\ v_{12}]$. Thus,

$$\begin{bmatrix} Y_{11}^* & Y_{12}^* \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = v_{11} Y_{11}^* + v_{12} Y_{12}^*.$$

The vector $\vec{v}$ is also called a combiner. It will also be called a zero-forcer if all other data than the intended one is cancelled at the output of the encoder. It is easy to see that the choice of $\overrightarrow{u_1} = \overrightarrow{v_1}$ and $\overrightarrow{u_2} = \overrightarrow{v_2}$ is sufficient for zero-forcing. In other words

$$U = \begin{bmatrix} \overrightarrow{u_1} \\ \overrightarrow{u_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } V = \begin{bmatrix} \overrightarrow{v_1} \\ \overrightarrow{v_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2}$$

will guarantee that the data symbols are transmitted without interference from the other user.

In general, we call the matrix U with rows $\overrightarrow{u_i}, 1 \leq i \leq n$ as a beamforming matrix. Similarly V with rows $\overrightarrow{v_i}, 1 \leq i \leq n$ is called the zero-forcing matrix. To illustrate, consider n users operating in TDMA mode. We can collect n symbols at the output of receiver k as a vector $\overrightarrow{y_k}$, given by

$$\overrightarrow{y_k} = \sum_{i=1}^{n} \alpha_{ik} d_i \overrightarrow{u_i}. \tag{3}$$

With $U = V = I_n$, we can have interference free operations, since $\langle \overrightarrow{y_k}, \overrightarrow{v_k} \rangle = \alpha_{kk}^* d_k$.

Exercise 1. *Verify that* $\langle \overrightarrow{y_k}, \overrightarrow{v_k} \rangle = \alpha_{kk}^* d_k$ *when* $U = V = I$.

4.1 CDMA

In the beamforming and zero-forcing employed in the previous section, the key property ensuring interference free transmission is that

$$\langle \overrightarrow{u_k}, \overrightarrow{v_i} \rangle = \delta_{k,i}.$$

Generalizing this, we can pick any orthonormal $n \times n$ matrix as U, and then take $V = U$. Recall that $\overrightarrow{u_i}$ (i^{th} row of U) as the beamforming vector at transmitter i, and $\overrightarrow{v_j}$ as the zeroforcer at receiver j. From (3), we have

$$\langle \overrightarrow{y_j}, \overrightarrow{v_j} \rangle = \sum_{i=1}^{n} \alpha^*_{ij}\, d_i \,\langle \overrightarrow{u_i}, \overrightarrow{v_j} \rangle = \alpha^*_{jj}\, d_j.$$

Notice that any unitary U is good enough, and the choice $U = I$ does indeed give TDMA. A popular technique which designs an orthogonal U using values from the set $\{-1, +1\}$ is called CDMA or plain CDMA.

4.2 Interference Alignment

Techniques like TDMA/CDMA takes a pessimistic view of the network, in the sense that each transmission is expected to create interference at all other receivers. Orthogonalizing n users require n transmissions in such cases, yielding a transmission efficiency of 1 symbol per user in every n transmissions. However, in practice, the network topology is often not that fully connected to warrant such an extreme pessimistic treatment. For example, consider the interference topology depicted in Figure 7.

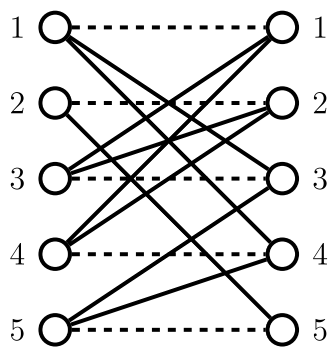


Figure 7: Interference Graph

In the interference graph shown, there are five transmitter receiver pairs participating in communication. Transmitter i intends to communicate to receiver i, shown by the dashed line. The solid lines represent the additive interference structure. Thus, user 1 causes additive disturbance at receivers 3 as well as 4. The link coefficients α_{ij} are taken to be identically unity for a simple exposition.

While TDMA/CDMA will achieve an efficiency of $\frac{1}{5}$ data symbol per transmission, better efficiencies are feasible in the above network. For example, users 1 and 2 can transmit simultaneously without interference. However, what more can be done is slightly unclear at this point. Let us build a more formal mechanism to analyze this model. To this end, let U be a $m \times n$ beamforming matrix. Thus, the beamformer at transmitter i is $\overrightarrow{u_i}$, which is $m-$ dimensional. Our intention is get m as small as possible. By collecting m samples as a vector at receiver i,

$$\overrightarrow{y_i} = \sum_{j \in \mathcal{A}_i} d_j \overrightarrow{u_j},$$

where $\mathcal{A}_i$ is the set of transmitters which are connected to receiver i. After combining or zeroforcing

$$\begin{aligned}\langle \overrightarrow{y_i}, \overrightarrow{v_i} \rangle &= \langle \sum_{j \in \mathcal{A}_i} d_j \overrightarrow{u_j}, \overrightarrow{v_i} \rangle \\ &= d_i \langle u_i, v_i \rangle + \sum_{j \in \mathcal{A}_i, j \neq i} d_j \langle \overrightarrow{u_j}, \overrightarrow{v_i} \rangle.\end{aligned}$$

Clearly, interference free operation can be achieved if we design beamformers $\overrightarrow{u_i}, 1 \le i \le n$ and combiners $\overrightarrow{v_j}, 1 \le j \le n$ such that

$$\langle \overrightarrow{u_j}, \overrightarrow{v_i} \rangle = \delta_{i,j} \text{ for } j \in A_i. \tag{4}$$

Notice that (4) will imply

$$UV^\dagger = \begin{bmatrix} \overrightarrow{u_1} \\ \overrightarrow{u_2} \\ \cdot \\ \cdots \\ \overrightarrow{u_n} \end{bmatrix} \begin{bmatrix} v_1^\dagger & v_2^\dagger & \cdots & v_n^\dagger \end{bmatrix} = \begin{bmatrix} 1 & \times & 0 & 0 & \times \\ \times & 1 & \times & \times & 0 \\ 0 & 0 & 1 & \times & \times \\ 0 & 0 & \times & 1 & \times \\ \times & \times & 0 & 0 & 1 \end{bmatrix}, \tag{5}$$

where $\times$ inside the matrix stands for a *don't care* condition. In other words, we are free to fill those with any values that we wish, a reminiscent of the so called Matrix Completion problem. In this particular example, we take an easy way out by filling all the don't care values by 1 in the first 4 columns. Fortunately, this makes two of the columns to repeat, and taking the fifth column as the difference of the second and third column only changes the don't care values. We can then write,

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 1 & -1 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 \end{bmatrix},$$

which gives a rank-2 decomposition. The interpretation of this from the transmission side is as follows. Users 1, 2 and 5 send their respective data symbols in every odd time-slots. Even time-slots are occupied by users 2, 3 and 5. By multiplying by $\overrightarrow{v_i}$ at receiver i data symbol d_i can be recovered.

Sources: For details of Interference alignment please refer, S.A.Jafar, Topological Interference Management Through Index Coding, Trans. Information Theory, 2011. Some of the material is from a recent tutorial conducted at IIT Bombay by Babak Hassibi, Caltech, 2017

5 Matrix Decomposition

In the above interference management problem, there were two aspects that we used in getting the decomposition. The first was the matrix completion problem, while maintaining a low rank to the matrix. The second was in finding a decomposition. We elaborate more on the latter part now. In particular we will learn the singular value decomposition (SVD).

5.1 SVD

Theorem 1. *A $m \times n$ matrix A can be decomposed as $A = U\Delta V^\dagger$, where U and V are unitary and Δ in an $m \times n$ diagonal matrix.*

Before we embark of the proof, the reader should be reassured that the proof uses nothing but some elementary techniques from linear algebra. In particular Eigen vectors and Eigen values of the matrix $AA^\dagger$ play key role. Recall that x is an Eigen-vector corresponding to Eigen value λ of $AA^\dagger$ if $||x|| = 1$ and

$$Ax = \lambda x.$$

We recapitulate some essential properties in the following three lemmas.

Lemma 1. *All the Eigen values of $AA^\dagger$ are non-negative, and the Eigen vectors form an orthonormal set.*

Proof. Notice that $x^\dagger AA^\dagger x = x^\dagger \lambda x = \lambda ||x||^2 = \lambda$. On the other hand, $x^\dagger AA^\dagger x = ||A^\dagger x||^2 \geq 0$, being the norm of a vector. This proves the first assertion. For the second take $\lambda_1 \neq \lambda_2$ as two Eigen values, with x and y the corresponding Eigen vectors. It is easy to see that

$$y^\dagger AA^\dagger x = \left(AA^\dagger y\right)^\dagger x.$$

The LHS above is nothing but $y^\dagger \lambda_1 x = \lambda_1 y^\dagger x$, whereas the RHS is $\lambda_2 y^\dagger x$. For equality $y^\dagger x = 0$, which is the intended result.

For the case where $\lambda_1 = \lambda_2 =, \cdots, = \lambda_l$, and other Eigen values different, clearly $u_1, \cdots, u_l$ will be orthogonal to all other Eigen vectors u_i with $\lambda_i \neq \lambda_1$. Thus $u_1, \cdots, u_l$ will span a sub-space. We will choose the orthonormal basis of this sub-space as the Eigen vectors. □

Let us now write $[\lambda_1, \lambda_2, \cdots, \lambda_r, 0, \cdots, 0]$ in descending order as the ordered Eigen values of $AA^\dagger$, and let $U = [u_1, u_2, \cdots, u_m]$ denote the corresponding Eigen vectors. Here r is called the rank of the matrix A with $r \leq \min\{m, n\}$.

Looks like we are being partial to $AA^\dagger$. To change that perception, we can collect the ordered Eigen values of $A^\dagger A$ as $[\lambda'_1, \cdots, \lambda'_r, 0, \cdots, 0]$ and let the corresponding Eigen vectors be $V = [v_1, \cdots, v_n]$. By the above arguments V contains an orthonormal set of vectors as well.

Lemma 2. *The matrices $AA^\dagger$ and $A^\dagger A$ have the same non-zero Eigen values, and we can take $v_j = (\sqrt{\lambda_j})^{-1} A^\dagger u_j$.*

Proof. By definition of Eigen vector u_i

$$A^\dagger AA^\dagger u_i = A^\dagger \lambda_i u_i = \lambda_i A^\dagger u_i.$$

Thus by taking $v_i = \frac{A^\dagger u_i}{||A^\dagger u_i||}$, we get

$$A^\dagger A v_i = \lambda_i v_i,$$

thus proving both the statements, since $||Au_i||^2 = u_i^\dagger AA^\dagger u_i = \lambda_i$. □

The following lemma is now straightforward.

Lemma 3.

$$u_i^\dagger A v_j = \sqrt{\lambda_j}\delta_{i,j},$$

where $\delta_{\{.\}}$ is the Kronecker delta.

Proof. By Lemma 2, we have $u_i^\dagger A v_j = (A^\dagger u_i)^\dagger v_j = \sqrt{\lambda}\, v_i^\dagger v_j = \sqrt{\lambda}\delta_{i,j}$. $\square$

Now, in order to obtain the proof for SVD, let us compute $U^\dagger AV$ with U and V defined as above. Assume $m \geq n$ for simplicity.

$$\begin{aligned} U^\dagger AV &= U^\dagger A\left[v_1, \cdots, v_n\right] && (6)\\ &= U^\dagger\left[Av_1, \cdots, Av_n\right] && (7)\\ &= \begin{bmatrix} u_1^\dagger \\ u_2^\dagger \\ . \\ . \\ u_m^\dagger \end{bmatrix} \left[Av_1, \cdots, Av_n\right] && (8) \end{aligned}$$

The entry at index (i,j) of this matrix is nothing but $u_i^\dagger A v_j = \sqrt{\lambda_j}\delta_{i,j}$. Thus

$$U^\dagger AV = \Delta,$$

where

$$\Delta = \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & \cdots & 0 \\ & & & & \\ 0 & 0 & \cdots & 0 & \sqrt{\lambda_n} \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Notice that $U^\dagger AV = \Delta$ will imply that $A = U^\dagger \Delta V^\dagger$, since U and V are unitary matrices, i.e. $U^\dagger U = UU^\dagger = \mathbb{I}$, and $V^\dagger V = VV^\dagger = \mathbb{I}$. Notice however that $UU^\dagger$ and $VV^\dagger$ may be of different dimensions.