Lecture Notes for Math 104: Fall 2010 (Extremely Rough Draft)

Jacob Bernstein

CHAPTER 1

Introduction

These notes are a (slightly) edited version of the material that I presented in class. I emphasize that there has been essential no proof reading of this material. I have tried to keep these notes in the order in which I presented the material in class. However, in some cases I have added some additional material where as needed (though I have tried to highlight when this has been done order to avoid confusion) and omitted other off hand remarks. In particular any section with NIC (for not in class) in in the section heading was not something discussed in class. Note the numbering of theorems and results is not consistent with any numbering scheme used in class and I hope this does not cause undo confusion. You should view this as supplemental to the course texts. As the proof reading has been pretty much non-existant, I would also greatly appreciate any comments and do let me know about mathematical errors or typos you may find.

CHAPTER 2

First Lecture

We begin by recalling some basic definitions and facts. In the class treat both real $\mathbb{R}$ and complex $\mathbb{C}$ linear algebra. The former is probably more familiar and "physical" though the latter is more general (and becomes necessary for the correct statement of some theorems). Recall $\mathbb{C}$ is the complex numbers, we will review these next lecture. Most aspects of linear algebra that we discuss won't really depend on whether we work over $\mathbb{R}$ or $\mathbb{C}$. However, as Trefethen and Bau usually works over $\mathbb{C}$ we will do so as well. Unless otherwise stated everything we will do holds over $\mathbb{R}$ and we will often illustrate concepts over $\mathbb{R}$ as the geometry. There will be some cases were the distinction matters and we will point these out! Recall that the complex numbers are just the real numbers with an additional "imaginary" number I which we treat like a regular number except $I^2 = -1$. (Note we don't use i as we want to reserve that for other purposes).

1. Vectors

For us a *vector* will be a n-tuple of real or complex numbers. i.e. $\mathbf{v}$ is can be represented by $(v_1, \ldots, v_n)$ for $v_i \in \mathbb{C}$). We will then say $\mathbf{v} \in \mathbb{C}^n$. If all the v_i are real then we have $\mathbf{v} \in \mathbb{R}^n$ is the set of n-dimensional vectors over the reals. We will also say $\mathbf{v}$ is a *real vector*. Rather than write vectors as n-tubles we will always write them as columns.

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

While the difference between tuples and columns is notational there is a more fundamental difference between the two. An n-tuple should really be though of as just an (ordered) list of numbers while the vector has some additional geometric and algebraic meaning. It is a subtle point, but conceptually important to make this distinction.

As we know we can add vectors. Let $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$ with

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

then

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{bmatrix}$$

Geometrically, if we have (say) $\mathbb{R}$ vectors this corresponds to laying the vectors end to end. Notice $\mathbf{v}+\mathbf{w}=\mathbf{w}+\mathbf{v}$ and we see this geometrically as well.

We may also multiply vectors by a scalar number $\lambda \in \mathbb{C}$

$$\lambda \mathbf{v} := \begin{bmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{bmatrix}$$

Geometrically, for $\mathbb{R}$ vectors this corresponds to stetching the vector by a factor of $|\lambda|$ and reversing direction if $\lambda < 0$. We can also multiply on the left side by a scalar and

$$\mathbf{v}\lambda := \lambda \mathbf{v}.$$

Finally, scalar multiplication interactions with sums in the usual way, namely:

$$\lambda(\mathbf{v}+\mathbf{w}) = \lambda \mathbf{v} + \lambda \mathbf{w}.$$

Let us recall some important vectors:

$$0 := \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

the additive identity. This geometrically this corresponds to "doing nothing" and $0+\mathbf{v}=\mathbf{v}+0=\mathbf{v}$ and $\lambda 0 = 0$. The "standard basis vectors"

$$\mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad i\text{th slot}$$

If $n=3$ and we consider real vectors then $\mathbf{e}_1 = \mathbf{i}$, $\mathbf{e}_2 = \mathbf{j}$, $\mathbf{e}_3 = \mathbf{k}$. Note we can always write a vector $\mathbf{v} \in \mathbb{C}^n$ (uniquely) as

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i = \begin{bmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{bmatrix}$$

For $v_i \in \mathbb{C}$ (or $\in \mathbb{R}$ if $\mathbf{v}$ is a real vector). That is the $\mathbf{e}_i$ are a *basis* of $\mathbb{C}^n$ (we will come back to this latter)..

2. Linear Transformations

Vectors are the basic object in linear algebra but are not all that interesting in and of themselves. More interesting is to study transformations of vectors. In the context of linear algebra we restrict attention to *Linear Transformations*. That is transformations that respect the linear structure (i.e. addition and scalar multiplication). More precisely a linear transformation is a function

$$T : \mathbb{C}^n \to \mathbb{C}^m$$

so that $T(\mathbf{v}+\mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$ and $T(\lambda\mathbf{v}) = \lambda T(\mathbf{v})$. This class will mostly consist of studying linear transformations. However, just as we think of a vector as a concrete list, we will also think of a linear transformation as a concrete object– namely as an array of numers called a matrix.

Recall a complex valued matrix is just an $m \times n$ array of complex numbers (when all the entries are real we say it is *real matrix*):

$$A := \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & | \cdots & | & \mathbf{a}_n \end{bmatrix}$$

where here $\mathbf{a}_j$ are columns:

$$\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

This suggests we think of the columns as vectors, which we will very often do. We will write $A \in \mathbb{C}^{m\times n}$ to say that A is an $m \times n$ matrix with complex entries and $A \in \mathbb{R}^{m\times n}$ if the entries are real.

How do we go between T and the matrix A that represents it? The easiest way is to see what T does to the standard basis. As $T(\mathbf{e}_j)$ is a vector in $\mathbb{C}^m$ we can expand it in standard basis vectors $\mathbf{e}_i$ of $\mathbb{C}^m$ as:

$$T(\mathbf{e}_j) = \mathbf{a}_i = \sum_{i=1}^{m} a_{ij}\mathbf{e}_i$$

but that is just

$$T(\mathbf{e}_j) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

Where this is standard matrix multiplication (see below). Notice that the ith column of A just the image of $\mathbf{e}_i$. By linearity then for arbitrary $\mathbf{v} = \sum_{j=1}^{n} v_j\mathbf{e}_j$ one has

$$T(\mathbf{v}) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}v_j\mathbf{e}_i$$

That is

$$T(\mathbf{v}) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_m \end{bmatrix}$$

More compactly one can write:

$$T(\mathbf{v}) = \begin{bmatrix}\mathbf{a}_1 & | & \cdots & | & \mathbf{a}_n\end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_m \end{bmatrix}$$
$$= v_1\mathbf{a}_1 + \ldots + v_n\mathbf{a}_n$$

here the $\mathbf{a}_j = T(\mathbf{e}_j)$ are the columns of the matrix.

As with vectors we will usually just talk directly about a $m \times n$ matrix but just as with should always remember that this is representing some sort of linear transformation.

3. Algebra of matrices

Recall that we can add matrices, multiply them by a scalar and multiply a $m \times n$ matrix on the left by a $n \times k$ matrix. Addition and multiplication by a scalar are unambiguous. But in case you're a little rusty: Set

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}.$$

Then

$$A + B = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix}$$

where $c_{ij} = a_{ij} + b_{ij}$. If $\lambda \in \mathbb{C}$ then

$$\lambda A = A\lambda = \begin{bmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & \ddots & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{bmatrix}$$

Slightly more complicated is matrix multiplication: Let B now be

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + \ldots + a_{1n}b_{n1} & \cdots & a_{11}b_{1k} + \ldots + a_{1n}b_{nk} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \ldots + a_{mn}b_{n1} & \cdots & a_{m1}b_{1k} + \ldots + a_{mn}b_{nk} \end{bmatrix}$$

or more compactly

$$AB = \begin{bmatrix} A\mathbf{b}_1 & | \cdots & | A\mathbf{b}_k \end{bmatrix}$$

here $\mathbf{b}_j$ are the columns of B. Notice that this (perhaps mysterious) formula for matrix multiplication comes from looking at the linear map that arises from the

composition of two linear maps. Notice also that if we let $C = AB$ and then expaning things out we see that the columns of C can be expressed as linear combinations of the columns of B.

Most of the usual algebraic rules apply except that matrix multiplication is *not* commutative. Just to list them, let $A, B \in \mathbb{C}^{m\times n}, C, D \in \mathbb{C}^{n\times k}, E \in \mathbb{C}^{k\times l}, \lambda \in \mathbb{C}$. Then

$$\begin{aligned} A + B &= B + A \\ \lambda A &= A\lambda \\ \lambda(A + B) &= \lambda A + \lambda B \\ A(\lambda C + D) &= \lambda AC + BD \\ (A + B)C &= AC + BC \\ (AC)D &= A(CD) \end{aligned}$$

Finally, we recall some important matrices. The first is the zero matrix. This is the matrix in each $\mathbb{C}^{m\times n}$ which has all zero entries and which we denote by 0. This corresponds to the linear transformation that sends everything to 0. If $A \in \mathbb{C}^{m\times n}, B \in \mathbb{C}^{l\times m}$ and $C \in \mathbb{C}^{n\times k}$ and $\lambda \in \mathbb{C}$ then

$$\begin{aligned} A + 0 = 0 + A &= A \\ \lambda 0 = 0\lambda &= 0 \\ B0 = 0 &= 0C. \end{aligned}$$

We also introduce the the identity matrix $I \in \mathbb{C}^{m\times m}$ (which we will also denote by Id). This is the matrix

$$I = \begin{bmatrix} |\mathbf{e}_1| & \cdots & |\mathbf{e}_i| & \cdots & |\mathbf{e}_n| \end{bmatrix}$$

whose columns are the standard basis. This matrix corresponds to the transformation that preserves all vectors. For $B \in \mathbb{C}^{l\times m}$ and $C \in \mathbb{C}^{m\times k}$ we have

$$\begin{aligned} BI &= B \\ IC &= C \end{aligned}$$

CHAPTER 3

Second Lecture

In this lecture we review some of the properties of complex numbers.

1. Complex Numbers

Let us look at the following equation:

$$x^2 + 1 = 0 \tag{1.1}$$

Naively $x = \sqrt{-1}$ would seem to be a solution. However, for any real xR $x^2+1 \geq 1$ so this equation can never be solved over the reals. One of the great realizations in mathematics was that in fact there is a solution if we just expand our concept of numbers. Indeed, we can introduce an "imaginary" number I. This is not a real number and is just a symbol. However, by pretending it has all the algebraic properties of a usual number along with satisfying $I^2 = -1$, will lead to a consistent theory. Formally doing so will allow us to say (up to a sign) $I = \sqrt{-1}$.

In practice what this means is that we should be able to mulitply I by any real number a to get a new "imaginary" number Ia. The set of all such numbers is usually refered to as the set of (purely) "imaginary" numbers and is denoted $I\mathbb{R}$ (it is more standard to use i but I want to avoid confusion with indices). We can also add real and complex numbers to get new numbers that (usually) are neither real nor complex that is let $a, b \in \mathbb{R}$ and then $z = a + Ib$ is (for $a, b \neq 0$) neither real nor imaginary. We denote the set of all such numbers by $\mathbb{C}$ and call them complex numbers.

Since we seek an algebraically consistent set of numbers, we must be a little careful how we define various algebraic operations. To add complex numbers we have:

$$(a + Ib) + (c + Id) = (a + c) + I(b + d)$$

and to multiply we have

$$(a + Ib)(c + Id) = (a + Ib)c + (a + Ib)Ib = ac - bd + I(ad + bc).$$

One important operation that is new is complex conjugation. The idea here is to take a complex number $z = a + ib$ and associate its complex conjugate $\bar{z} = a - ib$. The reason we do this is then that $z\bar{z} = a^2 + b^2$ is then always real (and non-negative). Notice that if $z = a + Ib$ then

$$a = \Re z = \frac{z + \bar{z}}{2}$$

and

$$b = \Im z = \frac{z - \bar{z}}{2I}.$$

The fact that $z\bar{z}$ is a non-negative real number makes it tempting to think of it as a length. This we do and define the *modulus* of the complex number z to be

$$|z| = \sqrt{z\bar{z}}.$$

Notice if $z = a + Ib$ for $a, b \in \mathbb{R}$ then $|z| = \sqrt{a^2 + b^2}$. It is straightforward to see that $z = 0$ if and only if $|z| = 0$.

The complex conjugate also gives an unambiguous way to to divide a complex number by a non-zero complex number. Indeed,

$$\frac{z_1}{z_2} = \frac{z_1\bar{z}_2}{|z_2|^2}.$$

The right hand side consists of multiplication of two complex numbers and then division by a real number all of which is straightforward.

We introduced the complex numbers in order to find roots to $x^2+1 = 0$. Indeed, we can now check that I and $-I$ are the (only) two solutions of this equation. It turns out that once one has I all polynomials (even with complex coefficients) have a solution.

THEOREM 1.1. *(Fundamental Theorem of Algebra) Let*

$$p(x) = \sum_{i=0}^{n} a_i x^i$$

be a polynomial of order n (can consider $a_i \in \mathbb{C}$) with $a_n \neq 0$. Then $p(x) = 0$ has exactly n solutions (counting multiplicity) $z_1, \ldots, z_n$ in $\mathbb{C}$. That is

$$p(x) = a_n(x - z_1)(x - z_2)\cdots(x - z_n)$$

2. Geometry of Complex Numbers

Real numbers are usual represented as points on a line. What is the right way to think of representing complex numbers geometrically? Notice that $z = a + Ib$ is really just a pair of numbers (a, b). It can also be thought of as a vector:

$$\begin{bmatrix} a \\ b \end{bmatrix}.$$

That is the complex number z can be represented as a 2-dimensional real vector. Notice complex addition is then the same as vector addition, However, complex numbers can be multiplied and there is no clear way to interpret this for vectors.

This graphical representation gives another way to describe a complex number $z = a + Ib$. Namely, we can associate an angle θ and a radius $r \geq 0$ to z so that $z = r\cos\theta + i\sin\theta$. Explictily, $r = |z| = \sqrt{a^2 + b^2}$ and $\tan\theta = b/a$. The number θ is called the *argument* of z and is defined only up to 2π.

This representation is particlurly useful when combined with the imporant fact know as Euler's formula (see below). For $t \in \mathbb{R}$ one has:

$$e^{It} = \cos t + I\sin t$$

More generally, one has that

$$e^{(\mu+I\nu)t} = e^{\mu t}(\cos(\nu t) + I\sin(\nu t)) \tag{2.1}$$

One way to justify this formual is to note that it ensures that

$$\frac{d}{dt}e^{\lambda t} = \lambda e^{\lambda t}.$$

even when $\lambda \in \mathbb{C}$.

3. Linear Algebra of Complex Numbers

As we we say we can think of a complex number $z = a + Ib$ as a real vector

$$\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$$

We then have 1 corresponding to $\mathbf{e}_1$ and I corresponding to $\mathbf{e}_2$. Where $\mathbf{e}_1, \mathbf{e}_2$ is the standard basis of $\mathbb{R}^2$.

It turns out that many of the natural operations on z as a complex number can be interpreted as a linear transformation on $\mathbf{v}$. We will use this to illustrate some ideas from last time. Lets consider the map $z \to \bar{z}$ that is let C be the function so that $C(z) = \bar{z}$. On vectors this is the map:

$$C : \begin{bmatrix} a \\ b \end{bmatrix} \to \begin{bmatrix} a \\ -b \end{bmatrix}$$

One checks that

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

and so C yields a linear trasformation on $\mathbb{R}^2$ with matrix

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

Notice if we apply complex conjugation twice we get back where we started. This corresonds to the fact that the square of the associated matrix is the identity.

Fix a complex number $w = c + Id$ and consider the function $M_w(z) = wz$. As vectors this yields the map

$$M_w : \begin{bmatrix} a \\ b \end{bmatrix} \to \begin{bmatrix} ac - bd \\ ad + bc \end{bmatrix}$$

We can again check that

$$\begin{bmatrix} ac - bd \\ ad + bc \end{bmatrix} = \begin{bmatrix} c & -d \\ d & c \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

And hence M_w can be thought of a linear map on $\mathbb{R}^{2\times 2}$ with associated matrix:

$$\begin{bmatrix} c & -d \\ d & c \end{bmatrix}$$

The rules of complex multiplication imply that $M_{w_1+w_2} = M_{w_1} + M_{w_2}$. In particular, if A, A_1 and A_2 are the matrices in $\mathbb{R}^{2\times}$ corresponding to $M_{w_1+w_2}$, M_{w_1} and M_{w_2} (respectively) then one can verify that $A = A_1 + A_2$. Similarly, one has $M_{w_1}(M_{w_2}(z)) = M_{w_1 w_2}(z)$ and so if A, A_1 and A_2 are the matrices in $\mathbb{R}^{2\times}$ corresponding to $M_{w_1 w_2}$, M_{w_1} and M_{w_2} then one checks that $A = A_1 A_2$. An important consequence is then that we can encode a complex number $z = a + Ib$ as a 2×2 real matrix and all the complex algbra directly corresponds to matrix algebra.

Now consider Euler's formula for $w = c + Id$. That is write $w = re^{i\theta}$ with $r \geq 0$ and $\theta \in [0, 2\pi)$. As a consequence $M_w(z) = M_r(M_{e^{I\theta}}(z)) = M_{e^{I\theta}}(M_r(z))$ In particular

$$\begin{bmatrix} c & -d \\ d & c \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$$

Notice we we've broken our matrix into a simple diagonal matrix and a matrix that is also very simple (it is orthogonal). We will generalize this sort factoriztion for arbitrary matrices (where the diagonal will become upper triangular). This is called the QR factorization and is at the heart of all sorts of applications of linear algebra.

4. Functions of a complex variable: NIC

The algebraic properties of complex numbers allow one to define polynomial functions of a complex variable. That is for a polynomial

$$p(x) = \sum_{i=0}^{n} a_i x^i$$

it is clear what $p(z)$ means for $z = a + Ib$ a complex number. For more general functions the same result can be accomplished by using a taylor series expansion (when one exists). For instance if a function f has a taylor series expansion

$$f(x) = \sum_{i=0}^{\infty} a_i x^i$$

on for $|x| < R$ (here R is the radius of convergence). Then the function f makes sense at complex values z so that $|z| < R$ by setting

$$f(z) := \sum_{i=0}^{\infty} a_i z^i.$$

One has to of course check this sum converges in an appropriate sense, but that is beyond the scope of these notes.

The point is that this gives a rigorous justification for the Euler formula. Namely, the Taylor series expansion for e^t is

$$e^t = \sum_{n=0}^{\infty} \frac{t^n}{n!}$$

which has infinite radius of convergence. We then have

$$e^{It} = \sum_{n=0}^{\infty} \frac{(It)^n}{n!}$$

But $I^{4k} = 1, I^{4k+1} = I, I^{4k+2} = -1$ and $I^{4k+3} = -I$ so this gives

$$e^{It} = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{(2n)!} + I \sum_{n=1}^{\infty} \frac{(-1)^(n-1) t^{2n-1}}{(2n-1)!}$$

But the Taylor series expansions of of $\cos t$ and $\sin t$ are

$$\cos t = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{(2n)!} \quad \sin t \qquad\qquad = \sum_{n=1}^{\infty} \frac{(-1)^(n-1) t^{2n-1}}{(2n-1)!}$$

So

$$e^{It} = \cos t + I \sin t$$

as claimed. The formula for a general complex number can be verified in a similar manner.

CHAPTER 4

Third Lecture

We recall the definitions of basic linear algebra concepts such as spans of vectors, linear independence, basis vectos and so on. We will also translate these concepts into corresponding properties of matrices.

1. Basic Linear Algebra

Suppose we have a set of vectors $\mathbf{v}_1, \cdots, \mathbf{v}_k$ in $\mathbb{C}^n$. One very natural question to ask is can we write every vector as a linear combination of the $\mathbf{v}_j$? If yes, one further asks "how many" different ways are there to express the same vector in terms of the$\mathbf{v}_j$. If no, which vectors fail to be so expressable and how many are there.

To start making this precise we define the *span* of a set of vectors $\{\mathbf{v}_j\}$ to be the set of all vectors. $Span(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \left\{\mathbf{w} : \mathbf{w} = \sum_j c_j \mathbf{v}_j\right\}$. This is the largest set of vectors we form from the set $\{\mathbf{v}_j\}$ by only using linear algebra operations. We say the $\mathbf{v}_j$ are *linearly independent* if

$$c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k = 0 \iff 0 = c_1 = \ldots = c_k$$

Otherwise we say the $\mathbf{v}_j$ are linearly dependent.

A simple example: The vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

are linearly independent in $\mathbb{R}^3$. Their span is the plane $z = 0$. To show this rigrously note that:

$$a\mathbf{v}_1 + b\mathbf{v}_2 = (a+b)\mathbf{e}_1 + (a-b)\mathbf{e}_2 = \begin{bmatrix} a+b \\ a-b \\ 0 \end{bmatrix}$$

For this to equal 0 need $a + b = 0$ and $a - b = 0$. That is $a = 0$ and $b = 0$. Hence the vectors are linearly independent. Checking the spanning property is similar. Notice all this comes down to is solving ansystem of equations.

Another example: The vectors

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

are not linearly independent. Indeed, $\mathbf{v}_3 = 3\mathbf{v}_1 - 3\mathbf{v}_2$. However any pair of the vectors is linearly independent the span of all three vectors is a plane in $\mathbb{R}^3$.

A final example: In general any collection of standard basis vectors $\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}$ in $\mathbb{C}^n$ is linearly independent if $k = n$ then the set spans $\mathbb{C}^n$.

How do we determine if a given set of vectors spans some set? Is linearly independent? As we sketched in the example above this is really a question about systems of linear equations. Consequently, the best method to do this is to turn the question into an equivalent question about matrices. We will then answer the corresponding question for matrices. This will also give additional information. So how do we turn this into a question about matrices? The point is that taking the span span looks like taking lots of matrix multipications. That is let V be the $n \times k$ matrix:

$$V = \begin{pmatrix} \mathbf{v}_1 & | & \cdots & | & \mathbf{v}_k \end{pmatrix}$$

Then let

$$\mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

then

$$V\mathbf{c} = \sum_{j=1}^{k} c_j \mathbf{v}_j$$

I.e. $\mathbf{w}$ is in the span of the $\mathbf{v}_j$ if and only if there is a vector in $\mathbf{c} \in \mathbb{C}^k$ so that $\mathbf{w} = V\mathbf{c}$. We denote by $Range(V)$ or $R(V)$ the precisely the latter such set. That is let $A \in \mathbb{C}^{m \times n}$ be a matrix we define the range of A, $Range(A) = R(A) \subset \mathbb{C}^m$ by

$$R(A) = \{\mathbf{w} \in \mathbb{C}^m : \mathbf{w} = A\mathbf{c} \text{ for some } \mathbf{c} \in \mathbb{C}^n\}$$

We also refer to $R(A)$ this as the *Image* of A or the *Column space of* A. Notice the latter makes sense as $R(A)$ is the span of the columns of A.

Similarly, we can use V to see when the $\mathbf{v}_j$ are linearly independent. Again we have the $\mathbf{v}_j$ linearly independent if

$$0 = \sum_j c_j \mathbf{v}_j = V\mathbf{c} \iff \mathbf{c} = 0$$

That is $\{\mathbf{v}_j\}$ is linearly independent if and only if the only vectors that A multiplies against to give 0 is the zero vector. For $A \in \mathbb{C}^{m \times n}$, we define $Null(A) = N(A) \subset \mathbb{C}^n$, the null space of A, to be the set of vectors

$$N(A) = \{\mathbf{v} \in \mathbb{C}^n : A\mathbf{v} = 0\}$$

This is sometimes refered to as the *kernel* of A. Hence we see that the $\mathbf{v}_j$ are linearly indepependent when and only when $N(V) = \{0\}$.

Having transformed the problem to a question about matrices we need to discuss how to use this to actually solve the problem. The main way to do this is to use Gaussian elimination. We'll review this algorithm next lecture.

2. Spaces of Vectors and Basis Vectors

In order to make some of the preceding clearer, we introduce some further mathematical definitions as well as state some important facts. The later will be proved in a couple of lectures after we have some important tools.

For a matrix $A \in \mathbb{C}^{m \times n}$ we have defined the range of A, $R(A)$ as a subset of $\mathbb{C}^m$ and the null space, $N(A)$ as a subset of $\mathbb{C}^m$. These sets are well behaved with respect to the operations of linear algebra. More precisely, we say a set of vectors

$E \subset \mathbb{C}^n$ is a *vector space* (or *vector sub-space*) if for any pair of vectors $\mathbf{v}, \mathbf{w} \in E$ and any scalar $\lambda \in \mathbb{C}$ one has $\lambda\mathbf{v} \in E$ and $\mathbf{v} + \mathbf{w} \in E$. That is E is closed uner the operations of linear algebra. Notice by taking $\lambda = 0$ we must alwasy have $0 \in E$. Note that $\mathbb{C}^n$ is a vector space as is $\{0\}$ the set consisting only of the zero vector $0 \in \mathbb{C}^n$.

Notice that the span of any set of vectors $vV_i \in \mathbb{C}^n$, $1 \leq i \leq k$ is a vector space. To see this let $\mathbf{v}, \mathbf{w} \in Span(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ then $\mathbf{v} = \sum_{i=1}^k c_i\mathbf{v}_i, \mathbf{w} = \sum_{i=1}^k d_i\mathbf{v}_i$. Then using the algebraic properties of vectors one has $\mathbf{v} + \mathbf{w} = \sum_{i=1}^k (c_i + d_i)\mathbf{v}_i$ which is then clearly in the span. A similar argument shows $\lambda\mathbf{v}$ is in the span. In particular, as the range space of A is the span of the columns of A, $R(A) \subset \mathbb{C}^m$ is a vector space. The null space of A, $N(A) \subset \mathbb{C}^n$ is also a vector space. To see this let $\mathbf{v}, \mathbf{w} \in N(A)$. Then $A\mathbf{v} = A\mathbf{w} = 0$. Now $A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w} = 0$ and $A(\lambda\mathbf{v}) = \lambda(A\mathbf{v}) = \lambda 0 = 0$, hence $\mathbf{v} + \mathbf{w} \in N(A)$ and $\lambda\mathbf{v} \in N(A)$.

For a given vector space E we say that a set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in E$ are a *basis* of E if $E = Span(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ and the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent. An easy example is that that the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis of $\mathbb{C}^n$. One important consequence of $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ being a basis of $E \subset \mathbb{C}^n$ is that any vector $\mathbf{v} \in E$ can be expressed uniquely in terms of the $\mathbf{v}_i$ that is

$$\mathbf{v} = \sum_{i=1}^k c_i\mathbf{v}_i = V\mathbf{c}$$

where here $V \in \mathbb{C}^{n\times k}$ is the matrix with columns the $\mathbf{v}_i$ and $\mathbf{c} \in \mathbb{C}^k$ are the *coefficients* of $\mathbf{v}$ with respect to the basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$. It is a simple excercise to see that $N(V) = \{0\}$ and $R(A) = E$. In particular, if we want to find the coefficients $\mathbf{c}$ for a given vector $\mathbf{v}$ we must solve the equation:

$$V\mathbf{c} = \mathbf{v}.$$

Which consists of n equations in k unknowns.

CHAPTER 5

Fourth Lecture

In this lecture we reviewed Gaussian elimination. We focused on the difference between row operations and column operations and how those could be used to determine different information about (respectively) the null space of a matrix and the column space.

1. Elementary Row and Column Operations

The basic tool we will use are a certain set of operations on the rows and columns of a matrix. These will preserve important features of the matrix while also simplifying the matrix.

Fix a matrix $A \in \mathbb{C}^{m\times n}$. We've already seen how to write A in terms of its columns:

$$A = \begin{pmatrix} \mathbf{a}_1 & | & \dots & | & \mathbf{a}_n \end{pmatrix}$$

here $\mathbf{a}_j$ is vector in $\mathbb{C}^m$ or a $m \times 1$ matrix. it is also helpful to write it in terms of its rows

$$A = \begin{pmatrix} c_1 \\ - \\ \vdots \\ - \\ c_m \end{pmatrix}$$

where here the c_j is a $1 \times n$ matrix (not a vector!).

Starting with A we define an elementary column operation to be one of the following operations: a) Swapping two columns of A, b) scaling the first column by a non-zero scaler and c) adding the second column to the first column. In general we will be iteratively applying a sequence such operations to A.

More precisely, we take and produce a new $m \times n$ matrix A' by doing one of the preceding operations as follows:

$$A = \begin{pmatrix} \mathbf{a}_1 & |\dots| & \mathbf{a}_i & |\dots| & \mathbf{a}_j & |\dots| & \mathbf{a}_n \end{pmatrix}$$

under the swapping operation, a), goes to

$$A' = \begin{pmatrix} \mathbf{a}_1 & |\dots| & \mathbf{a}_j & |\dots| & \mathbf{a}_i & |\dots| & \mathbf{a}_n \end{pmatrix}.$$

Here we are free to choose any $1 \leq i < j \leq n$ we like. Ander the scaling operation, b), one has A going to

$$A' = \begin{pmatrix} \lambda\mathbf{a}_1 & |\dots| & \mathbf{a}_i & |\dots| & \mathbf{a}_j & |\dots| & \mathbf{a}_n \end{pmatrix}$$

where here $\lambda \neq 0$ is a scaler in $\mathbb{C}$. Finally, under the addition operation, c), A goes to

$$A' = \begin{pmatrix} \mathbf{a}_1 + \mathbf{a}_2 & |\mathbf{a}_2| & \dots| & \mathbf{a}_i & |\dots| & \mathbf{a}_j & |\dots| & \mathbf{a}_n \end{pmatrix}.$$

Notice that by combining the swapping operation with the scaling operation, one obtains an operation given by scaling any column by a non-zero scalar. Similarly, by combining the swapping operation with the addition operation one gets an operation wherein any column may be added to any other. This larger set of operations is often refered to as *elementary column operations*.

The nice feature of elementary column operations is that they preserve the range space of a matrix.

THEOREM 1.1. *Let $A \in \mathbb{C}^{m\times n}$. If A' is obtained from A by a (finite) sequence of elementary column operations then $R(A) = R(A')$*

REMARK 1.2. In general $N(A) \neq N(A')$, though it is true that $dimN(A) = dimN(A')$.

PROOF. We verify this only when A' is obtained from A by one elementary column operation. The theorem then follows by induction. Lets first verify that $R(A') = R(A)$ when A' is obtained from A by swapping the ith and jth column. Let $\mathbf{v} \in R(A)$ then there is a $\mathbf{w} \in \mathbb{C}^n$ so that $\mathbf{v} = A\mathbf{w}$. We may write $\mathbf{w} = w_1\mathbf{e}_1 + \cdots + w_i\mathbf{e}_i + \cdots + w_j\mathbf{e}_j + \cdots + w_n\mathbf{e}_n$ where $\mathbf{e}_k$ is the kth standard basis vector. Now let A' be obtained from A by swapping the ith and jth columns. If we set $\mathbf{w}' = \mathbf{w} - w_i\mathbf{e}_i - w_j\mathbf{e}_j + w_j\mathbf{e}_i + w_i\mathbf{e}_j$ then one verifies that $A'\mathbf{w}' = \mathbf{v}$. Hence $R(A) \subset R(A')$. However, reversing the argument works just as well so $R(A) = R(A')$.

Now suppose that A' is obtained by A by scaling by $\lambda \neq 0$ the first column of A. If $\mathbf{v} \in R(A)$ then $\mathbf{v} = A\mathbf{w}$ where $\mathbf{w} = \sum_{i=1}^n w_i\mathbf{e}_i$ then if we set $\mathbf{w}' = \frac{w_1}{\lambda}\mathbf{e}_1 + \sum_{i=2}^n w_i\mathbf{e}_i$ then $A'\mathbf{w}' = A\mathbf{w} = \mathbf{v}$. Hence in this case also $R(A) \subset R(A')$. Again the argument is reversible so $R(A) = R(A')$.

Finally, suppose A' is obtained from A by adding the second column to the first. If $\mathbf{v} \in R(A)$ then $\mathbf{v} = A\mathbf{w}$ where $\mathbf{w} = \sum_{i=1}^n w_i\mathbf{e}_i$. Now set $\mathbf{w}' = w_1\mathbf{e}_1 + (w_2 - w_1)\mathbf{e}_2 + \sum_{i=3}^n w_i\mathbf{e}_i$. Then one has $A'\mathbf{w}' = A\mathbf{w} = \mathbf{v}$ so $R(A) \subset R(A')$. Again the argument is reversible so $R(A) = R(A')$. □

In a corresponding way we may define the elementary row operations. Roughly speaking, an *elementary row operation* is one of the following operations: a) swapping two rows, b) scaling the first row by a non-zero scalar, or c) adding the second row to the first row. More precisely, for a matrix A a new matrix A' is obtained by an elementary row operation applied from A if it is given by one of the following formulas:

$$A = \begin{pmatrix} c_1 \\ \vdots \\ c_i \\ \vdots \\ c_j \\ \vdots \\ c_m \end{pmatrix}, A' = \begin{pmatrix} c_1 \\ \vdots \\ c_j \\ \vdots \\ c_i \\ \vdots \\ c_m \end{pmatrix}, A' = \begin{pmatrix} \lambda c_1 \\ \vdots \\ c_i \\ \vdots \\ c_j \\ \vdots \\ c_m \end{pmatrix}, A' = \begin{pmatrix} c_1 + c_2 \\ \vdots \\ c_i \\ \vdots \\ c_j \\ \vdots \\ c_m \end{pmatrix}$$

Unlike the elementary column operations, the elementary row operations preserve the null space, though they change the column space.

THEOREM 1.3. *Let $A \in \mathbb{C}^{m\times n}$ be a matrix. If A' is obtained from A by a (finite) sequence of elementary row operations then $N(A) = N(A')$.*

REMARK 1.4. In general $R(A) \neq R(A')$, though it is true that $dim R(A) = dim R(A')$. We will prove this later.

2. Reduced Echelon Form and Gaussian elimination

Elementary row operations and elementary column operations can be applied to a matrix A to produce new matrices A' and A'' that are "simpler" in a certain sense. In order to make this precise, we need a notion of what a "simple" matrix should be.

We make the following defintions:

DEFINITION 2.1. Let $A \in \mathbb{C}^{m\times n}$ and write

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

we say A is in *row reduced echelon form* (rref) if

(1) A is upper triangular, i.e. if $a_{ij} = 0$ when $i > j$.
(2) The first non-zero entry of each row of A is 1 (note some rows may be all zeros). This entry is called a *pivot*.
(3) The only non-zero entry in a column containing a pivot is the pivot entry.

DEFINITION 2.2. Let $A \in \mathbb{C}^{m\times n}$ and write

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

we say A is in *column reduced echelon form* (cref) if

(1) A is lower triangular, i.e. if $a_{ij} = 0$ when $i < j$.
(2) The first non-zero entry of each column of A is 1 (note some columns may be all zeros). This entry is called a *pivot*.
(3) The only non-zero entry in a row containing a pivot is the pivot entry.

The point about the echelon forms is that (as well discuss below) they are easier to extract information about the range and null space from. The point is that that one can always obtain a matrix A' which is in rref from A by elementary row operations. I won't discuss the details the algorithm to do this – namely Gaussian elimination – as you learned it in Math 51. In a similar manner, and by essentially the same algorithm for any matrix A one can obtain a matrix A' from A by column operations and so that A' is in cref.

The existence of such the Gaussian elimination algorithm is then a proof of the following theorems which we will use in this class.

THEOREM 2.3. *For any matrix $A \in \mathbb{C}^{m\times n}$ there is a unique matrix A' obtained by a finite number of elementary row operations from A so that A' is in rref. We write $A' = rref(A)$.*

Similarly, there is a unique matrix A'' obtained by a finite number of elementary column operations from A and so that A'' is in cref. We write $A'' = cref(A)$.

3. Pivots

The pivots of $cref(A)$ and $rref(A)$ allow one to determine (respectively) the range or the null space of A more easily. For instance one can check that the set of columns of $cref(A)$ containing a pivot forms a basis of $R(cref(A)) = R(A)$. Thus, $dimR(A)$ is the number of pivots in $cref(A)$. A more involved argument shows that, in $rref(A)$ looking at the null space, non-pivot columns correspond to "free variables" while pivot columns correspond to"pivot variables" that are determined by the free variables. In particular, one expects the dimension of the $N(rref(A)) = N(A)$ to be the number of "free variables".

4. Examples

I'll illustrate some of the preceding with an example. Let

$$A = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

column operation steps

$$\begin{pmatrix} 2 & 3 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1/2 & 1/2 & 1/2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & -1/2 & 0 \end{pmatrix}$$

so A has rank 2 and the range is spanned by the vectors

$$\begin{bmatrix} 1 \\ 0 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1/2 \end{bmatrix}$$

Notice that $A' = cref(A)$ has nullity 1 and null space spanned by $\mathbf{e}_2$ which is NOT in the null space of A. Row operations give:

$$\begin{pmatrix} 2 & 3 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 3/2 & 1/2 \\ 0 & 1 & 1 \\ 0 & 1/2 & 1/2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

It is straightforward to see that for $A' = rref(A)$, $Null(A')$ is spanned by $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ and hence so is $Null(A)$. In particular as expected the nullity is 1. I leave it to you to see that the$R(A)$ is not the same as $R(A')$.

5. Column and row operations as matrix multiplications

We claim (we will come back to this later in the course that there are $n \times n$ (non-singular) matrices S_{ij}, M_λ and C (corresponding to elementary row column operations a), b) and c) respectively) so that multiplication of A on the right by the matrix produces A', i.e. $A' = AC$. Where A' is obtained from A by one of the elementary column operations.

Similarly, there are $m \times m$ (non-singular) matrices so that multiplication of A on the left by these matrices yields the elementary row operations.

CHAPTER 6

Fifth Lecture

We present some (selected) proofs of important linear algebra facts.

THEOREM 0.1. *Every vector space $E \subset \mathbb{C}^n$ admits a basis.*

PROOF. This is such a fundamental result that it can be a bit difficult to prove and so we don't do it here. □

LEMMA 0.2. *If A is a $m \times n$ matrix and $m < n$ (i.e. A is short and wide) then there is a non-zero vector $\mathbf{v} \in \mathbb{C}^n$ so $A\mathbf{v} = 0$. (i.e. $N(A)$ is non-trivial).*

PROOF. Let $A' = rref(A)$. We verified last time that $N(A') = N(A)$, so we just need to find a non-zero vector in $N(A')$. Write

$$A' = \begin{bmatrix} \mathbf{a}'_1 | & \cdots & | \mathbf{a}'_n \end{bmatrix}$$

so $\mathbf{a}'_i$ are the columns of A'. Since there is only one pivot (at most) in each column and row) and $m < n$ there must be a column $\mathbf{a}'_{j_0}$ without a pivot. Write

$$\mathbf{a}_{i_0} = \begin{bmatrix} a_{1j_0} \\ \vdots \\ a_{mj_0} \end{bmatrix}$$

Let

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

where

$$v_i = \begin{cases} -a_{ij_0} & \text{if } \mathbf{a}'_i \text{ has a pivot and } i < j_0 \\ 0 & \text{if } \mathbf{a}'_i \text{ has no pivot and } i < j_0 \\ 1 & \text{if } i = j_0 \\ 0 & \text{if } i > j_0 \end{cases}$$

Notice that $\mathbf{v} \neq 0$ and $A'\mathbf{v} = 0$ hence $\mathbf{v} \in N(A)$. □

THEOREM 0.3. *If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is a basis of $E \subset \mathbb{C}^n$ and $\mathbf{w}_1, \ldots \mathbf{w}_l$ is also a basis for E then $k = l$.*

PROOF. We argue by contradiction. Up to a relabelling we may assume that $k > l$. The fact that the $\mathbf{w}_i$ are a basis means there are c_{ij} so that $\mathbf{v}_j = \sum_{i=1}^{k} c_{ij}\mathbf{w}_i$. Now let V be the $n \times k$ matrix whose columns are $\mathbf{v}_j$ and W be the $n \times l$ matrix whose columns are the $\mathbf{w}_i$. Let C be the $l \times k$ matrix with entries c_{ij}. Then we have

$$V = WC$$

Now C is a $l \times k$ matrix and $l < k$ so by Lemma 0.2 there is a non-trival $\mathbf{x}$ so that $C\mathbf{x} = 0$ but then $V\mathbf{x} = 0$ but this implies the columns of V (i.e. the $\mathbf{v}_j$s) are linearly dependent. A contradiction. □

LEMMA 0.4. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_l$ be linearly independent vectors in E. Then $dim(E) \geq l$.*

PROOF. This is very similar to the previous argument. We argue by contradiction. Suppose that $dim(E) = k < l$. By Theorem 0.1, there is a basis of E and by Theorem 0.3 we may write it as $\mathbf{w}_1, \ldots, \mathbf{w}_k$. As before writing the $\mathbf{v}_j$ in terms of the $\mathbf{w}_i$ yields a contradiction to Lemma 0.2 Hence $dim(E) = k \geq l$. □

THEOREM 0.5. *Any set of linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_l$ in a vector space E can be extended to a basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ where $k = dim(E) \geq l$.*

PROOF. Set $m = k - l$ i.e. the difference between number of linear inependent vectors $\mathbf{v}$ we start with and the number we want to end up with. By Lemma 0.4 $m \geq 0$. By Theorems 0.1 and 0.3 and Lemma 0.4 when $m = 0$ then the vectors we started with form a basis. To really see this just need to verify that $\mathbf{v}_1, \ldots, \mathbf{v}_l$ generate E. To that end pick any $\mathbf{w} \in E$. If $\mathbf{w} \notin Span\{\mathbf{v}_1, \ldots, \mathbf{v}_l\}$ then $\{\mathbf{w}, \ldots, \mathbf{v}_1, \ldots, \mathbf{v}_l\}$ is linearly independent, then Lemma 4 implies $\dim E \geq l + 1$ contradiction $m = 0$.

We now prove the theorem by induction on m. That is fix $m \geq 0$: suppose that we know that for any vector space E' and lin indep vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{l'}$ in F with $dim(E') = k'$ satisfies $k' - l' = m$ then $\mathbf{a}_1, \ldots, \mathbf{a}_{l'}$ can be extended to a basis of F $\mathbf{a}_1, \ldots, \mathbf{a}_{k'}$. (i.e. we added m new vectors) We want to conclude that for any other vector space E'' with lin indep vectors $\mathbf{b}_1, \ldots, \mathbf{b}_{l''}$ so that $dim(E'') = k''$ satisfies $k'' - l'' = m+1$ then $\mathbf{b}_1 \ldots, \mathbf{b}_{l''}$ extends to a basis $\mathbf{b}_1, \ldots, \mathbf{b}_{k''}$ (i.e. we added $m+1$ new vectors). We see this as follows: By Theorem **??** the $\mathbf{b}_1, \ldots, \mathbf{b}_{l'}$ cannot span E'' as otherwise they would form a basis but $m + 1 > 0$. Hence pick any vector $\mathbf{w}$ in E'' not in the span of the $\mathbf{b}_1, \ldots, \mathbf{b}_{l'}$. Now if set $\mathbf{b}_{l''+1} = \mathbf{w}$ then the new set of vectors $\mathbf{b}_1, \ldots, \mathbf{b}_{l''}, \mathbf{b}_{l''+1}$ is a) Linearly independent in F'' and hence together with F'' satisfies the induction hypotheses.

This proves the theorem. □

REMARK 0.6. If you are having problems with this think about the case $m = 1$ and $m = 2$.

Some applications to matrices.

THEOREM 0.7. *Let A be a $m \times m$ matrix . Then A is full rank if and only if the columns of A form a basis of $\mathbb{C}^m$*

PROOF. $\Rightarrow$. We need to check that the columns are a basis that is the generate and are linearly independent. We know that A has full rank implies the dimension of $R(A)$ is n which implies by Theorems 0.1 and 0.3 that $R(A)$ has basis $\mathbf{w}_1, \ldots, \mathbf{w}_m \in \mathbb{C}^m$. As the $\mathbf{w}_i$s are linearly indpendet, by Theorem 5 we can extend $\mathbf{w}_1, \ldots, \mathbf{w}_m$ to a basis of $\mathbb{C}^m$ but by Theorem 0.3 this extension must be trival hence $R(A) = \mathbb{C}^m$ and since the columns spance $R(A)$ they span $\mathbb{C}^n$. We check that the columns are linearly indendent as follows: if they failed to be linearly independent then one could remove one of the columns and have $m-1$ vectors spanning $\mathbb{C}^n$. In particular, there would be a $m \times (m-1)$ matrix A' with $R(A') = \mathbb{C}^m$. In particular, if $\mathbf{e}_i$ is the standard basis of $\mathbb{C}^m$ we can find $\mathbf{c}_i \in \mathbb{C}^{m-1}$ so that $\mathbf{e}_i = A'\mathbf{c}_i$. Let C be the

$m-1 \times m$ matrix with columns $\mathbf{c}_i$ then $Id = A'C$, but C is short and wide and so by Lemma 2 there is a non-trival $\mathbf{x} \in \mathbb{C}^m$ so that $C\mathbf{x} = 0$. Since $Id\mathbf{x} = \mathbf{x} \neq 0$ this is a contradiction.

$\Leftarrow$. Since the columns form a basis their span has dimension m. Hence $dimR(A) = n$ and hence A has full rank. □

COROLLARY 0.8. *If A is an $m \times m$ matrix with linearly independent columns then A has full rank.*

PROOF. As the columns are linearly independent and there are m of them, they must span $\mathbb{C}^m$ hence they are a basis. □

CHAPTER 7

Sixth Lecture

In this lecture we discussed how to interpret matrices as systems of linear equations. We also discussed non-singular matrices.

1. Systems of Linear Equations

We disucss one of the classic applications of linear algebra. Namely solving systems of linear equations. We've already seen these sorts of questions in other guises.

The basic set up is let $A \in \mathbb{C}^{m\times n}$ be a $m \times n$ matrix. We can interpret this matrix as a system of m linear equations in n unknowns by letting $\mathbf{x} \in \mathbb{C}^n$ be a vector of variables and $\mathbf{b} \in \mathbb{C}^m$ be fixed and try and look for solutions to:

$$A\mathbf{x} = \mathbf{b}$$

Two important and natural questions immediately arise. Is there always a solution? And if so is it unique? In general both anwers are false so it is a good idea to be able to quantify them. Of course one would also like to find a solution if it exists, but that is a more computational question.

It follows from the definitions pretty much directly that our system has a solution when and only when $\mathbf{b} \in R(A)$. On the other hand if $\mathbf{x} = \mathbf{v}$ is a solution and $\mathbf{v}' \in N(A)$ then it is also clear $\mathbf{x} = \mathbf{v} + \mathbf{v}'$ is also a solution. Similarly, if $\mathbf{x} = \mathbf{v}$ and $\mathbf{xw}$ are two solutions then $\mathbf{v} - \mathbf{w} \in N(A)$. In other words the null space precisely characterizes how solutions fail to be unique while the range characterizes which inputs lead to solutions.

We point out that thinking about the A as a linear transformation leads to some important (and equivalent) notions. Namely, that of *surjective* and *injective* maps. Consider the linear map $\hat{A}$ associated to A from $\mathbb{C}^n$ to $\mathbb{C}^m$ given by

$$\begin{aligned} \hat{A} : \mathbb{C}^n &\to \mathbb{C}^m \\ \mathbf{x} &\mapsto A\mathbf{x} \end{aligned}$$

We say that $\hat{A}$ is surjective (or onto) precisely when for all $\mathbf{b} \in \mathbb{C}^m$ there is an $\mathbf{a} \in \mathbb{C}^n$ so $\hat{A}(\mathbf{a}) = \mathbf{b}$ that is one can always solve $A\mathbf{x} = \mathbf{b}$ for any $\mathbf{b}$. We say $\hat{A}$ is injective (or 1-1) when $\hat{A}(\mathbf{x}) = \hat{A}(\mathbf{y})$ implies that $\mathbf{x} = \mathbf{y}$ that is there is at *most* one solution to $A\mathbf{x} = \mathbf{b}$ (there may be none).

We note the following equivalent properties for $A \in \mathbb{C}^{m\times n}$: A is surjective as a linear map, $A\mathbf{x} = \mathbf{b}$ always has a solution, $R(A) = \mathbb{C}^m$, the columns of A span $\mathbb{C}^m$. Similiarly: A is injective: $A\mathbf{x} = \mathbf{b}$ has at most one solution, $N(A) = \{0\}$, the columns of A are linearly independent.

2. Non-Singular Matrices

We won't at present discuss further the mechanics of solving systems. The standard approach to this is Gaussian elimination a topic that should have been covered in great depth in Math 51. Rather we will specialize to a very special case. Namely when A is surjective *and* injective, i.e. $R(A) = \mathbb{C}^n$ and $N(A) = \mathbb{C}^m$.

One easy to see fact about A in this case is that A is $m \times m$ i.e. square. This follows as each column is an m vector and there are n of them in $m \times n$ matrix. Injective implies the columns of A are linearly-independent while surjective implies they span $\mathbb{C}^n$. That is they form a basis of $\mathbb{C}^m$ and so $m = n$. We call such a matrix A *Non-Singular*. A $m \times m$ matrix that is not non-singular is called singular. Notice that by Theorem 0.7 and Corollary 0.8 any $m \times m$ matrix is non-singular provided that either columns linearly-independent *or* span

Let A be a $m \times m$ non-singular matrix. One of the most imporant facts about such A is that there exists another matrix which we denote by A^{-1} so that $AA^{-1} = I$. We see this as follows: Let $\mathbf{a}_i$ be the unique solution to $A\mathbf{x} = \mathbf{e}_i$. Then one has:

$$A^{-1} = \begin{bmatrix} \mathbf{a}_1 & | \dots | & \mathbf{a}_m \end{bmatrix}$$

Then we check that $AA^{-1} = I$. A^{-1} is the inverse and so also say that A is invertible

We list some useful facts and indicate an rough idea of the proofs.

- A^{-1} is non-singular. To see this note that $N(A^{-1}) = \{0\}$. Indeed, if $\mathbf{x} \in N(A^{-1})$ then $\mathbf{x} = AA^{-1}\mathbf{x} = A0 = 0$. Hence, A^{-1} is non-singular.
- $A^{-1}A = I$. To see this, we note that $A^{-1} = A^{-1}I = A^{-1}AA^{-1}$. Hence, $I = A^{-1}(A^{-1})^{-1} = A^{-1}AA^{-1}(A^{-1})^{-1} = A^{-1}A$. Here $(A^{-1})^{-1}$ exists as A^{-1} is non-singular.
- If $BA = Id$ or $AB = Id$ then A is non-singular and $B = A^{-1}$. To see this multiply (on the right or left) by A^{-1}.
- $(A^{-1})^{-1} = A$.
- If A, B non-singular then so is AB and $(AB)^{-1} = B^{-1}A^{-1}$. Conversely, if AB is non-singular then both A and B are.

We are going to see non-singular matrices again and again. This is because and this is really important *the columns of a non-singular matrix form a basis* and *a basis gives a non-singular matrix by taking them as columns*! More precisely, given a basis $\mathbf{a}_i$ of $\mathbb{C}^m$ we can write some vector

$$\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \sum_{i=1}^{m} b_i \mathbf{e}_i = I\mathbf{b}$$

as

$$\mathbf{b} = \sum_{i=1}^{m} c_i \mathbf{a}_i = A\mathbf{c}$$

Here A is $m \times m$ matrix with colums $\mathbf{a}_i$. The c_i are the coefficients of $\mathbf{b}$ in the basis $\mathbf{a}_i$. We say that the c_is (equvialently the $\mathbf{c}$) are the *coefficients of* $\mathbf{b}$ *with respect to the basis* $\{a_i\}$

The point is the matrix A tells us how to determine the coefficients of $\mathbf{b}$ with respect to the *standard basis* in terms of the coefficients $\mathbf{c}$. On the other hand, multiplying by A^{-1} we have $A^{-1}\mathbf{b} = \mathbf{c}$ in other words A^{-1} tells us how to write

the c_i in terms of the b_i. In other words how to write the coefficients of $\mathbf{b}$ in terms of the basis $\mathbf{a}$ in terms of the coefficients of the standard basis. For this reason non-singular matrices are sometimes said to give a *change of basis*.

Why is changing basis good? Well some problems are easier to understand in a given basis. More importantly a matrix X often has a natural basis (usually different form the standard basis) on which it behaves particularly simple. For example, let us write

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

We claim that the three vectors form a basis of $\mathbb{C}^3$. Now let us try and write

$$\mathbf{w} = \begin{bmatrix} 2, 3, 1 \end{bmatrix}$$

in terms of this basis. That is find c_1, c_2, c_3 so that $\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$ This amounts to looking at

$$\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Then we have

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

computing out

$$\begin{bmatrix} 1 & 0 & 1 \\ -1/2 & 1/2 & -1/2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

3. How to tell if A is non-singular

There are a number of ways to recognize that an $m \times m$ matrix A is non-singular we won't discuss all of them. When $A \in \mathbb{C}^{m \times m}$ then A is nonsingular is equivalent to...

- $N(A) = \{0\}$ equivalently the columns of A are linearly independent equivalently for all $\mathbf{b}$, $A\mathbf{x} = \mathbf{b}$ has at most one solution.
- $R(A) = \mathbb{C}^m$ equivalently the columns span $\mathbb{C}^m$ equivalently for all $\mathbf{b}$ $A\mathbf{x} = \mathbf{b}$ has a solution.
- There is a $m \times m$ matrix B so $AB = I$ or $BA = I$
- $\det(A) \neq 0$

The last condition is that the determinant of the matrix. We won't discuss this much further as it is a concept that while useful theoretically, almost never gets used in the algorithms we will study. As I consequence, I'll defer defining the determinant till we use it (if ever).

Lets see another example. Lets say I tell you that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is a basis of $\mathbb{C}^3$. Then I define $\mathbf{w}_1 = \mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3$, $\mathbf{w}_2 = \mathbf{v}_2 - \mathbf{v}_3$ and $\mathbf{w}_3 = \mathbf{v}_1 + \mathbf{v}_3$. How can we determine if $\mathbf{w}_i$s are a basis of $\mathbb{C}^3$? To answer this lets rewrite this as a matrix

problem.

$$\begin{bmatrix}\mathbf{w}_1 & |\mathbf{w}_2| & \mathbf{w}_3\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1 & |\mathbf{v}_2| & \mathbf{v}_3\end{bmatrix} \begin{bmatrix}1 & 0 & 1\\ -2 & 1 & 0\\ 1 & -1 & 1\end{bmatrix}$$

I.e. for each column multiplying by V replaces standard basis by basis $\mathbf{v}_i$. For $\mathbf{w}_i$ to be a basis W needs to be non-singular. V is non-singular so it is enougth that the right hand side is non-singular. Row reducing shows that it is.

Now how do we write $\mathbf{v}_i$'s in terms of the $\mathbf{w}_i$'s?

$$\begin{bmatrix}\mathbf{w}_1 & |\mathbf{w}_2| & \mathbf{w}_3\end{bmatrix} \begin{bmatrix}1 & 0 & 1\\ -2 & 1 & 0\\ 1 & -1 & 1\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{v}_1 & |\mathbf{v}_2| & \mathbf{v}_3\end{bmatrix}$$

CHAPTER 8

Seventh Lecture and Eighth Lecture

In this lecture we introduced the hermitian inner product and adjoint. These are generalizations to the complex settings of the dot product and transpose which are defined for real vectors and matrices. This is an amalgamation of the seventh and eigth lectures.

1. Notation

We remind our selves of some notation before preceding further

Let $A \in \mathbb{C}^{m\times n}$ be a $m \times n$ matrix. We've often expressed A as a set of columns:

$$A = \begin{bmatrix} \mathbf{a}_1 | & \dots & | \mathbf{a}_n \end{bmatrix}$$

here $\mathbf{a}_i$ is also a $m \times 1$ matrix We can also write A in terms of its rows

$$A = \begin{bmatrix} a_1' \\ \vdots \\ a_m' \end{bmatrix}$$

where a_j' is a $1 \times n$ matrix. As we've seen multiplying A on the right by a vector

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

is the same as getting a linear combination of the columns, that is

$$A\mathbf{v} = \sum_{j=1}^{n} v_j \mathbf{a}_j$$

In a similar manner multiplying A on left by a row $q = \begin{bmatrix} q_1 & \dots & q_m \end{bmatrix}$ gives a linear combination of the rows:

$$q\mathbf{v} = \sum_{i=1}^{m} q_i a_i' \in \mathbb{C}^{1\times n}$$

We then have rules for matrix multiplication if $B \in \mathbb{C}^{n\times k}$ and

$$B = \begin{bmatrix} \mathbf{b}_1 | & \dots & | \mathbf{b}_k \end{bmatrix}$$

then

$$AB = \begin{bmatrix} A\mathbf{b}_1 | & \dots & | A\mathbf{b}_n \end{bmatrix}$$

Similarly, if we write

$$B = \begin{bmatrix} b_1' \\ \vdots \\ b_m' \end{bmatrix}$$

then

$$AB = \begin{bmatrix} a_1'B \\ \vdots \\ a_m'B \end{bmatrix}$$

In either case this yields:

$$AB = \begin{bmatrix} a_1'\mathbf{b}_1 & \cdots & a_1'\mathbf{b}_k \\ \vdots & a_i'\mathbf{b}_j & \vdots \\ a_m'\mathbf{b}_1 & \cdots & a_m'\mathbf{b}_k \end{bmatrix}$$

2. Adjoints

We now introduce an important formal operation on matrices. Let

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{C}^{m\times n}, B = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix} \in \mathbb{C}^{n\times m}$$

We say that B is the *hermitian conjugate* or *adjoint* of A when and only when $b_{ij} = \bar{a}_{ji}$ and write $B = A^*$. Another way to think about this is: Let $\mathbf{v} \in \mathbb{C}^m$ be the vector $\mathbf{v} = (v_1, \ldots, v_m)$. we define $\mathbf{v}^* = [\bar{v}_1, \ldots, \bar{v}_m]$. For a matrix $A \in \mathbb{C}^{m\times n}$ we write $A = \begin{bmatrix} \mathbf{a}_1 | & \cdots & |\mathbf{a}_n \end{bmatrix}$ and define

$$A^* = \begin{bmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_n^* \end{bmatrix}$$

When $A \in \mathbb{R}^{m\times n}$ is a real matrix we have that $\bar{a}_{ij} = a_{ij}$ and so all are doing when taking the adjoint is flipping. You've probaby see this before and called it the transpose $A^\top$.

An important class of matrices are the *hermitian* and *symmetric* matrices. A hermitian matrix is one so that $A^* = A$. A symmetric matrix is just a hermitian matrix that has real entries (and so $A^\top = A$). Notice any such matrix (in either case) is square.

3. Inner Products

We may think of a vector $\mathbf{v} \in \mathbb{C}^n$ as a $n \times 1$ matrix. Hence we can write $\mathbf{v}^*$ to get a $1 \times n$ matrix. We define the innerproduct of two vectors $\mathbf{v}$ and $\mathbf{w} \in \mathbb{C}^n$ as:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^*\mathbf{w}$$

Notice that when the vectors have real entries (i.e. $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$) this is just the usual dot product. This suggests we use the inner product to define a notion of "length" of a vector. That is set:

$$||\mathbf{v}||_2 = \sqrt{\mathbf{v}^*\mathbf{v}} \geq 0.$$

We point out that $\mathbf{v}^*\mathbf{v} \geq 0$ for any vector so the squareroot is okay. We included complex conjugation precisely to achieve this. When $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we can really geometrically think of $||\mathbf{v}||_2$ as the length of $\mathbf{v}$ and $\mathbf{v}^*\mathbf{w} = ||\mathbf{v}||_2||\mathbf{w}||_2 \cos\alpha$ where α is the angle between $\mathbf{v}$ and $\mathbf{w}$. For complex vectors this geometric interpretation doesn't make as much sense, but is still useful for intuition.

Some useful properties (left as an exercise). Bilinearity of the innerproduct:

$$(\mathbf{v}_1 + \mathbf{v}_2)^*\mathbf{w} = \mathbf{v}_1^*\mathbf{w} + \mathbf{v}_2^*\mathbf{w}$$
$$\mathbf{v}^*(\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v}^*\mathbf{w}_1 + \mathbf{v}^*\mathbf{w}_2$$
$$(\alpha\mathbf{v})^*(\beta\mathbf{w}) = \bar{\alpha}\beta\mathbf{v}^*\mathbf{w}$$

These all follow from properties of matrix multiplication. Note this implies $||\lambda\mathbf{v}||_2 = |\lambda|||\mathbf{v}||_2$. The innerproduct satisfies the following inequality known as the Cauchy-Schwarz inequality:

$$|\mathbf{v}^*\mathbf{w}| \leq ||\mathbf{v}||_2||\mathbf{w}||_2.$$

with equality if and only if the $\mathbf{v}$ and $\mathbf{w}$ are collinear. The length also satisfies the so called *triangle inequality*:

$$||\mathbf{v} + \mathbf{w}||_2 \leq ||\mathbf{v}||_2 + ||\mathbf{w}||_2.$$

Note that both of these are easily shown for real vectors and have nice geometric intrepretations, but they also hold for complex vectors.

For general matrices $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times k}$ one has $(AB)^* = B^*A^*$. In particular if $\mathbf{v} \in \mathbb{C}^m$ and $\mathbf{w} \in \mathbb{C}^n$ then $\mathbf{v}^*(A\mathbf{w}) = (A^*\mathbf{v})^*\mathbf{w}$. That is

$$\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, A^*\mathbf{w}\rangle$$

This last fact is really key and we will return to it soon.

4. Orthogonality

One useful thing to use the inner product for is to tell if two vectors are *orthogonal*. We say $\mathbf{v}$ and $\mathbf{w}$ are orthogonal if $\mathbf{v}^*\mathbf{w} = 0$. If the vectors are real then this means geometrically that they are perpendicular. An example $\mathbf{e}_i$ and $\mathbf{e}_j$ are orthogonal when $i \neq j$.

We say that two sets of vectors E and F (not neccesarily vector spaces) are *orthogonal* if whenever $\mathbf{v} \in E$ and $\mathbf{w} \in F$ one has $\mathbf{v}^*\mathbf{w} = 0$. A set of non-zero vectors S is *orthogonal* if for any $\mathbf{v} \in S, \mathbf{w} \in S$ with $\mathbf{v} \neq \mathbf{w}$ one has $\mathbf{v}^*\mathbf{w} = 0$. This set is orthonormal if in addition $||\mathbf{v}||_2 = 1$ for all $\mathbf{v} \in S$. Examples include the standard basis and the vectors

$$1/\sqrt{2}\begin{bmatrix} I \\ -I \end{bmatrix}, 1/\sqrt{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{C}^2.$$

Probably the most important thing about orthogonality for our purposes is that it is a simple condition that ensures linear independence.

THEOREM 4.1. *(2.1 in T-B) The vectors in an orthogonal set S are linearly independent.*

PROOF. Suppose one has $\mathbf{v}_i \in S$ $(i = 1..l)$ that are linearly dependent. Then we can write $\mathbf{v}_k = \sum_{i=1}^{l} c_i\mathbf{v}_i$ where $c_k = 0$ and $\mathbf{v}_k$. Then $\mathbf{v}_k^*\mathbf{v}_k = \mathbf{v}_k^*\sum_{i=1}^{l} c_i\mathbf{v}_i = \sum_{i=1}^{l} c_i\mathbf{v}_k^*\mathbf{v}_i$ by the bilinearity . The orthogonality says that the left hand side equls $c_k\mathbf{v}_k^*\mathbf{v}_k = 0$. However, this implies $\mathbf{v}_k = 0$ which is impossible. □

One consequence is that if $S \subset \mathbb{C}^m$ is an orthogonal set then S is a basis of $span(S)$. We refer to such a basis as an *orthonormal basis* if in addition for $\mathbf{v} \in S$ $||\mathbf{v}||_2 = 1$. If $S \subset \mathbb{C}^m$ and there are m vectors in S then S is a basis of $\mathbb{C}^m$. Note by normalizing one can always go from an orthogonal set to an orthonormal set. One good property of an orthonormal basis is that it is easy to find the coefficients

of a vector with respect to the basis using the inner product. Namely, if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are an orthormal basis of $E \subset \mathbb{C}^m$ then We can write $\mathbf{v} \in \mathbb{C}^m$ as

$$\mathbf{v} = \sum_{i=1}^{k} \langle \mathbf{v}_i, \mathbf{v} \rangle \mathbf{v} = \sum_{i=1}^{k} (\mathbf{v}_i^* \mathbf{v}) \mathbf{v}$$

i.e. the coefficients are $\mathbf{v}_i^* \mathbf{v}$.

5. Unitary Matrices

We now introduce an important class of matrices that are related to what we just discussed. The idea is that while a basis corresponds to a non-singular matrix, an *Orthonormal basis* corresponds to a *Unitary* matrix. Since expanding components in an orthonormal basis is easier than doing it for a generic basis, so finding inverses for unitary matrices is easier than for non-singular matrices.

We say $Q \in \mathbb{C}^{m \times m}$ is *unitary* if $Q^* = Q^{-1}$ (if $Q \in \mathbb{R}^{m \times m}$ say is it is *orthogonal*). That is if $Q^*Q = QQ^* = I$. It straight forward to check that if $Q = \begin{bmatrix} q_1 & | \ldots | & q_m \end{bmatrix}$ then $Q^* = \begin{bmatrix} q_1^* \\ \vdots \\ q_m^* \end{bmatrix}$ so

$$Q^*Q = \begin{bmatrix} q_1^* q_1 & \cdots & q_1^* q_m \\ \vdots & \ddots & \vdots \\ q_m^* q_1 & \cdots & q_m^* q_m \end{bmatrix}$$

So Q unitary if and only if the columns form an orthonormal basis. In particular $Q^*\mathbf{b}$ gives the coefficients of $\mathbf{b}$ in the orthonormal basis given by the columns of Q. Two important additional properties of unitary matrices are that they preserve innerproduct and $2-norm$ i.e. $(Q\mathbf{v})^*(Q\mathbf{w}) = \mathbf{v}^* Q^* Q \mathbf{w} = vV^* \mathbf{w}$. and So $||Q\mathbf{v}||_2 = ||\mathbf{v}_2||$.

For real matrices this has the geometric interpretation that Q is given by a rigid motion fixing the origin. For instance rotation about the origin or reflection through any line (or plane etc) through the origin.

Example: we check that rotation is Unitary.

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

then $A^*A = A^\top A = Id$ is a straight forward computation.

CHAPTER 9

Ninth and Tenth Lectures

In this Lecture I started to discuss complementary decompositions of vector spaces as well as projectors.

1. Orthogonal Projections

One of the key uses of inner products is that they allow one to decompose arbitrary vectors into orthogonal components. This often simplifies a problem substantially. We say this already when one has a basis but the idea is more general.

The basic idea: Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an orthonormal set of vectors. For $\mathbf{w}$ an arbitrary vector we have that $\mathbf{v}_i^*\mathbf{w}$ is a scalar. If we write

$$r = \mathbf{w} - (\mathbf{v}_1^*\mathbf{w})\mathbf{v}_1 - (\mathbf{v}_2^*\mathbf{w})\mathbf{v}_2 - \ldots - (\mathbf{v}_k^*\mathbf{w})\mathbf{v}_k$$

Then it is straight forward to check that r is orthogonal to S that is we can write:

$$\mathbf{w} = r + (\mathbf{v}_1^*\mathbf{w})\mathbf{v}_1 + (\mathbf{v}_2^*\mathbf{w})\mathbf{v}_2 + \ldots + (\mathbf{v}_k^*\mathbf{w})\mathbf{v}_k = r + \sum_{i=1}^{k}(\mathbf{v}_i\mathbf{v}_i^*)\mathbf{w}$$

so all the summand vectors are orthogonal. Note the second equality just uses the fact that scalar multiplication can be commuted.

Notice that if the $\mathbf{v}_i$ form a basis then $r = 0$. That is we have expressed $\mathbf{w}$ in terms of the basis $\mathbf{v}_i$ in a relatively painless manner. This is one of the real powers of inner products and orthogonality.

I point out also that when I rewrote the expansion in terms of $(\mathbf{v}_i\mathbf{v}_i^*)\mathbf{w}$ I wasn't doing much mathematically but the interpretation is important. The poin this $\mathbf{v}_i\mathbf{v}_i^*$ is now a square matrix P_i that one can check preserves any vector $\lambda\mathbf{v}_i$ (i.e. $P_i(\lambda\mathbf{v}_i) = \lambda\mathbf{v}_i$ and has $Null(P_i)$ the space of vectors orthogonal to $\mathbf{v}_i$. That is P_i is the projection matrix onto $\mathbf{v}_i$. These are important special cases of a more general type of matrix that will be important for us.

Notice that philosphically the two expansions are different. The first we view $\mathbf{w}$ as coefficients $\mathbf{v}_i^*\mathbf{w}$ times the $\mathbf{v}_i$ plus some left over term r while in the second we view $\mathbf{w}$ as the sum of vectors $(\mathbf{v}_i\mathbf{v}_i^*)\mathbf{w}$ given by projecting plus some left over term r. projections

We come up with a more general concept of orthogonal projection if we let

$$P = \sum_{i=1}^{k}(\mathbf{v}_i\mathbf{v}_i^*)$$

then $\mathbf{w} = P\mathbf{w} + r$. Here P is $m \times m$ matrix with rank k. P gives projection onto the span of the $\mathbf{v}_i$. For instance if $k = 2$ this is projection onto a plane. We point out that $P^2 = P$ and $P^* = P$. I.e. P is *idempotent* and *hermitian*. We will return to this soon.

2. Sums of vector spaces

In order to get a better sense of what is going on with orthogonal projection we need some ideas about vector subspaces. It will also help to take a slightly more general point of view.

To that end let us suppose that we have two vector spaces $E_1, E_2 \subset \mathbb{C}^n$. We denote by $E_1 + E_2$ the vector space so that $\mathbf{w} \in E_1 + E_2$ when and only when there are $\mathbf{v}_1 \in E_1, \mathbf{v}_2 \in E_2$ so that $\mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2$. It is straightforward to check that $E_1 + E_2$ is a vector space and I leave it as an excercise.

An important fact is that if $E_1 \cap E_2 = \{0\}$ then each $\mathbf{w} \in E_1 + E_2$ can be written *UNIQUELY* as $\mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2$ where $\mathbf{v}_1 \in E_1$ and $\mathbf{v}_2 \in E_2$. To see this suppose that $\mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{w} = \mathbf{v}_1' + \mathbf{v}_2'$ where $\mathbf{v}_i, \mathbf{v}_i' \in E_i$. By equating the two sides we have $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_1' + \mathbf{v}_2'$ that is $\mathbf{v}_1 - \mathbf{v}_1' = \mathbf{v}_2' - \mathbf{v}_2$ we denote the common value by $\mathbf{v}$. Notice the left hand side is in E_1 while the right hand side is in E_2 and so $\mathbf{v} \in E_1 \cap E_2$ and so $\mathbf{v} = 0$. In other words $\mathbf{v}_1 = \mathbf{v}_1'$ and $\mathbf{v}_2 = \mathbf{v}_2'$.

We will mostly be interested in situations where E_1 and E_2 span $\mathbb{C}^n$ that is $E_1 + E_2 = \mathbb{C}^n$ and $E_1 \cap E_2 = \{0\}$. In this case we say that E_1 and E_2 are *complimentary*. The idea here is now that any vector $\mathbf{w} \in \mathbb{C}^n$ can be written as $\mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2$ where $\mathbf{v}_i \in E_i$.

For example: Let $\{\mathbf{b}_i\}$ be a basis of $\mathbb{C}^n$, $i = 1, \ldots, n$ if $E_1 = span\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and $E_2 = span\{\mathbf{b}_{k+1}, \ldots, \mathbf{b}_n\}$ then E_1 and E_2 are complementary subspaces.

One important task is: Given two complementary vector spaces E_1, E_2 in $\mathbb{C}^n$ we know that for any $\mathbf{w} \in \mathbb{C}^n$ we have $\mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_i \in E_i$ and this decomposition is unique. The question is to what extent can we determine $\mathbf{v}_1$ from $\mathbf{w}$.

We claim that in fact there is a fairly straightforward answer to this question. Namely there is a $n \times n$ matrix P so that $\mathbf{v}_1 = P\mathbf{w}$. Such a P is called a *projector*. We will return to them in a bit.

Before disucssing projectors we wish to point out one final thing. If E_1 and E_2 are orthogonal subspaces in $\mathbb{C}^n$ and they span $\mathbb{C}^n$ then they are automatically complementary (any $\mathbf{v} \in E_1 \cap E_2$ would satisfy $\langle \mathbf{v}, \mathbf{v} \rangle = 0$). In this case E_1 and E_2 are said to be *orthogonal complements*. We've already seen that it is fairly straightforward to find the orthogonal projector in this case. Indeed, let us use our fact that we can find $\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$ an orthonormal basis of E_1 then as we've seen for any $\mathbf{w} \in \mathbb{C}^n$

$$\mathbf{w} = (\sum_{j=1}^{k} \mathbf{q}_j \mathbf{q}_j^*)\mathbf{w} + \mathbf{r}$$

where $\mathbf{r}$ is orthogonal to the $\mathbf{q}_i$ and hence to E_1 and so lies in E_2. In particular our projector in this case is $P = \sum_{j=1}^{k} \mathbf{q}_j \mathbf{q}_j^*$.

3. Projectors

As mentioned in the previous section for any pair of complementary spaces $E_1, E_2 \subset \mathbb{C}^n$ there are matrices P_1 and P_2 in $\mathbb{C}^{n \times n}$ so that $P_1\mathbf{w} \in E_1$ and $P_2\mathbf{w} \in E_2$ and $\mathbf{w} = P_1\mathbf{w} + P_2\mathbf{w}$. We then call the P_i *projectors*.

To see this we argue as follows. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis of $\mathbb{C}^n$. For each i the fact that E_1 and E_2 is orthogonal allows us to write

$$\mathbf{e}_i = \mathbf{a}_i + \mathbf{b}_i$$

so that $\mathbf{a}_i \in E_1$ and $\mathbf{b}_i \in E_2$ and the $\mathbf{a}_i$ and $\mathbf{b}_i$ are neccesarily unique. We then set:

$$P_1 = \begin{bmatrix}\mathbf{a}_1| & \cdots & |\mathbf{a}_n\end{bmatrix}, P_2 = \begin{bmatrix}\mathbf{b}_1| & \cdots & |\mathbf{b}_n\end{bmatrix}.$$

And claim that P_1 and P_2 are the desired matrices. To see this it suffices to show

LEMMA 3.1. *Let $\mathbf{v} \in E_1$ then $P_1\mathbf{v} = \mathbf{v}$ and $P_2\mathbf{v} = 0$.*

PROOF. Write $\mathbf{v} = \sum_{i=1}^n v_i\mathbf{e}_i = \sum_{i=1}^n v_i(\mathbf{a}_i + \mathbf{b}_i) = \sum_{i=1}^n v_i\mathbf{a}_i + \sum_{i=1}^n v_i\mathbf{b}_i$. Notice that the first summand is in E_1 while the second is in E_2. Since we can also write $\mathbf{v} = \mathbf{v} + 0$ where the first summand is in E_1 and the second is in E_2 by the uniquess of the decomposition (as E_1 and E_2 are complementary we have

$$\mathbf{v} = \sum_{i=1}^n v_i\mathbf{a}_i$$

and

$$0 = \sum_{i=1}^n v_i\mathbf{b}_i$$

One the other hand, $P_1\mathbf{v} = P_1\left(\sum_{i=1}^n v_i\mathbf{e}_i\right) = \sum_{i=1}^n v_i\mathbf{a}_i$ and $P_2\mathbf{v} = \sum_{i=1}^n v_i\mathbf{b}_i$. □

COROLLARY 3.2. *If $\mathbf{w} \in \mathbb{C}^n$ then $P_1\mathbf{w} \in E_1$ and $P_2\mathbf{w} \in E_2$ and $\mathbf{w} = P_1\mathbf{w} + P_2\mathbf{w}$.*

PROOF. The columns of P_1 are in E_1 so $R(E_1) \subset E_1$ and similarly $R(P_2) \subset E_2$. Now write $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$ with $\mathbf{w}_1 \in E_1$ $\mathbf{w}_2 \in E_2$. We see that $P_1\mathbf{w} = P_1(\mathbf{w}_1 + \mathbf{w}_2) = P_1\mathbf{w}_1 + P_1\mathbf{w}_2 = \mathbf{w}_1$ by the proceeding lemma. Similarly, $P_2\mathbf{w} = \mathbf{w}_2$. □

Notice that this proof is not constructive so we don't have a good way to find P.

4. Projectors

It is useful to formalize the notion of a projector as a property inherent to a matrix. This allows us to more easily answer and manipulate questions about complementary subspaces. To that end, we say a $n \times n$ matrix P is an *oblique projector* (or just *projector*) if $P^2 = P$ (such a property is often called begin idempotent). Notice that Lemma 3.1 implies that the matrix P_1 is a projector in this sense.

Once you have a projector P, you can get a different projector $Q = I - P$ called the *complementary projector*. We check this as follows: $Q^2 = (I - P)^2 = I - P - P + P^2 = I - P = Q$. Note that P is then the complementary projector of Q. A useful fact is that $N(P) = R(Q)$ and $N(Q) = R(P)$. Check this: $\mathbf{w} = Q\mathbf{v} = \mathbf{v} - P\mathbf{v}$ then $P(\mathbf{v}) - P^2(\mathbf{v}) = 0$. So $R(Q) \subset N(P)$. On the other hand if $P\mathbf{v} = 0$ Then $Q\mathbf{v} = \mathbf{v} - P\mathbf{v} = \mathbf{v}$. Note if P_1 is the projector of the proceeding section then P_2 is the complementary projector.

As we saw given a pair of complementary spaces E_1 and E_2 we obtain complementary projectors P_1 and P_2. The converse is also true, namely, suppose that we are given a projector P an $n \times n$ matrix and let Q be the complementary projector. The previous fact allows us to see immediately that $R(P)$ and $R(Q)$ are complementary subspaces of $\mathbb{C}^n$. Indeed, for any vector $\mathbf{w} \in \mathbb{C}^n$ we have $\mathbf{w} = P\mathbf{w} + Q\mathbf{w}$. To check this, we need to check that any vector $\mathbf{w}$ can be written as the sum of avector in the range of P and a vector in the range of Q and that this is unique. The first is obvious $P\mathbf{w} + Q\mathbf{w} = P\mathbf{w} + (I - P)\mathbf{w} = \mathbf{w}$. The second uses our fact.

I.e. if $\mathbf{w} \in R(P) \cap R(Q)$ then $\mathbf{w} \in R(Q)$. But then by above $\mathbf{w} \in N(P)$. But $\mathbf{w} \in R(P)$ so $\mathbf{w} = P\mathbf{v}$ so $0 = P\mathbf{w} = P^2\mathbf{v} = P\mathbf{v} = \mathbf{w}$.

That is given a projector we obtain a pair of complementary subspaces for which the projector tells us how to decompose.

5. Orthogonal Projectors Revisited

We now return to the concept of *Orthogonal Projector*. We say a projector is orthogonal provided the $R(P)$ and $R(Q)$ are orthogonal subspaces, i.e. are orthogonal complements. It is important to note that orthogonal projectors are NOT orthogonal or unitary matrices.

They are however hermitian matrices and in fact this characterizes them. That is we have the following

THEOREM 5.1. *A projector P is an orthogonal projector if and only if $P^* = P$.*

PROOF. ($\Leftarrow$) Let $Q = I - P$ be complementary projector. If $\mathbf{w} \in R(Q)$ then $\mathbf{w} = Q\mathbf{v} = \mathbf{v} - P\mathbf{v}$ for some $\mathbf{v}$. Now let $\mathbf{a} \in R(P)$ so $\mathbf{a} = P\mathbf{b}$. Then $\langle \mathbf{w}, \mathbf{a} \rangle = \langle \mathbf{v} - P\mathbf{v}, P\mathbf{b} \rangle = \langle P^*\mathbf{v} - P^*P\mathbf{v}, \mathbf{b} \rangle$ Now using $P = P^*$ this gives $\langle P\mathbf{v} - P^2\mathbf{v}, \mathbf{b} \rangle = \langle 0, \mathbf{b} \rangle = 0$. ($\Rightarrow$) In order to show this we must use the fact that any $E \subset \mathbb{C}^n$ a vector space admits an orthonormal basis. We will show this in a couple of lectures. We know that $R(P)$ and $R(Q)$ are orthogonal compliments. Let $\mathbf{x}_1, \ldots, \mathbf{x}_k$ be an orthonormal basis of P and let $\mathbf{x}_{k+1}, \ldots, \mathbf{x}_n$ be an orthonormal basis of $R(Q)$. Notice that then $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is then an orthonormal basis of $\mathbb{C}^n$. Now, $P\mathbf{x}_j = 0$ for $k+1 \leq j \leq n$ as such $\mathbf{x}_j \in R(Q) = Null(P)$. On the other hand $P\mathbf{x}_j = \mathbf{x}_j$ for $1 \leq j \leq q$. This is because $\mathbf{x}_j = P\mathbf{x}'_j$ but then $P\mathbf{x}_j = P^2\mathbf{x}'_j = P\mathbf{x}'_j = \mathbf{x}_j$. Thus, in terms of the basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ P looks pretty nice. Let $X = \begin{bmatrix} \mathbf{x}_1 | & \cdots & |\mathbf{x}_n \end{bmatrix}$ be the $n \times n$ matrix with columns $\mathbf{x}_i$. Notice that X is unitary. Given a general vector $\mathbf{v}$, $\mathbf{v} = \sum_j c_j \mathbf{x}_j$ where $\mathbf{c} = X^{-1}\mathbf{v} = X^*\mathbf{v}$. In particular, $P\mathbf{v} = P(\sum_j c_j \mathbf{x}_j) = XId_k\mathbf{c} = XId_kX^*\mathbf{v}$. Here Id_k is matrix with 1 along diagonal for first k rows and then 0s elsewhere. In other words $P = XId_kX^*$. Then $P^* = (XId_kX^*)^* = (X^*)^*Id_k^*X^* = XId_kX^* = P$. □

One final remark: Given a subspace E, there are lots of complementary subspaces. These correspond to different oblique projectors P with $R(P) = E$. However, it is not too hard to see that there is only one orthogonal projector $P^\perp$ with $R(P^\perp) = E$. Equivalently, there is only one complementary subspace that is orthogonal.

CHAPTER 10

Eleventh and Twelfth Lectures

In this lecture I started talking about the four fundamental spaces associated to a matrix.

1. Orthogonal Complement

One bit of notation I do want to introduce. Given $E \subset \mathbb{C}^n$ a vector space I will denote by

$$E^{\perp} = \{\mathbf{v} \in \mathbb{C}^n : \langle \mathbf{v}, \mathbf{w} \rangle = 0, \mathbf{w} \in E\}$$

this is the orthogonal complement of E. It is clear that $E^{\perp}$ is a vector space and that E and $E^{\perp}$ are orthogonal. We claim also that $E + E^{\perp} = \mathbb{C}^n$, that is $E, E^{\perp}$ are orthogonal complements. We also claim that if P is an orthogonal projector with $R(P) = E$ then $E^{\perp} = R(I - P)$. Similarly, given E there is exatly one orthogonal projector P so that $R(P) = E$ and then $E^{\perp} = R(I - P)$.

2. Four fundamental spaces of a matrix

Let A be a $m \times n$ matrix. We then have two natural vector spaces associated to A. Namely $N(A) \subset \mathbb{C}^n$ and $R(A) \subset \mathbb{C}^m$. The null space and column space of A. Notice that it is important to think of these as being in *different* spaces (even if $m = n$). We now introduce two more important subspaces associated to A as we will see these turn out to be orthogonal complements of the original two.

The first of these is the *row space* of A. We denote this by $Row(A) \subset \mathbb{C}^n$ and we set $Row(A) := R(A^*)$. Notice this is essentially the span of the rows, however we have taken an adjoint in order to make the rows vectors. The second of these is called the *Left Null Space* and is denoted by $L - Null(A) \subset \mathbb{C}^m$ and we set $L - Null(A) := N(A^*)$.

Notice that the row space sits in $\mathbb{C}^n$ along with $N(A)$, while the left null space sits in $\mathbb{C}^m$ along with $R(A)$. We justify the terminology for left null space as follows: Basically it consists the rows which when multipliel agains A on the left give 0 (using the adjoint to turn rows into vectors).

It turns out that $Row(A) = R(A^*)$ and $N(A)$ are orthogonal complements in $\mathbb{C}^n$ and $L - Null(A)$ and $R(A)$ are orthogonal complements in $\mathbb{C}^m$. That is $Row(A) = N(A)^{\perp}$ and $L - Null(A) = R(A)^{\perp}$. Taken together all four spaces are known as the four fundamental spaces of the matrix A.

To prove this it suffices to restrict attention to $N(A)$ and $R(A^*)$. As we will see this is enought. We first verify these two spaces are orthogonal: Take $\mathbf{v} \in R(A^*)$ and $\mathbf{w} \in N(A)$. So $\mathbf{v} = A^*\mathbf{v}'$ for $\mathbf{v}' \in \mathbb{C}^m$. Then $\langle \mathbf{v}, \mathbf{w} \rangle = \langle A^*\mathbf{v}', \mathbf{w} \rangle = \langle \mathbf{v}', A\mathbf{w} \rangle = \langle \mathbf{v}', 0 \rangle = 0$.

In order to complete the claim we must still show that $R(A^*) + N(A) = \mathbb{C}^n$. To do this we will actually show something about the dimension of these spaces.

Namely, $dimR(A^*)+dimN(A) = n$ i.e. the dimension of the row space is the same as the dimension of the null space.

To see this we use Gaussian elimination. The point is that for each *column operation* we can do on A there is a corresponding *row operation* we can do on A^*. More precisely, suppose one gets $B \in \mathbb{C}^{m\times n}$ from A by a column operation. Then one gets B^* from A^* by a row operation.

As an example consider

$$A = \begin{bmatrix}\mathbf{a}_1| & \cdots & |\mathbf{a}_n\end{bmatrix}$$

and let

$$B = \begin{bmatrix}\mathbf{a}_1+\mathbf{a}_2| & \cdots & \mathbf{a}_n\end{bmatrix}$$

(i.e adding second column to the first) then

$$A^* = \begin{bmatrix}\mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_n^*\end{bmatrix}$$

and

$$B^* = \begin{bmatrix}(\mathbf{a}_1+\mathbf{a}_2)^* \\ \vdots \\ \mathbf{a}_n^*\end{bmatrix} = \begin{bmatrix}\mathbf{a}_1^*+\mathbf{a}_2^* \\ \vdots \\ \mathbf{a}_n^*\end{bmatrix}$$

which is adding second row to the first. Similarly row operations on A become row column operations on A^*.

A consequence of this fact is that $(rref(A))^* = cref(A^*)$. That is:

LEMMA 2.1. *Let $A \in \mathbb{C}^{m\times n}$ then $(rref(A))^* = cref(A^*)$ and $(cref(A))^* = rref(A^*)$.*

PROOF. It is straightforward to check that if a matrix B is in row reduced echelon form (rref) then B^* is in column reduced echelon form (cref). (Go back to the definition to convince yourself). Since hence $rref(A)^*$ is in cref and since $rref(A)$ is obtained from A by a finite number of row operations, $rref(A)^*$ is obtained from A^* by a finite number of column operations. By the uniqueness of $cref(A^*)$ we then see that $cref(A^*) = rref(A)$. □

As a consequence fo this, the number of pivots in $rref(A)$ is the same as the number of pivots of $cref(A*)$ and number of pivots of $cref(A)$ is same as number of pivots of $rref(A*)$. An important fact we have already used was that for an arbitrary $m \times n$ matrix B, $dimR(B)$ was the number of pivots (say k) of $cref(B)$. Similarly, if the number of pivots of $rref(B)$ is l then $dimN(A) = n - l$. Hence $dimN(A)$ is $n - k$ where k is the number of pivots of $rref(A)$. However, $rref(A)^*$ has the same number of pivots as $cref(A^*)$ and so we have $dimR(A^*) = k$. Hence $dimN(A) + dimRow(A) = n$ as claimed.

We can now show that $R(A^*)$ and $N(A)$ are orthogonal complements. Notice we've already shown they are orthogonal. Pick a basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of $R(A^*)$ and a basis $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ of $N(A)$. Notice the numbers of vectors is right as $dimRow(A)+dimN(A) = n$. We claim $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis. Its enough to check that it is linearly independent. Suppose that $\sum_{i=1}^n c_i\mathbf{v}_i = 0 = \sum_{i=1}^k c_i\mathbf{v}_i + \sum_{i=k+1}^n c_i\mathbf{v}_i$. But then (by uniqueness of the decomposition) $\sum_{i=1}^k c_i\mathbf{v}_i = 0$ and $\sum_{i=k+1}^n c_i\mathbf{v}_i$. Then by linear independence in $R(A^*)$ and $N(A)$ $c_i = 0$ for all i.

We now can conclude that $L-Null(A)$ and $R(A)$ are orthogonal complements. To see this it is enough to notice that $R(A) = Row(A^*)$ and $L-Null(A) = N(A^*)$. And use what we already showed only for A^* instead of A.

3. Relations amongst the Fundamental Spaces

We can now get useful relationships between the sizes of the fundamental spaces of A.

THEOREM 3.1. *Let $A \in Cx^{m\times n}$ then $dimR(A) = dimR(A^*)$ i.e the dimension of the row space is the same as that of the column space.*

PROOF. Pick $\mathbf{v}_1, \ldots, \mathbf{v}_k$ a basis of $R(A^*)$ and $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ a basis of $N(A)$. As we saw above the set $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of $\mathbb{C}^n$. Now let $\mathbf{w}_i = A\mathbf{v}_i$. Notice that $\mathbf{w}_i = 0$ for $k+1 \leq i \leq n$. We claim however, that for $1 \leq i \leq k$, the $\mathbf{w}_i$ form a basis of $R(A)$. Lets check they are linearly independent. Suppose $0 = \sum_{i=1}^k c_i\mathbf{w}_i = A\sum_{i=1}^k c_i\mathbf{v}_i$. Hence $\mathbf{v} = \sum_{i=1}^k c_i\mathbf{v}_i \in N(A)$. But $\mathbf{v} \in R(A^*)$ (by our set up) so must have $\mathbf{v} = 0$. However, as the $\mathbf{v}_i$ are a basis, all the $c_i = 0$. Let's check they span $R(A)$. Pick $\mathbf{w} \in R(A)$. Write $\mathbf{w} = R\mathbf{v}$. Now as $N(A)$ and $R(A^*)$ are complementary in $\mathbb{C}^n$ we can write $\mathbf{v} = \mathbf{a} + \mathbf{b}$ where $\mathbf{a} \in N(A)$ and $\mathbf{b} \in R(A^*)$. Then $\mathbf{w} = A\mathbf{v} = A(\mathbf{a}+\mathbf{b}) = A\mathbf{a} + A\mathbf{b} = A\mathbf{b}$. Now write $\mathbf{b} = \sum_{i=1}^k c_i\mathbf{v}_i$. Then $\mathbf{w} = A\mathbf{b} = \sum_{i=1}^k c_i\mathbf{w}_i$ so the $\mathbf{w}_i$ span $R(A)$. Thus $dimR(A) = k = dimR(A^*)$. □

COROLLARY 3.2. *Let $A \in \mathbb{C}^{m\times n}$ then $dimR(A)$ is the number of pivots in $rref(A)$.*

COROLLARY 3.3. *(Rank-Nullity Theorem) Let $A \in \mathbb{C}^{m\times n}$ then $dimR(A) + dimN(A) = n$.*

4. Other Facts about the fundamental spaces

A good example of using the four fundamental subspaces is the following fact:

PROPOSITION 4.1. *Let $A \in \mathbb{C}^{m\times n}$ then $A^*A\mathbf{v} = 0$ if and only if $A\mathbf{v} = 0$.*

PROOF. If $A\mathbf{v} = 0$ then it is clear that $A^*A\mathbf{v} = 0$. On the other hand, if $A^*A\mathbf{v} = 0$ then $A\mathbf{v}$ is in $N(A^*)$ i.e. in $L - Null(A)$. On the other hand $A\mathbf{v}$ is clearly in $R(A)$. That is $A\mathbf{v} \in L - Null(A) \cap R(A)$ but these two spaces are complements so $A\mathbf{v} = 0$. □

Let us pick out a matrix factorization from the proof of Theorem 3.1. Pick an orthonormal basis $\mathbf{v}_i$ of $\mathbb{C}^n$ so that $\mathbf{v}_1, \cdots, \mathbf{v}_k$ is an orthonormal basis of $R(A^*)$ and $\mathbf{v}_{k+1}, \cdots, \mathbf{v}_n$ is an orthonormal basis of $N(A)$ (well see why we can do this next lecture). Similarly, pick a basis of $\mathbb{C}^m$ $\mathbf{w}_j$ so that $\mathbf{w}_1, \ldots, \mathbf{w}_k$ is a orthonormal basis of $R(A)$ and $\mathbf{w}_{k+1}, \ldots, \mathbf{w}_m$ is an orthonoraml basis of $L - Null(A)$. Then

$$A = W\begin{bmatrix}\hat{A} & 0\\ 0 & 0\end{bmatrix}V^{-1} = W\begin{bmatrix}\hat{A} & 0\\ 0 & 0\end{bmatrix}V^*$$

Where $\hat{A}$ is a $k\times k$ non-singular matrix. And the values 0 specify a $k\times(n-k)$ matrix with all zero entries, a $(m-k)\times k$ matrix with all zero entries and a $(m-k)\times(n-k)$ matrix with all zero entries.

CHAPTER 11

Thirteenth Lecture

We discussed the Gram-Schmidt Orthogonalization and began discussing the QR factorization of a matrix.

1. Gram-Schmidt Orthogonalization

We've mentioned a number of times already that given a basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of $E \subset \mathbb{C}^n$ one can construct an orthonormal basis $\mathbf{q}_1, \ldots, \mathbf{q}_k$ of E. We will give you a simple algorithm for doing this. By doing so we give a proof of the existence of such a basis (since we already know every space has some basis).

There are a number of ways to do this, I'm going to start with the *classical Gram-Schmidt* procedure. This is the easiest orthogonalization procedure from a theoretical point of view, however computationally it has some problems (it is unstable, in other words the rounding errors on a computer can cause major problems).

The basic idea is to start with a given basis and to produce an orthonormal basis. The method to do so is iterative. Namely let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be a basis of $E \subset \mathbb{C}^n$. We proceed as follows: Start with $\mathbf{v}_1$ and let $E_1 = span\{\mathbf{v}_1\}$. We want to find an orthonormal basis of E_1. This is easy: set $\mathbf{q}_1 = \mathbf{v}_1/||\mathbf{v}_1||_2$. Notice that $\mathbf{v}_1 \neq 0$ (otherwise it couldn't be part of a basis. Now let $E_2 = span(\mathbf{v}_1, \mathbf{v}_2) = span(\mathbf{q}_1, \mathbf{v}_2)$. We want to find an orthonormal basis of E_2 this is a little bit harder as $\mathbf{q}_1$ and $\mathbf{v}_2$ need not be orthogonal. But notice that $P_{\mathbf{q}_1}\mathbf{v}_2 \neq \mathbf{v}_2$ and if we let $\hat{\mathbf{q}}_2 = \mathbf{v}_2 - P_{\mathbf{q}_1}\mathbf{v}_2 = P_{\perp \mathbf{q}_1}\mathbf{v}_2 = \mathbf{v}_2 - \langle \mathbf{q}_1, \mathbf{v}_2\rangle \mathbf{q}_1$ then $\hat{\mathbf{q}_2} \in E_2$ is non-zero and $\langle \mathbf{q}_1, \hat{\mathbf{q}}_2\rangle = 0$. Hence we can set $\mathbf{q}_2 = \hat{\mathbf{q}}_2/||\mathbf{q}_2||_2$. The reason this works is if $P_{\mathbf{q}_1}\mathbf{v}_2 = \mathbf{v}_2$ then one would have that $\mathbf{v}_2 \in span(\mathbf{q}_1) = span(\mathbf{v}_1)$ i.e. $\mathbf{v}_2$ and $\mathbf{v}_1$ would be linearly dependent.

Inductively, we have a method that takes $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and gives $\{\mathbf{q}_1, \ldots, \mathbf{q}_l, \mathbf{v}_{l+1}, \ldots, \mathbf{v}_k\}$ where $E_j = span\{\mathbf{q}_1, \ldots, \mathbf{q}_j\} = span\{\mathbf{v}_1, \ldots, \mathbf{v}_j\}$ and $\{\mathbf{q}_1, \ldots, \mathbf{q}_j\}$ is an orthonormal basis of E_j here $1 \leq j \leq l$. We now wish to produce $\mathbf{q}_{l+1}$ from $\mathbf{v}_{l+1}$ so that now $\mathbf{q}_1, \ldots, \mathbf{q}_{l+1}$ is a orthonormal basis for $E_{l+1} = span\mathbf{v}_1, \ldots, \mathbf{v}_{l+1}$.

To do this we again note that if we set $\hat{\mathbf{q}}_{l+1} = \mathbf{v}_{l+1} - P_{E_l}\mathbf{v}_{l+1} = \mathbf{v}_{l+1} - \sum_{j=1}^{l}\langle \mathbf{q}_j, \mathbf{v}_{l+1}\rangle \mathbf{q}_j$. Then $\hat{\mathbf{q}}_{l+1}$ is non-zero and orthogonal to each $\mathbf{q}_j$ $1 \leq j \leq l$. Setting $\mathbf{q}_{l+1} = \hat{\mathbf{q}}_{l+1}/||\hat{\mathbf{q}}_{l+1}||_2$. Then provides the iteative step. Again we have used the fact that the $\mathbf{v}_i$ are linearly independent to ensure that $\hat{\mathbf{q}}_{l+1} \neq 0$.

Iterating this k times produces the desired $\mathbf{q}_1, \ldots, \mathbf{q}_k$.

We can write this algorithm in pseudo-code as:

For $j = 1$ to k $\mathbf{a}_j = \mathbf{v}_j$
    For $i = 1$ to $j - 1$
        $r_{ij} = \mathbf{q}_i^* \mathbf{v}_j$
        $\mathbf{a}_j = \mathbf{a}_j - r_{ij}\mathbf{q}_i$
    $r_{jj} = ||\mathbf{a}_j||_2$ $\mathbf{q}_j = \mathbf{a}_j/r_{jj}$

Where this has numerical problems is when the $\mathbf{v}_i$ are close to parallel.

2. The QR factorization

We are now going to apply this idea of orthogonalization to a matrix. The idea is to look at a matrix $A \in \mathbb{C}^{m\times n}$ and try and get an orthonormal basis for the column space of A. But we are actually going to be more careful than that.

Consider the columns of A so we have

$$A = \begin{bmatrix}\mathbf{a}_1| & \cdots & |\mathbf{a}_n\end{bmatrix}$$

We get a whole sequence of spaces $E_1 = span(\mathbf{a}_1), E_2 = span(\mathbf{a}_1, \mathbf{a}_2), \ldots, E_n = span(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. So $E_1 \subset E_2 \subset \ldots \subset E_n = R(A)$. What we want to do is in sense get an orthonormal basis for all of these subsets. That is find an orthonormal set $\mathbf{q}_1, \ldots, \mathbf{q}_k$ so that $E_1 = span(\mathbf{q}_1), E_2 = span(\mathbf{q}_1, \mathbf{q}_2), \ldots, E_n = span(\mathbf{q}_1, \ldots, \mathbf{q}_k)$. Being able to do this will be equivalent to given a good factorization of the matrix A. Notice that by a dimension count we are implicity assuming that the $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly independent.

Well starting from an arbitrary $A \in \mathbb{C}^{m\times n}$ and assume for this discussion that $m \geq n$ and that A has full rank (i.e. n). This later condition implies that the columns are linearly independent. If we can find $\mathbf{q}_i \in \mathbb{C}^m$ as desired then we have

$$\mathbf{a}_1 = r_{11}\mathbf{q}_1, \mathbf{a}_2 = r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2, \ldots, \mathbf{a}_n = r_{1n}\mathbf{q}_1 + \ldots, r_{nn}\mathbf{q}_n$$

Notice this is equivalent to the matrix factorization

$$\begin{bmatrix}\mathbf{a}_1| & \cdots & |\mathbf{a}_n\end{bmatrix} = \begin{bmatrix}\mathbf{q}_1| & \cdots & |\mathbf{q}_n\end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & & & \\ & & & \\ \ddots & & \vdots & \\ 0 & \cdots & 0 & r_{nn}\end{bmatrix}$$

That is

$$A = \hat{Q}\hat{R}$$

where $\hat{Q} \in \mathbb{C}^{m\times n}$ has columns that are the orthonormal vectors $\mathbf{q}_i$ for $1 \leq i \leq n$ and $\hat{R} \in \mathbb{C}^{n\times n}$ is upper triangular. This is called the *reduced QR factorization*

For certain purposes it is convenient to have a different form of the factorization. That is we want to replace the $\hat{Q}$ term by a unitary term Q. As $m \geq n$, we can do this by adding extra elements to Q that complete the columns fo Q to an orthonormal basis of $\mathbb{C}^m$. Namely let $\mathbf{q}_{n+1}, \ldots, \mathbf{q}_m \in \mathbb{C}^m$ be an orthonormal basis of $R(A)^\perp = L - N(A)$. Then one has that $\mathbf{q}_1, \ldots, \mathbf{q}_m$ is an orthonormal basis of $\mathbb{C}^m$ so in particular

$$Q = \begin{bmatrix}\mathbf{q}_1| & \cdots & |\mathbf{q}_m\end{bmatrix} \in \mathbb{C}^{m\times m}$$

Is unitary.

Then setting

$$R = \begin{bmatrix}\hat{R} \\ 0\end{bmatrix},$$

so $R \in C^{m\times n}$ is still upper triangular we obtain the full or unreduced QR factorization as

$$A = QR.$$

Notice that the span of the "silent" columns in the full QR factorization are precisely an orthonormal basis of $R(A)^{\perp} = L - Null(A)$.

One important geometric interpretation the full QR factorization allows is the following: The range of the matrix R is precisely the n-dimensional subspace of $\mathbb{C}^m$ where the last $m - n$ entries are zero. For instance if $m = 3$ and $n = 2$ then the range of R is exactly the plane with third component zero. The matrix Q then acts as a sort of "rotation" which allows us to obtain all other n-dimensional subspaces. One way to think of this is that the Q tells us where $R(A)$ sits in $\mathbb{C}^m$ while R tells us how vectors in $R(A)$ and in $\mathbb{C}^n$ are identified.

CHAPTER 12

Fourteenth Lecture

We introduced the QR factorization in the last lecture. We discuss it in a bit more depth.

1. The QR factorization

Recall, the QR factorization worked by starting with a matrix $A \in \mathbb{C}^{m\times n}$ where $m \geq n$ and A with full rank (i.e. $\dim R(A) = n$). We write

$$A = \begin{bmatrix}\mathbf{a}_1| & \cdots & |\mathbf{a}_n\end{bmatrix}.$$

The *reduced QR factorization* is a factorization:

$$A = \hat{Q}\hat{R}$$

where $\hat{Q} \in \mathbb{C}^{m\times n}$ has columns

$$\hat{Q} = \begin{bmatrix}\mathbf{q}_1| & \cdots & |\mathbf{q}_n\end{bmatrix}$$

where $\{\mathbf{q}_i\} \in \mathbb{C}^m$ are an orthonormal set of vectors and $\hat{R} \in \mathbb{C}^{n\times n}$ is upper triangular. It is straight forward to verify that $span(\mathbf{a}_1, \cdots, \mathbf{a}_k) = span(\mathbf{q}_1, \cdots, \mathbf{q}_k)$ for $1 \leq k \leq n$.

For certain purposes it is convenient to have the so called *full QR factorization* Here

$$A = QR$$

where

$$Q = \begin{bmatrix}\mathbf{q}_1| & \cdots & |\mathbf{q}_n & \mathbf{q}_{n+1}| & \cdots & |\mathbf{q}_m\end{bmatrix}$$

is now in $\mathbb{C}^{m\times m}$ and is unitary. We then have $R \in \mathbb{C}^{m\times n}$ still upper triangular. Notice that then the bottom rows must be all zero then. The additional vectors $\mathbf{q}_{n+1}, \ldots, \mathbf{q}_m$ are "silent" and are arbitrary as long as they an orthonormal basis of $R(A)^\perp = L - Null(A)$.

We are also interested in the case where A does not have full rank. In this case there is still a QR factorization. We just have to modify our algorithm a bit.

THEOREM 1.1. *Every $A \in \mathbb{C}^{m\times n}$ with $(m \geq n)$ has a full QR factorization.*

PROOF. We will actually construct a reduced QR factorization of A and then complete it to a full QR factorization as needed. The proof is just the Gram-Schmidt algorithm. However, we can no longer ensure that the columns of A are linearly independent. In particular it may happen that $span(\mathbf{a}_1, \ldots, \mathbf{a}_j) = span(\mathbf{a}_1, \ldots, \mathbf{a}_{j+1})$.

More precisely. Start with $\mathbf{a}_1$ if $\mathbf{a}_1 = 0$ choose $\mathbf{q}_1$ an arbitrary unit vector and in this case take $r_{11} = 0$. If $\mathbf{a} \neq 0$ set $\mathbf{q}_1 = \mathbf{a}_1/||\mathbf{a}_1||_2$ take $r_{11} = ||\mathbf{a}_1||_2$. Notice in both cases:

$$\mathbf{a}_1 = r_{11}\mathbf{q}_1.$$

Now proceed inductively: That is suppose we've gotten $\mathbf{q}_1, \ldots, \mathbf{q}_k$ from $\mathbf{a}_1, \ldots, \mathbf{a}_k$. Notice that in this case $span(\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}) \supset span(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1})$ We want to find $\mathbf{q}_j$. To do so, we (as before) set $r_{ij} = \langle \mathbf{a}_j, \mathbf{q}_j \rangle$ and $\hat{\mathbf{q}}_j = \mathbf{a}_j - \sum_i = 1^{j-1} r_{ij}\mathbf{q}_j$. By the bilinearity of the inner product $\langle \hat{\mathbf{q}}_j, \mathbf{a}_j \rangle = 0$. If $\hat{\mathbf{q}}_j = 0$ we take $r_{jj} = 0$ and pick $\mathbf{q}_j$ to be any unit vector orthogonal to $\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}$ otherwise set $r_{jj} = ||\hat{\mathbf{q}}_j||_2$ and $\mathbf{q}_j = r_{jj}^{-1}\hat{\mathbf{q}}_j$. Then

$$\mathbf{a}_j = \sum_{i=1}^{j} r_{ij}\mathbf{q}_j$$

Hence, if we set

$$\hat{Q} = \begin{bmatrix} \mathbf{q}_1 | & \cdots & \mathbf{q}_n \end{bmatrix}$$

and

$$\hat{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix}$$

we obtain a reduced QR factorization of A. To get the full QR factorization we can add the silent columns as before. That is we find $\mathbf{q}_{n+1}, \ldots, \mathbf{q}_m$ forming an orthonormal basis of $span(\mathbf{q}_1, \ldots, \mathbf{q}_m)$. □

Remark 1.2. Notice that $R(A) \subset span(\mathbf{q}_1, \ldots, \mathbf{q}_m)$ with equality only when A has full rank. One consequence is that the silent columns $\mathbf{q}_{n+1}, \ldots, \mathbf{q}_m$ while lying in $L - Null(A)$ no longer need to form a basis. Another consequence is that A has full rank when and only when $r_{ii} \neq 0$ for all $i = 1, \cdots, n$.

Remark 1.3. One may wonder how to find $\mathbf{q}_j$ when $r_{jj} = 0$. Notice that we want $\mathbf{q}_j$ to be orthogonal to each $\mathbf{q}_i$ for $1 \leq i \leq j - 1$. That is if we set

$$\hat{Q}_{j-1} = \begin{bmatrix} \mathbf{q}_1 | & \cdots & \mathbf{q}_{j-1} \end{bmatrix} \in \mathbb{C}^{m \times (j-1)}$$

we need $\hat{Q}_{j-1}\mathbf{q}_j = 0$ and $||\mathbf{q}_j||_2 = 1$. Another way to think about this that $\mathbf{q}_j \in \hat{Q}_{j-1}^*$) so $\mathbf{q}_j$ can be found by Gaussian elimination (though there is likely a more efficient algorithm).

The full QR factorization tends not to be unique. This is because silent columns are not specified by the algorithm. While this is not an issue with the reduced QR factorization. There is still non-uniqueness in this case. To see this, note one can multiply the ith column of $\hat{Q}$ by some $\lambda \in \mathbb{C}$ so that $|\lambda| = 1$ and get a new matrix $\hat{Q}'$ which still has orthonormal columns, if one multiplies the ith row of $\hat{R}$ by λ^{-1} to get $\hat{R}'$ then this is still upper triangular and $\hat{Q}\hat{R} = \hat{Q}'\hat{R}'$. This corresponds to the arbitrary choice that one makes in the Gram-Schmidt algorithm.

However, there is uniqueness if A is full rank and one demands $\hat{R}$ have a special form.

Theorem 1.4. *For each $A \in Cx^{m \times n}$ with $m \geq n$ and so that A has full rank there is a unique reduced QR factorization*

$$A = \hat{Q}\hat{R}$$

So that the diagonal entries of $\hat{R}$ are positive real numbers (i.e. $r_{ii} > 0$).

PROOF. If we look at the proof of the preceeding theorem we see that everything is determined except the"sign" of r_{ii}. If we insist that $r_{ii} > 0$ then we are done. □

2. Solving a System via QR factoriation.

One thing the QR factorization allows us to do is to solve systems. Let $A \in \mathbb{C}^{m\times m}$ be non-singular matrix (i.e. of full rank). And fix $\mathbf{b} \in \mathbb{C}^m$. We want to solve

$$A\mathbf{x} = \mathbf{b}$$

The usual way is to use Gaussian elimination, which in a sense is a better approach to this specific problem. First theroem of this lecture there is a QR factorization of A and as $m = n$ the reduced is the same as the full so we write

$$A = QR$$

Here Q is unitary and R is upper triangular with no non-zero entries on the diagonal. This latter fact follows as A is non-singular.

Hence, one has

$$R\mathbf{x} = Q^*\mathbf{b}$$

Now we are solving a system of equations where the system consists of an upper triangular matrix. This can easily be solved by back-substitution.

An example: Let

$$A = \begin{bmatrix} 0 & -3 & 0 \\ 0 & 4 & 1 \\ 4 & 0 & 1 \end{bmatrix}$$

And lets solve

$$A\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

To start

$$\mathbf{a}_1 = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} \Rightarrow \mathbf{q}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

and so $r_{11} = 4$. Now $\langle \mathbf{a}_2, \mathbf{q}_1 \rangle = 0$ and so $r_{12} = 0$ and

$$\mathbf{q}_2 = \begin{bmatrix} -3/5 \\ 4/5 \\ 0 \end{bmatrix}$$

and $r_{22} = 5$. Finally, $\langle \mathbf{a}_3, \mathbf{q}_1 \rangle = 1$ and $\langle \mathbf{a}_3, \mathbf{q}_2 \rangle = 4/5$ so $r_{13} = 1$ and $r_{23} = 4/5$ thus

$$\hat{\mathbf{q}}_3 = \begin{bmatrix} 12/25 \\ 9/25 \\ 0 \end{bmatrix}$$

so $r_{33} = 3/5$ and

$$\mathbf{q}_3 = \begin{bmatrix} 4/5 \\ 3/5 \\ 0 \end{bmatrix}$$

Hence

$$A = \begin{bmatrix} 0 & -3/5 & 4/5 \\ 0 & 4/5 & 3/5 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 & 1 \\ 0 & 5 & 4/5 \\ 0 & 0 & 3/5 \end{bmatrix}$$

And so

$$R\mathbf{x} = A^*\mathbf{b} = \begin{bmatrix} 0 & 0 & 1 \\ -3/5 & 4/5 & 0 \\ 4/5 & 3/5 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -3/5 \\ 4/5 \end{bmatrix}$$

Then

$$\begin{bmatrix} 4 & 0 & 1 \\ 0 & 5 & 4/5 \\ 0 & 0 & 3/5 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ -3/5 \\ 4/5 \end{bmatrix}$$

So $x_3 = 4/3$, $5x_2 = -3/5 - 16/15 = -5/3$ so $x_2 = -1/3$. Then $4x_1 = -4/3$ so

$$\mathbf{x} = \begin{bmatrix} -1/3 \\ -1/3 \\ 4/3 \end{bmatrix}$$

CHAPTER 13

Fifteenth Lecture

We use the QR factorization to study a problem about solve overdetermined systems of linear equations "approximately"..

1. Least Squares Method

Recall we say that a system of linear equations is overdetermined if there are more equations then unknowns. That is one has $A \in \mathbb{C}^{m\times n}$ with $m > n$ and look at

$$A\mathbf{x} = \mathbf{b}$$

for a fixed $\mathbf{b} \in \mathbb{C}^m$. By the rank-nullity theorem $dimR(A) \leq n < m$ so for "most" $\mathbf{b}$ this equation has no solution.

In this case what we do is study the so called *residual*

$$\mathbf{r} = \mathbf{b} - A\mathbf{x} \in \mathbb{C}^m$$

The idea is to try and find the $\mathbf{x}$ that makes the residual as small as possible.

In order to do this we need to have a notion of "size" for vectors. We will discuss this more later but for now we take the 2-norm, that is we try and minimize:

$$||\mathbf{b} - A\mathbf{x}||_2.$$

That is we try and find $\mathbf{x}$ so that $\mathbf{r}$ has least length, in other words so$A\mathbf{x}$ is the closest vector in $R(A)$ to $\mathbf{b}$. This turns out to be natural from both geometric point of view and from more experience. It also has the advantage of being mathematically and algorithmically very tractable.

So how do we find the $\mathbf{x}$ that minimizes the residual? It turns out that there is a nice characterization in terms of linear algebra that we have already developed:

THEOREM 1.1. *Let $A \in \mathbb{C}^{m\times n}$ $(m \geq n)$ and $\mathbf{b} \in \mathbb{C}^m$. A vector $\mathbf{x} \in \mathbb{C}^n$ minimizes the residual $||\mathbf{r}||_2 = ||\mathbf{b} - A\mathbf{x}||_2$ if and only if $\mathbf{r}$ is orthogonal to $R(A)$ that is $\mathbf{r} \in L - Null(A)$ (i.e. $A^*\mathbf{r} = 0$).*

PROOF. $\Rightarrow$ By our hypothesis for any $\mathbf{y} \in R(A)$, and $t \in \mathbb{R}$ for $t \neq 0$ then $||\mathbf{b} - (A\mathbf{x} + t\mathbf{y})||_2 \geq ||\mathbf{r}||_2$. We can square both sides so obtain:

$$\begin{aligned} ||\mathbf{b} - A\mathbf{x} - t\mathbf{y}||_2^2 &\geq ||\mathbf{r}||_2^2 \\ ||\mathbf{r} - t\mathbf{y}||_2^2 &\geq ||\mathbf{r}||_2^2 \\ ||\mathbf{r}||_2^2 - t(\langle \mathbf{r}, \mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{r}\rangle + t^2||\mathbf{y}||_2^2 &\geq ||\mathbf{r}||_2^2 \end{aligned}$$

Here the last line follows by expanding out the inner product. Thus, after dividing by t (since it is not 0)

$$-(\langle \mathbf{r}, \mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{r}\rangle) + t||\mathbf{y}||_2^2 \geq 0.$$

By letting $t \to 0$ get

$$-(\langle \mathbf{r}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{r} \rangle) \geq 0$$

Notice that by replacing $\mathbf{y}$ by $-\mathbf{y}$ be get

$$-(\langle \mathbf{r}, -\mathbf{y} \rangle + \langle -\mathbf{y}, \mathbf{r} \rangle) \geq 0$$
$$\langle \mathbf{r}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{r} \rangle \geq 0$$

So $\langle \mathbf{r}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{r} \rangle = (\langle \mathbf{r}, \mathbf{y} \rangle + \overline{\langle \mathbf{r}, \mathbf{y} \rangle} = 0$ this means $\langle \mathbf{r}, \mathbf{y} \rangle$ is purely imaginary. By replacing $\mathbf{y}$ by $\pm\mathbf{y}$ be get

$$-(\langle \mathbf{r}, \pm I\mathbf{y} \rangle + \langle \pm I\mathbf{y}, \mathbf{r} \rangle) \geq 0$$
$$-(\pm I\langle \mathbf{r}, \mathbf{y} \rangle \mp I\langle \mathbf{y}, \mathbf{r} \rangle) \geq 0$$
$$\mp I(\langle \mathbf{r}, I\mathbf{y} \rangle - \langle I\mathbf{y}, \mathbf{r} \rangle) \geq 0$$

This implies $\langle \mathbf{r}, \mathbf{y} \rangle$ is purely real and hence combining with the above is 0.

$\Leftarrow$. We need to show that if $\mathbf{r}$ is orthogonal to $R(A)$ then for any point $\mathbf{y} \in R(A)$ one has $||\mathbf{b} - \mathbf{y}||_2 \geq ||\mathbf{r}||_2$. To do so we note that $A\mathbf{x} - \mathbf{y} \in R(A)$ and $\mathbf{b} - A\mathbf{x} = \mathbf{r}$ is orthogonal to this so

$$||\mathbf{b} - \mathbf{y}||_2^2 = ||\mathbf{b} - A\mathbf{x} + A\mathbf{x} - \mathbf{y}||_2^2 = ||\mathbf{b} - A\mathbf{x}||_2^2 + ||A\mathbf{x} - \mathbf{y}||_2^2 \geq ||\mathbf{r}||_2^2$$

Here we used the Pythagorean theorem. □

A useful consequence of this theorem is then: Let $P \in \mathbb{C}^{m \times m}$ be a orthogonal projection onto $R(A)$. Then $\mathbf{r} = \mathbf{b} - A\mathbf{x}$ minimizes the residual norm if and only if $\mathbf{x}$ solves

$$A\mathbf{x} = P\mathbf{b}.$$

which we know has at least one solution (since $P\mathbf{b} \in R(A)$). Notice that $\mathbf{x}$ is unique when only when $N(A) = \{0\}$ i.e. if A has full rank. We will usually assume this.

One other way to think about this is as a *right approximate inverse* Idea is let $A \in \mathbb{C}^{m \times n}$ with $m \geq n$ and A of full rank. For each $\mathbf{e}_i$ a standard basis vector of $\mathbb{C}^m$ let $\mathbf{b}_i$ be the unique vector in $\mathbb{C}^n$ so that

$$A\mathbf{b}_i = P\mathbf{e}_i$$

and set

$$B = \begin{bmatrix} \mathbf{b}_1 | & cdots & | \mathbf{b}_m \end{bmatrix} \in \mathbb{C}^{n \times m}$$

Now $AB = P$. In other words, if we let Q be the compelementary projector to P (so Q projects orthogonally onto $R(A)^\perp = L - Null(A)$ then

$$AB = I - Q$$

So for instance if the left null space is zero we have an actual inverse.

2. Least Squares from QR factorization

So how do we use this in practice? We need to find the the projection onto $R(A)$. The key is getting an orthonormal basis. We've used this before (but maybe not said it so clearly). Basically, let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be an orthonormal basis of $R(A)$. Then one checks that

$$P = \sum_{i=1}^{k} \mathbf{v}_i \mathbf{v}_i^* = VV^* \in \mathbb{C}^{m \times m}$$

gives orthogonal projection onto $R(A)$. It suffices to verify that $P^2 = P$, $P^* = P$ and that $R(P) = R(A)$. I leave this as an excersize.

Thus we need to find an orthonormal basis of $R(A)$. The QR factorization provides a good way to do this. To make this work we need to work with full rank A in $\mathbb{C}^{m\times n}$ ($m \geq n$). If we take the reduced QR factorization ie.

$$A = \hat{Q}\hat{R}$$

then if we set

$$P = \hat{Q}\hat{Q}^*$$

then $P \in \mathbb{C}^{m\times m}$ is orthogonal projection onto $R(A)$. (Recall the columns of $\hat{Q}$ are an orthonormal basis of $R(A)$). Notice if A is not full rank, then we can't ensure that the columns of $\hat{Q}$ are not necessarily inside of $R(A)$. This is one reason to start with A of full rank. So we are solving

$$\hat{Q}\hat{R}\mathbf{x} = P\mathbf{b} = \hat{Q}\hat{Q}^*\mathbf{b}.$$

That is

$$\hat{Q}\hat{R}\mathbf{x} - P\mathbf{b} = \hat{Q}\hat{Q}^*\mathbf{b} = \hat{Q}\left(\hat{R}\mathbf{x} - \hat{Q}^*\mathbf{b}\right) = 0.$$

As $\hat{Q}$ has columns which are orthonormal, the columns are all linearly independent so $N(\hat{Q}) = 0$. Thus the equation we want to solve is:

$$\hat{R}\mathbf{x} = \hat{Q}^*\mathbf{b}.$$

This yields the following algorithm for solving the problem (at least for full rank $A \in \mathbb{C}^{m\times n}$):

(1) Compute the (reduced) QR factorization $A = \hat{Q}\hat{R}$.
(2) Compute the vector $\mathbf{b}' = \hat{Q}^*\mathbf{b}$
(3) Solve the upper-triangular system $\hat{R}\mathbf{x} = \mathbf{b}'$.

Notice that (1) is the most computationally intensive step. We can do it by either the Gram-Schmidt algorithm already discussed or some other approachs we discuss in the next lecture. For the last step one uses back substitution.

3. Application:NIC

Least-Squares is used in many different contexts. I'll present one important instance, namely fitting a polynomial to data. The basic set is to start with m points $(x_1, y_1), \ldots, (x_m, y_m)$ in $\mathbb{R}^2$ (or $\mathbb{C}^2$). We assume $x_i \neq x_j$ for $i \neq j$. With out this hypthoesesis the points wouldn't lie on any graph of any function of x. We look for a polynomial P of degree $n-1$ so that $P(x_i) = y_i$.

If we write

$$P(x) = c_0 + c_1 x + \ldots + c_{n-1}x^{n-1}$$

then we are looking for $c_0, \ldots, c_{n-1}$ so that $P(x_i) = y_y$. Finding such c_i is a linear problem (even though polynomials tend to be very non-linear). Indeed, if we write:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

We see we are really solving a system of linear equations. We shorten this to

$$X\mathbf{c} = \mathbf{y}$$

Here $X \in \mathbb{C}^{m\times n}$ is the called the Vandermonde matrix.

It turns out that the condition that $x_i \neq x_j$ implies that X is full rank (Excercise!). In particular, if $m = n$ we can always find the desired P so taht $P(x_i) = y_i$.

It turns out choosing such a polynomial is less than ideal. The problem is that the graph doesn't interpolate the points well. That is, in between adjacent values x_1, x_2 the graph might become very far from y_1 and y_2. A related issue is that if the x_i and y_i are changed slightly, the approximating polynomial might change radically. Since data is noisy this is not desirable. It turns out that this issue is lessened if one uses a lower degree polynomial. That is take $m < n$. In this case one has an overdetermined system of equations so one has to look at "approximate" solutions as above.

CHAPTER 14

Sixteenth and Eighteenth Lectures

We discuss alternate methods of computing the QR factorization. These are better suited for implementation on a computer.

1. Modified Gram-Schmidt

We have seen how to compute the QR factorizations using the Gram-Schmidt algorithm and this is perfectly fine from a theoretical point of view. Practically however, the algorithm handles rounding errors very poorly (mainly an issue when the initial basis contains vectors that are nearly parallel). To see how to get around this we first give a variant of Gram-Schmidt that is better behaved.

Lets think for a moment about what Gram-Schmidt itself does. Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ be a of linearly independent vectors in $\mathbb{C}^m$. The Gram-Schmidt algorithm produces,

$$\mathbf{q}_1 = \frac{P_1\mathbf{a}_1}{||P_1\mathbf{a}_1||_2}, \mathbf{q}_2 = \frac{P_2\mathbf{a}_2}{||P_2\mathbf{a}_2||_2}, \ldots, \mathbf{q}_k = \frac{P_k\mathbf{a}_k}{||P_k\mathbf{a}_k||_2}$$

where here P_1 is the identity matrix and for $j \geq 2$ each $P_j \in \mathbb{C}^{m\times m}$ is orthogonal projection onto $span(\mathbf{q}_1, \ldots, \mathbf{q}_{j-1})^{\perp}$. As the $\mathbf{q}_i$ are an orthonormal set of vectors we can write:

$$\hat{Q}_{j-1} = \begin{bmatrix}\mathbf{q}_1| & \cdots & |\mathbf{q}_k\end{bmatrix}$$

and then $\hat{Q}_{j-1}\hat{Q}^*_{j-1}$ gives projection onto $span(\mathbf{q}_1, \ldots, \mathbf{q}_{j-1})$. Thus P_j projection onto the orthogonal complement is given by

$$P_j = I - \hat{Q}_{j-1}\hat{Q}^*_{j-1}$$

Notice that each P_j is of rank $m - (j-1)$. An important observation is that one can factorize the P_j in terms of rank $m-1$ orthogonal projectors:

$$P_j = P_{\perp\mathbf{q}_{j-1}} \cdots P_{\perp\mathbf{q}_2} P_{\perp\mathbf{q}_1} I$$

This follows by noting that $P_j = I - \sum_{i=1}^{j} \mathbf{q}_i\mathbf{q}_i^*$ and $P_{\perp\mathbf{q}_i} = I - \mathbf{q}_i\mathbf{q}_i^*$ and taking a product. I leave the details as an exercise. This gives another algorithm which is less sensitive to rounding errors.

Idea of the algorithm: Start with $\mathbf{a}_j$ set $\mathbf{v}_j^{(1)} = \mathbf{a}_j$. One iteratively computes as follows: At the ith step, set $\mathbf{v}_j = \mathbf{v}_j^{(j)}$, $r_{ii} = ||\mathbf{v}_i||_2$ and $\mathbf{q}_i = r_{ii}^{-1}\mathbf{v}_i$ and set $\mathbf{v}_j^{(i+1)} = P_{\perp\mathbf{q}_i}\mathbf{v}_j^{(i)}$.

Another way to think about this is that there are upper triangular matrices R_1, $R_2, \ldots, R_n$ (in $\mathbb{C}^{n\times n}$) so that

$$\begin{bmatrix}\mathbf{q}_1 & \cdots & |\mathbf{q}_{j-1} & |\mathbf{v}_j^{(j)}| & \cdots & |\mathbf{v}_n^{(j)}\end{bmatrix} R_j = \begin{bmatrix}\mathbf{q}_1 & \cdots & |\mathbf{q}_j & |\mathbf{v}_{j+1}^{(j+1)}| & \cdots & |\mathbf{v}_n^{(j+1)}\end{bmatrix}$$

We refer to Trefethen and Bau Lecture 8 for the exact form of the R_j. Then one has

$$AR_1R_2\dots R_n = \hat{Q}$$

One can check that the product of upper triangular matrices is still upper triangular and one of your homework exercises was to show the inverse was upper triangular so with

$$\hat{R} = (R_1R_2\dots R_n)^{-1}$$

one obtains a reduced QR factorization:

$$A = \hat{Q}\hat{R}$$

2. Householder Reflections

So we saw how to determine a (reduced) QR factorization by repeated multiplications by upper triangular matrices that in the end produces a unitary matrix. Another approach is to use unitary matrices to produce an upper triangular matrix. This is known as the Householder algorithm.

Basic idea is to write

$$A = \begin{bmatrix}\mathbf{a}_1| & \cdots & |\mathbf{a}_n\end{bmatrix}$$

we want to find a $Q_1 \in \mathbb{C}^{m\times m}$ that is unitary and so that

$$Q_1A = \begin{bmatrix}r_{11}\mathbf{e}_1 & |\mathbf{a}_2^{(2)}| & \cdots & |\mathbf{a}_n^{(2)}\end{bmatrix}$$

Then find a $Q_2 \in \mathbb{C}^{m\times m}$ that is unitary and so that

$$Q_2Q_1A = \begin{bmatrix}r_{11}\mathbf{e}_1 & |r_{12}\mathbf{e}_1 + r_{22}\mathbf{e}_2 & |\mathbf{a}_3^{(3)}| & \cdots & |\mathbf{a}_n^{(3)}\end{bmatrix}$$

and so on. The end result will be $Q_1, \dots, Q_n$ all unitary so that

$$Q_nQ_{n-1}\dots Q_1A = R$$

where $R \in \mathbb{C}^{m\times n}$ is upper triangular. We get the QR factorization by setting

$$Q = (Q_nQ_{n-1}\dots Q_1)^*$$

where we use that the product of unitary matrices is still unitary as is the adjoint of a unitary matrix. We leave this fact to you to check.

We now discuss how to find such unitary matrices. The first property that we want is for the Q_k to preserve the first $k-1$ columns of $Q_{k-1}\cdots Q_1A$. To do so we may take Q_k to be of the form

$$Q_k = \begin{bmatrix}I_{k-1} & 0\\ 0 & F\end{bmatrix}$$

where $I_{k-1} \in \mathbb{C}^{(k-1)\times(k-1)}$ is the identity and $F \in \mathbb{C}^{(m-k+1)\times(m-k+1)}$. This works as the first $k-1$ columns of $Q_{k-1}\cdots Q_1A$ are upper triangular so the F term doesn't effect those columns. For Q_k to be unitary it must have orthonormal columns and hence F must have orthonormal columns and so also be unitary.

Let $\mathbf{x} \in \mathbb{C}^(m-k+1)$ denote the vector obtained from $\mathbf{a}_k^{(k)} \in \mathbb{C}^m$ by omitting the first $k-1$ entries. One has

$$Q_k\mathbf{a}_k^{(k)} = \begin{bmatrix}*\\ F\mathbf{x}\end{bmatrix}$$

where $*$ represents $m-k-1$ entries. In particular, to to find F (and hence Q_k it suffices to ensure $F\mathbf{x} = r_{kk}\mathbf{e}_1$.

As unitary matrices preserve distance, one needs $|r_{kk}| = ||\mathbf{x}||_2$ and so we start by taking $r_{kk} = ||\mathbf{x}||_2$. One geometric way to do this would be by rotation. This is not optimal from a pratical point of view. Instead, reflection is a better choice. Namely, let $\mathbf{v}_+ = ||\mathbf{x}||_2\mathbf{e}_1 - \mathbf{x}$ and let $E_+ = span(\mathbf{v}_+)$ and $H_+ = E_+^\perp$. We can then take F to be reflection across H_+. As it will then be a unitary matrix with the desired mapping property. (Note: this geometric intrepretation works best over the reals).

Lets figure out the matrix for the reflection. Let $P_{\mathbf{v}_+}$ denote orthogonal projection onto E_+ and P_{H_+} denote orthgonal projection onto H_+, i.e they are complementary orthogonal projectors. One verifes then that $F = I - 2P_{E_+}$ is unitary and has the desired behavior. In other words:

$$F = I - 2\frac{\mathbf{v}_+\mathbf{v}_+^*}{||\mathbf{v}_+||_2^2}$$

Notice there are other choices. For instance one could try and make $F\mathbf{x} = -||\mathbf{x}||_2\mathbf{e}_1$. Here we let $\mathbf{v}_- = -||\mathbf{x}||_2\mathbf{e}_1 - \mathbf{x}$ and work as above then we are reflecting across the hyperplane H_- which is orthgonal to $\mathbf{v}_-$. Mathematically both choices are equivalent (i.e. lead to the same answer) however, numerically it turns out to be better to choose the sign so that $F\mathbf{x}$ is as far as possible from $\mathbf{x}$. It is easy to see that this is equivalent to choosing $r_{kk} = -sign(x_1)||\mathbf{x}||_2$ where x_1 is first component of $\mathbf{x}$ and $sign(x_1) = 1$ if $x_1 \geq 0$ and $= -1$ if $x_1 < 0$. One way to see why this might be the case is to consider the real case when the dimension is 3 or larger, i.e. where $F \in \mathbb{R}^{n\times n}$ for $n \geq 3$. When this is the case, if $\mathbf{x}$ is near $||\mathbf{x}||_2\mathbf{e}_1$ and one trys to reflect $\mathbf{x}$ to $||\mathbf{x}||_2\mathbf{e}_1$ a very small perturbation of $\mathbf{x}$ could cause F to change a lot (think what happens if you rotation $\mathbf{x}$ around the $\mathbf{e}_1$ axis by 90°–the reflecting plane also rotates by 90°). On the other hand in this case reflecting to $-||\mathbf{x}||_2\mathbf{e}_1$ is not very sensitive to small perturbations (the plane will always be near the one perpendicular to $\mathbf{e}_1$.

CHAPTER 15

Nineteenth and Twentieth Lectures

In these two lectures we look at other notions of length of vectors then the 2-norm. We also discuss notions of length for matrices.

1. Vector Norms

We are familiar with the two norm already.

$$||\mathbf{v}||_2 = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v}^* \mathbf{v}}$$

We interpret this as the length of the vector $\mathbf{v}$. Some important properties of the two norm are that

$$\begin{aligned} ||\mathbf{v}||_2 \geq 0 \text{ and } ||\mathbf{v}||_2 = 0 &\iff \mathbf{v} = 0 \\ ||\lambda \mathbf{v}||_2 &= |\lambda| ||\mathbf{v}||_2 \\ ||\mathbf{v} + \mathbf{w}||_2 &\leq ||\mathbf{v}||_2 + ||\mathbf{w}||_2. \end{aligned}$$

It is sometimes necessary to have other notions of length besides the 2-norm. To do this we take the three preceeding properties as a definition. We say a function $|| \cdot || : \mathbb{C}^m \to \mathbb{R}$ is a *norm* if

$$\begin{aligned} ||\mathbf{v}|| \geq 0 \text{ and } ||\mathbf{v}|| = 0 &\iff \mathbf{v} = 0 \\ ||\lambda \mathbf{v}|| &= |\lambda| ||\mathbf{v}|| \\ ||\mathbf{v} + \mathbf{w}|| &\leq ||\mathbf{v}|| + ||\mathbf{w}||. \end{aligned}$$

There are many norms. For instance: The p-norms, let $\mathbf{x} = \sum_{i=1}^m x_i \mathbf{e}_i$

$$\begin{aligned} ||\mathbf{x}||_p := \left(\sum_{i=1}^m |x_i|^p) \right)^{1/p} \quad (1 \leq p < \infty) \\ ||\mathbf{x}||_\infty : - = \max_{1 \leq i \leq m} |x_i| \end{aligned}$$

Note that $||\mathbf{x}||_2 = \sqrt{\mathbf{x}^* \mathbf{x}}$ which agrees with the usual notion of 2 norm. It is a good exercise to check that the ∞-norm is a actually a norm. There are lots of other norms for instance let $W \in \mathbb{C}^{m \times m}$ be a diagonal matrix with positive entries w_{ii} on the diagonal We can define

$$||\mathbf{x}||_W = ||W\mathbf{x}||_2 = \sqrt{\sum_{1=1}^m |w_{ii} x_i|^2}$$

Unit ball is then some sort of ellipse.

2. Induced Matrix Norms

Associated to any pair of norms $||\cdot||_{(n)}$ on $\mathbb{C}^n$ and $||\cdot||_{(m)}$ on $\mathbb{C}^m$ (not necessarily p-norms) there is an *induced matrix norm*, $||\cdot||_{(m,n)}$ on $\mathbb{C}^{m\times n}$ This norm measures the maximum amount of "stretching" (as measured by the norms on $\mathbb{C}^m$ and $\mathbb{C}^n$) that multiplication by A can achieve that is

$$||A||_{(m,n)} = \sup_{\mathbf{x}\in\mathbb{C}^n,\mathbf{x}\neq 0} \frac{||A\mathbf{x}||_{(m)}}{||\mathbf{x}||_{(n)}} = \sup_{\mathbf{x}\in\mathbb{C}^n,||\mathbf{x}||_{(n)}=1} \frac{||A\mathbf{x}||_{(m)}}{||\mathbf{x}||_{(n)}}.$$

We leave it as an excercise to see that the two definitions are equivalent. Another way to think about this is to note that the induced norm is the smallest value C so that

$$||A\mathbf{x}||_{(m)} \leq C||\mathbf{x}||_{(n)}$$

for all $\mathbf{x} \in \mathbb{C}^n$. In general this definition is hard to use computationally (as formulated it is not an algebraic property). It is very intuitive though and has good mathematical properties.

We will often consider the case when $||\cdot||_{(n)} = ||\cdot||_p$ and $||\cdot||_{(m)} = ||\cdot||_p$ (i.e. both norms are p-norms). We then write $||A||_p$ instead of $||A||_{(m,n)}$. A simple example. Suppose that $m = n$ and A is a diagonal matrix

$$A = \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ \vdots & & \ddots & \\ & & & a_m \end{bmatrix}$$

Then $||A||_p = \max 1 \leq i \leq m|a_i|$. When $p = 2$ we can see this geometrically. As A maps a circle to an ellipse. And the longest vector in the image is the biggest axis.

Another example. Lets compute the 1-norm of a matrix. This turns out to be easy to determine in terms of the lengths of columns. We calim that with

$$A = \begin{bmatrix} \mathbf{a}_1 & |\cdots & |\mathbf{a}_n \end{bmatrix}$$

one has

$$||A||_1 = \max_{1\leq j\leq n} ||\mathbf{a}_j||_1$$

To see this we calculate for $\mathbf{x} = \sum_{j=1}^n x_j\mathbf{e}_j$ with $||\mathbf{x}||_1 = 1$. In this case we see that $\sum_{j=1} |x_j| = 1$. Then

$$||A\mathbf{x}||_1 = ||\sum_{j=1}^n x_j\mathbf{a}_j||_1 \leq \sum_{j=1}^n ||x_j\mathbf{a}_j||_1 = \sum_{j=1}^n |x_j|||\mathbf{a}_j||_1$$

But then

$$\leq \left(\max_{1\leq j\leq n} ||\mathbf{a}_j||_1\right) \sum_{j=1}^n |x_j| = \max_{1\leq j\leq n} ||\mathbf{a}_j||_1$$

This implies

$$||A||_1 \leq \max_{1\leq j\leq n} ||\mathbf{a}_j||_1$$

To get the equality we we suppose the maximum is achieved at the j_0 column i.e.

$$\max_{1\leq j\leq n} ||\mathbf{a}_j||_1 = ||\mathbf{a}_{j_0}||_1$$

then with $\mathbf{x}_0 = \mathbf{e}_{j_0}$ one has $||\mathbf{x}_0||_1 = 1$ and $||A\mathbf{x}_0||_1 = ||\mathbf{a}_{j_0}||_1$. In a similar fashion one can show that

$$||A||_\infty = \max_{1\leq i\leq m} ||\mathbf{a}_i^*||_1$$

I.e. is the maximum length of the rows. We leave this as an exercise.

Computing matrix p-norms for $1 < p < \infty$ is much harder. We will see a method to do this for 2 norms (which is the most important). One fact that can be useful in at least getting a bound on induced norms is a generalization of the Cauchy-Schwarz inequality called H olders Inequality:

$$|\langle \mathbf{x}, \mathbf{y}\rangle| \leq ||\mathbf{x}||_p ||\mathbf{y}||_q$$

provided $1/p + 1/q = 1$. When $p = q = 2$ this is the Cauchy-Schwarz inequality.

3. General Matrix Norms

There are many more norms on matrices then just the induced norms. In general we say a map $||\cdot|| : \mathbb{C}^{m\times n} \to \mathbb{R}$ is a matrix norm if it is just a norm on the vector space $\mathbb{C}^{mn}$. That is

$$\begin{aligned} ||A|| \leq 0 \text{ and } ||A|| = 0 &\iff A = 0 \\ ||A + B|| &\leq ||A|| + ||B|| \\ ||\lambda A|| &= |\lambda| ||A|| \end{aligned}$$

It is easy to see any induced norm satisfies these conditions.

One important norm that is not an induced norm is the so called *Frobenius norm*. This is given by

$$||A||_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2} = \left(\sum_{j=1}^{n} ||\mathbf{a}_j||_2^2 \right)^{1/2}$$

Which is just the 2-norm on $\mathbb{C}^{mn}$.

One other way to compute this (which is useful from a theoretical point of view) is

$$||A||_F = \sqrt{tr(A^*A)} = \sqrt{tr(AA^*)}$$

Here $tr(A) = \sum_{i=1}^{\min(m,n)} a_{ii}$. It is a simple exercise to check this.

General matrix norms do not interact with matrix multiplication. However, for induced norms and the Frobenius norm the norm of the product is controlled by the product of the norms. Indeed,

$$||AB||_{(l,n)} \leq ||A||_{(l,m)} ||B||_{(m,n)}$$

here $A \in \mathbb{C}^{l\times m}$ and $B \in \mathbb{C}^{m\times n}$. To see this just consider

$$||A\mathbf{x}||_{(l)} \leq ||A||_{(l,m)} ||B\mathbf{x}||_{(m)} \leq ||A||_{(l,m)} ||B||_{(m,n)} ||\mathbf{x}||_{(n)}$$

and

$$||AB||_F \leq ||A||_F ||B||_F$$

Final useful property the induced 2-norm and the Frobenius norm is that they are invariant under pre- or post-multiplication by a unitary matrix. That is let $Q \in \mathbb{C}^{m\times m}$ and $Q' \in \mathbb{C}^{n\times n}$ both be unitary. Then for $A \in \mathbb{C}^{m\times n}$ one has

$$||QA||_2 = ||A||_2 \text{ and } ||QA||_F = ||A||_F$$

and

$$||AQ'||_2 = ||A||_2 \text{ and } ||AQ'||_F = ||A||_F.$$

CHAPTER 16

Twenty-First Lecture

We introduce the singular value decomposition (SVD).

1. What is the SVD: A Geometric point of view

The SVD is a factorization of an arbitrary matrix that follows from geometric properties of linear maps. In particular, one tries to understand what the image of the unit sphere is under multiplication by A. To work geometrically we first focus on the reals. To that end, let $A \in \mathbb{R}^{m\times n}$ and consider the unit sphere in $\mathbb{R}^n$ i.e. the vectors $\mathbf{x} \in \mathbb{R}^n$ with $||\mathbf{x}||_2 = 1$. We denote this set by S and the consider it is AS. Formally, $AS = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = A\mathbf{x}, \mathbf{x} \in S\}$. We claim that AS is in general a hyperellipse (i.e. a higher dimenisonal analog of an ellipse).

For $n = m = 2$ this means that S is the unit circle and AS should be a rotation and stretching of some ellipse. Note that we are allowed to stretch so much that AS is actually a line segment.

To be more precise we suppose that $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ and suppose also that A has full rank (i.e. the columns linearly independent). We define the *singular values* $\sigma_1, \sigma_2, \ldots, \sigma_n$ to be the length of the principal semiaxes of AS. We usually order these so $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n > 0$. Note that $\sigma_n > 0$ as $N(A) = \{0\}$.

We define the *left singular vectors* of A to be the set of orthogonal unit vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ in $\mathbb{C}^m$ so that $\sigma_i\mathbf{u}_i$ is a principal semiaxis of AS. In particular $\sigma_1\mathbf{u}_1$ is the largest semiaxis of AS. The *right singular vectors* are the set of orthgonal unit vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in $\mathbb{C}^n$ so that $A\mathbf{v}_i = \mathbf{u}_i$. As $N(A) = \{0\}$ the $\mathbf{u}_i$ are unique. We mention that it is not a priori clear that the $\mathbf{u}_i$ need to be orthgonal, this is however true and is something we will show.

In terms of matrices:

$$AV = \hat{U}\Sigma$$

Here

$$V = \begin{bmatrix} |\mathbf{v}_1| & \cdots & |\mathbf{v}_n \end{bmatrix} \in \mathbb{C}^{n\times n}$$

While

$$\hat{U} = \begin{bmatrix} |\mathbf{u}_1| & \cdots & |\mathbf{u}_n \end{bmatrix} \in \mathbb{C}^{m\times n}$$

and

$$\Sigma = \begin{bmatrix} \sigma_1 & \cdots & \\ \vdots & \ddots & \vdots \\ & & \sigma_n \end{bmatrix}$$

This means one has a *reduced SVD* factorization:

$$A = \hat{U}\Sigma V^*.$$

As with the QR factorization we can form a full SVD by adding additionally columns to $\hat{U}$ to make a unitary square matrix U. This requires adding additional zeros to Σ. This gives

$$A = U\Sigma V^*$$

2. What is the SVD: an Algebraic Point of View

While the geometric point of view discussed above is imporatant to understanding the SVD it is hard to make rirgourous (and not easy to compute with). We will now discuss a more algebraic point of view.

We let m, n now be arbitary integers and let $A \in \mathbb{C}^{m\times n}$ also be arbitary. A (full) *Singular Value Decomposition* of A is a factorization

$$A = U\Sigma V^*$$

Where $U \in \mathbb{C}^{m\times m}$ is unitary. $V \in \mathbb{C}^{n\times n}$ is unitary and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal. We assume in addition that the diagonal elements of Σ are non-negative and ordered so $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ where $p = \min(m, n)$. That is we can write (here we have $m = n$):

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \\ 0 & \cdots & 0 & \sigma_p \end{bmatrix}$$

Notice that it is then clear that the image of the unit sphere under A is a hyperellipse.

The issue now is to see whether every matrix admits a singular value decomposition. This turns out to always be the case:

THEOREM 2.1. *Every matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition. Furthermore, the singular values $\{\sigma_j\}$ are uniquely determined and if A is square and the σ_j are distinct the ne the left and right singular vectors $\{u_j\}$ and $\{v_j\}$ are uniquely determined up to (complex) sign.*

PROOF. The method of proof is an induction on the dimension of A where what me really mean is an induction on $l = \min(m, n)$. Set $\sigma_1 = ||A||_2$. Because the unit sphere is a compact and the map $\mathbf{x} \to ||A\mathbf{x}||_2$ is continuous there must be vectors $\mathbf{v}_1' \in \mathbb{C}^n$ and $\mathbf{u}_1' \in \mathbb{C}^m$ with $||\mathbf{v}_1'||_2 = ||\mathbf{u}_1'||_2 = 1$ and so that $A\mathbf{v}_1' = \sigma_1\mathbf{u}_1'$. You should take this for granted as it is beyond the scope of this class to discuss it further. Consider a basis extension of $\mathbf{v}_1'$ to $\{\mathbf{v}_j'\}$ an orthonormal basis of $\mathbb{C}^n$ and a basis extension of $\mathbf{u}_1'$ to $\{\mathbf{u}_j'\}$ an orthonormal basis of $\mathbb{C}^m$. Let U_1 and V_1 denote the matrices with columns $\{\mathbf{v}_j'\}$ and $\{\mathbf{u}_j'\}$. Then one has

$$U_1^* A V_1 = S = \begin{bmatrix} \sigma_1 & \mathbf{w}^* \\ 0 & B \end{bmatrix}$$

Here $B \in \mathbb{C}^{(m-1)\times(n-1)}$ and $\mathbf{w} \in \mathbb{C}^{n-1}$. One estimates

$$\left|\left|\begin{bmatrix} \sigma_1 & \mathbf{w}^* \\ 0 & B \end{bmatrix}\begin{bmatrix} \sigma_1 \\ w \end{bmatrix}\right|\right|_2 \geq \sigma_1^2 + \mathbf{w}^*\mathbf{w} = (\sigma_1^2 + \mathbf{w}^*\mathbf{w})^{1/2}\left|\left|\begin{bmatrix} \sigma_1 \\ w \end{bmatrix}\right|\right|_2, .$$

This means $\sigma_1 = ||A||_2 = ||S||_2 \geq (\sigma_1^2 + \mathbf{w}^*\mathbf{w})^{1/2}$ which can only occur if $\mathbf{w}^*\mathbf{w} = ||\mathbf{w}||_2^2 = 0$. In particular,

$$U_1^* A V_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix}$$

If $n = 1$ or $m = 1$ we are done – this is the base case $l = 1$. Otherwise B describes an action on $span(\mathbf{v}_1)^\perp$. By the induction hypothesis one has an SVD of B

$$B = U_2 \Sigma_2 V_2^*$$

One verifies that

$$A = U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}^* V_1^*$$

is an SVD of A. The point is that the first two matrices are unitary so their product is also unitary, same true for last two and the middle one is diagonal. Notice that $||B||_2 \leq ||A||_2$ so the singular values are ordered as desired.

To verify uniqueness we note that σ_1 is uniquely determined by being equal to $||A||_2$. Now suppose that in addition to $\mathbf{v}_1$ there is another (linearly independent) vector $\mathbf{w}$ with $||\mathbf{w}||_2 = 1$ and $||A\mathbf{w}||_2 = \sigma_1$. Let

$$\mathbf{v}_2 = \frac{P_{\mathbf{v}_1^\perp} \mathbf{w}}{||P_{\mathbf{v}_1^\perp} \mathbf{w}||_2}$$

As $||A||_2 = \sigma_1$ one has $||A\mathbf{v}_2||_2 \leq \sigma_1$. We claim this is an equality.

To see this we note that as $\mathbf{v}_1$ is a left singular vector and $\mathbf{v}_2$ is perpendicular to $\mathbf{v}_1$ one has

$$\langle A\mathbf{v}_1, A\mathbf{v}_2 \rangle = 0$$

To see this we note that A has the SVD $A = U\Sigma V^*$ where $\mathbf{v}_1$ is the first column of V. Thus $\langle A\mathbf{v}_1, A\mathbf{v}_2 \rangle = \langle A^* A\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \sigma_1^2 \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$. Now we can write $\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ and since $\mathbf{v}_1$ and $\mathbf{v}_2$ are an orthonormal set $|c_1|^2 + |c_2|^2 = 1$ with both non-zero. Withouth equality the Pythagorean theorem would imply $||A\mathbf{w}||_2 < \sigma_1$ a contradiction. Thus $\mathbf{v}_2$ is a second right singular vector of A corresponding to σ_1. This implies that the singular values would not be distinct and so cannot occur. The result follows by induction. □

CHAPTER 17

Twenty-Second and Twenty-Third Lectures

We discussed some applications of the SVD.

1. Applications of the SVD

If we know the SVD of a matrix there is lots of useful information we can deduce about the matrix A. Let $A \in \mathbb{C}^{m\times n}$ and suppose that A has the SVD

$$A = U\Sigma V^*$$

Let us write $p = \min(m, n)$ so p is the number of singular values of A then let $r \leq p$ denote the number of non-zero singular values. We denote by $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ the non-zero singular values of A and

$$U = \begin{bmatrix}|\mathbf{u}_1| & \cdots & |\mathbf{u}_m\end{bmatrix} \text{ and } V = \begin{bmatrix}|\mathbf{v}_1| & \cdots & |\mathbf{v}_n\end{bmatrix}$$

the left and right singular vectors.

THEOREM 1.1. *The rank of A is r the number of non-zero singular values.*

PROOF. As Σ is diagonal, it is immediate that $\mathbf{e}_1, \ldots, \mathbf{e}_r$ is a basis of $R(\Sigma)$. Since U, V are intertible one then has that $U\mathbf{e}_i$ is a basis of $R(A)$. □

A more refined result is the following:

THEOREM 1.2. *$R(A) = span(\mathbf{u}_1, \ldots, \mathbf{u}_r)$ and $N(A) = span(\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n)$*

PROOF. For a diagonal matrix $R(\Sigma) = span(\mathbf{e}_1, \ldots, \mathbf{e}_r)$ while $N(\Sigma) = span(\mathbf{e}_{r+1}, \ldots, \mathbf{e}_n)$. The range of A has a basis $U\mathbf{e}_i = \mathbf{u}_i$ for $1 \leq i \leq r$. Similarly, by solving $V^*\mathbf{x} = \mathbf{e}_i$ for $r + 1 \leq i \leq n$ one obtains vectors in $N(A)$. We see that the solutions to this equation is $V\mathbf{e}_i = \mathbf{v}_i$. □

REMARK 1.3. Notice that this actually gives an orthonormal basis ofof $R(A)$. Namely, $\mathbf{u}_1, \ldots, \mathbf{u}_r$. This is *NOT* necessarily the same as the one obtained via QR factorization.

As a consequence, if one sets

$$\hat{U} = \begin{bmatrix}|\mathbf{u}_1| & \cdots & |\mathbf{u}_r\end{bmatrix}$$

then one has the matrix $P = \hat{U}\hat{U}^*$ giving orthogonal projection onto $R(A)$. In particular, we can use the SVD to solve least squares problems.

We can also use the SVD to compute 2-norms. Indeed,

THEOREM 1.4. *$||A||_2 = \sigma_1$ and $||A||_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$.*

PROOF. To see this we note that both these norms are invariant under pre- and post- multiplication by unitary maps, so $||A||_2 = ||\Sigma||_2$ and $||A||_F = ||\Sigma||_F$. Since Σ is diagonal it is easy to compute the norms in this case. □

One very important application is the SVD is that it allows one to get a good approximations of a given matrix in terms of lower rank matrices. This is important in trying to understand what the "dominant" part of the matrix is. It can also be thought of in terms of how much "compression" can be applied to the matrix.

The basic idea is that can express A is the sum of r rank on matrices

$$A = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^*$$

Which follows just by multiplying out the SVD. The point is there are lots of ways to write A as a sum of rank one matrices for instance

$$A = \sum_{j=1}^{n} \mathbf{a}_j \mathbf{e}_j^*$$

or

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \mathbf{e}_i \mathbf{e}_j^*$$

. However, the sum given by the SVD has the property of having the kth partial sum capturing as much "energy" of A as possible, that is of being the best approximation possible in the induced 2-norm or Frobenius norm.

To make this precise

THEOREM 1.5. *For any k with $0 \leq k \leq r$ define*

$$A_k = \sum j = 1^k \sigma_j \mathbf{u}_j \mathbf{v}_j^*$$

so $A_r = A$. Then:

$$||A - A_k||_2 = \inf_{B \in \mathbb{C}^{m \times n} rank(B) \leq k} ||A - B||_2 = \sigma_{k+1}$$

Here if $k = p = \min(m, n)$ we set $\sigma_{k+1} = 0$.

REMARK 1.6. That is we have that A_k is the best (in terms of the induced 2 norm approximation of A by a rank k matrix).

PROOF. Suppose one has a $B \in \mathbb{C}^{m \times n}$ with $rank(B) \leq k$ and $||A - B||_2 < ||A - A_k||_2 = \sigma_{k+1}$. By the rank-nullity theorem we see that there is a $(n-k)$-dimensional space W in $\mathbb{C}^n$ so that for $\mathbf{w} \in W$, $B\mathbf{w} = 0$. (i.e. $W \subset N(B)$). Now for $\mathbf{w} \in W$ $A\mathbf{w} = (A - B)\mathbf{w}$ and so

$$||A\mathbf{w}||_2 = ||(A - B)\mathbf{w}||_2 \leq ||A - B||_2 ||\mathbf{w}||_2 < \sigma_{k+1} ||\mathbf{w}||_2$$

Thus W is an $(n-k)$ dimensional subspace with $||A\mathbf{w}||_2 < \sigma_{k+1}||\mathbf{w}||_2$. However, by considering $W' = span(\mathbf{v}_1, \ldots, \mathbf{v}_{k+1})$ one obtains a $k+1$ dimensional space (all vectors are orthogonal hence linearly independent) with $||A\mathbf{w}'||_2 \geq \sigma_{k+1}||\mathbf{w}'||_2$ for all $\mathbf{w}' \in W'$. Now W' and W must have a non-zero vector in common (otherwise one would get $n+1$ linearly independent vectors in $\mathbb{C}^n$. But this is a contradiction. □

Notice when σ_{k+1} is small this means that $||A\mathbf{x} - A_k\mathbf{x}||_2$ is small (at least relative to $||\mathbf{x}||_2$). I.e. mulitplication by A is well approximated by multiplication by A_k. A similar result also holds for the Frobenius norm i.e.

THEOREM 1.7. *For any k with $0 \leq k \leq r$ one has*

$$||A - A_k||_F = \inf_{B \in \mathbb{C}^{m \times n} rank(B) \leq k} ||A - B||_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$$

Notice that when $\sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$ is small all of the entries of A are close to the entries of A_k. That is the array of numbers making up A are all well approximated by the array of numbers making up A_k. Notice that A takes mn numbers to determine (i.e. each entry) while A_k takes $(m+n+1)k$ to represent (i.e. the left and right singular vector and the singular value). If k is small relative to $p = \min(m,n)$ this is a significant savings.

2. Least squares via SVD: NIC

As we've seen the SVD of a matrix A gives a orthonormal basis of $R(A)$. More than that it gives an approach to solving least squares problems.

Assume that $A \in \mathbb{C}^{m\times n}$ with $m > n$. We assume also that $N(A) = \{0\}$ though this isn't neccesary. We want to solve the overdetermined problem

$$A\mathbf{x} = \mathbf{b}$$

in a least squares sense using the SVD. To that end, let A have reduced SVD

$$A = \hat{U}\hat{\Sigma}V^*$$

with

$$\hat{U} = \begin{bmatrix} |\mathbf{u}_1| & \cdots & |\mathbf{u}_n \end{bmatrix} \in \mathbb{C}^{m\times n}$$

Now orthogonal projection onto $R(A)$ is given by $P = \hat{U}\hat{U}^*$. Hence to solve the equation in the least squares sense it is enough to solve

$$A\mathbf{x} = P\mathbf{b}$$

but this leads to

$$\hat{U}\hat{\Sigma}V^*\mathbf{x} = \hat{U}\hat{U}^*\mathbf{b}$$

since the columns of $\hat{U}$ are linearly independent this is equivalent to solving

$$\hat{\Sigma}V^*\mathbf{x} = \hat{U}^*\mathbf{b}$$

But this consists just of solving a diagonal system and multiplying by a unitary matrix.

CHAPTER 18

Twenty-Fourth Lecture

In this lecture we recall some definitions related to the study of eigenvectors and eigenvalues. This will allows us to compute the SVD of a matrix by solving a related eigenvalue problem (which is slightly more algebraically tractable).

1. Eigenvalues and Eigenvectors

We review some the very imporant linear concept of eigenvectors and eigenvalues. It is helpful to compare and contrast these with singular vectors and singular values.

Recall that for $A \in \mathbb{C}^{m\times m}$ we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* and $0 \neq \mathbf{v} \in \mathbb{C}^m$ is an *eigenvector* if

$$A\mathbf{v} = \lambda\mathbf{v}$$

that is multiplication of $\mathbf{v}$ by A scales $\mathbf{v}$ by λ. We call the set of all eigenvalues of A, $\Lambda(A)$ the *spectrum* of A. We point out that if A is singular then $0 \in \Lambda(A)$ and any non-zero vector in $N(A)$ is then an eigenvector (with eigenvalue 0).

When everything is real – i.e. both the matrix A and the eigenvalue λ and eigenvectors $\mathbf{x}$– then one can geometrically understand this as saying A scales $\mathbf{x}$ by $|\lambda|$ (and possible reverses its direction if $\lambda < 0$. However, it is possible for A to be real and for λ and $\mathbf{v}$ to be complex. This contrasts with much of what we have seen previously and should be kept in mind. For instance the matrix

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

which geometrically rotates by 90° has eigenvalues $\pm\sqrt{-1} = \pm I$. Roughly speaking, for real matrices, complex eigenvalues correspond geometrically to such a "rotation" (possible also with a scaling) while real eigenvalues correspond to pure scaling.

How do we find eigenvalues and eigenvectors? We can recast the question slightly and see that we are trying to find non-trivial solutions to

$$(A - \lambda I)\mathbf{x} = 0$$

That is we try to find λ so that $N(A - \lambda I) \neq \{0\}$ and then in this case try and find elements in the null space. The latter problem is easy (as it is just solving a linear system) and the difficulty arises mostly in the former. Indeed, determining the spectrum $\Lambda(A)$ is an essentially non-linear problem.

Phrase things in a manner that is amenable to algebraic investigation we must recall the determinant. This is a function:

$$\det : \mathbb{C}^{m\times m} \to \mathbb{C}$$

defined by

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

and inductively by

$$\det A = \sum_{i=1}^{m} (-1)^i a_{1i} \det A_{1i}.$$

Here A_{1i} is the matrix in $\mathbb{C}^{(m-1)\times(m-1)}$ obtained by omitting the first row and ith column. In other words, we have defined the determinant by expanded along the first row. The determinant has many properties that allow one to compute it in other ways. We refer to Strang for instance for more detailed discussion. You should be able to compute the determinant of small matrices (i.e. 2×2 and 3×3).

There is a big theory of determinants. The main property we will need is the fact that A is non-singular when and only when $\det(A) \neq 0$. Using this fact and an expansion of the determinant we see that $\lambda \in \Lambda(A)$ when and only when λ is a root of the polynomial

$$p_A(z) = \det(zI - A) = z^m + c_{m-1}z^{m-1} + \ldots + c_0$$

is a degree m polynomial. We call this the *characteristic* polynomial of A. The coefficients c_i are determined by the entries of A in an explicit (but non-linear) way.

This is one place that working over $\mathbb{C}$ greatly simplifies things. Indeed, the fundamental theorem of algebra tells us that over $\mathbb{C}$. $p_A(z)$ has exactly m roots (counting multiplicity). That is we can factor

$$p_A(z) = (z - \lambda_1)^{m_1} \ldots (z - \lambda_k)^{m_k}$$

where $k \leq m$, $m_i \geq 1$ and $\sum_i m_i = m$. We call the value $m_i = m_{\lambda_i}$ the *algebraic multiplicity* of the eigenvalue λ_i. Notice one cannot always produce such a factorization over $\mathbb{R}$.

For a $\lambda \in \Lambda(A)$ we say the *eigenspace* associated λ is the vector space

$$E_\lambda = N(\lambda I - A)$$

this is always a non-empty vector space all the non-zero vectors of E_λ are eigenvectors with eigenvalue λ. We let $g_\lambda = dim E_\lambda$ and call this number the *geometric multiplicity* of λ. One always has $1 \leq g_\lambda \leq m_\lambda$ (for a proof we refer to Trefethen-Bau Lecture 24). We say A is *non-defective* if $g_\lambda = m_\lambda$ for all $\lambda \in \Lambda(A)$.

The point above is that for a non-defective $A \in \mathbb{C}^{m\times m}$ one has the dimensions of the eigenspaces summing up to m (since the algebraic multiplicites have this property). In this case there is a set $\mathbf{x}_1, \ldots, \mathbf{x}_m$ a basis of $\mathbb{C}^m$ where each $\mathbf{x}_i$ is an eigenvector of A. In particular, there is a non-singular matrix

$$X = \begin{bmatrix} |\mathbf{x}_1 & \cdots & |\mathbf{x}_m \end{bmatrix}$$

so that

$$A = X\Lambda X^{-1}$$

where Λ is diagonal. Notice that unlike the SVD we have only one set of vectors, they are not neccesarily orthogonal, we must start with a square matrix, the diagonal matrix may have complex or negative entries and we aren't guarenteed of such a decomposition existing.

One important result we will need is the following:

THEOREM 1.1. *Let $A \in \mathbb{C}^{m\times m}$ be hermitian, i.e. $A^* = A$. Then A is non-defective, all the eigenvalues of A are real and one may choose a orthonormal basis of eigenvectors.*

COROLLARY 1.2. *There is a $Q \in \mathbb{C}^{m\times m}$ that is unitary so that*

$$A = Q\Lambda Q^*$$

where Λ is diagonal with real entries.

We will prove this later. The point is that hermitian matrices are rather nice from an eigenvalue point of view.

2. Eigenvalues and the SVD

Despite the differences noted above, there is a clear important relationship between eigenvalues and singular values. Indeed, for hermitian matrices they are (practically) the same. One thing that is useful about this is that eigenvalues and eigenvectors can be found algebraically (though this is not an easy problem for large matrices). This allows one to find singular values in an algebraic manner:

THEOREM 2.1. *Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ be the non-zero singular values of $A \in \mathbb{C}^{m\times n}$. Then σ_i^2 are precisely the non-zero eigenvalues of A^*A and of AA^* (i.e these matrices have the same non-zero eigenvalues).*

PROOF. : Let A have SVD $A = U\Sigma V^*$ then

$$A^*A = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma^* U^* U\Sigma V^* = V\Sigma^*\Sigma V^* = V\Sigma^2 V^*$$

This says exactly that A^*A has eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$ with associated eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$. Of course there may be more eigenvalues but these must all be zero. Similarly,

$$AA^* = (U\Sigma V^*)(U\Sigma V^*)^* = U\Sigma V^* V\Sigma^* U^* = U\Sigma\Sigma^* U^* = U\Sigma^2 U^*$$

This says that AA^* has eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$ with associated eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$. As above there may be more eigenvalues but they are zero. □

It is imporatant to note that this given the eigenvalues of A^*A one gets the singular values by taking the (positive) square root. More over, by taking the associated eigenvectors of A^*A one gets the right singular vectors of A.

CHAPTER 19

Twenty-Sixth Lectures

In the this lecture we further discuss properties of the Eigenvalues and Eigenvectors. In particular, we derive some consequences of the Schur factorization discussed last lecture. (Note Lecture Twenty-Five was accidentally overwritten).

1. Applications of the Schur Factorization

Recall, last time we showed that:

THEOREM 1.1. *Every square matrix $A \in \mathbb{C}^{m\times m}$ has a Schur factorization. That is*

$$A = QTQ^*$$

where $Q \in \mathbb{C}^{m\times m}$ is unitary and $T \in \mathbb{C}^{m\times m}$ is upper triangular.

REMARK 1.2. If T is diagonal then A is diagonizable and is indeed is *unitarly* diagonizable.

One nice thing about upper triangular matrices is that the entries on their diagonal are the eigenvalues:

THEOREM 1.3. *Let*

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots \\ 0 & x_{22} & \cdots \\ \vdots & & \ddots \end{bmatrix}$$

be upper triangular. Then $\Lambda(X) = \{x_{11}, \ldots, x_{mm}\}$.

PROOF. Expanding out the determinant one can compute the characteristic polynomial of X to be

$$P_X(z) = (z - x_{11}) \cdots (z - x_{mm}).$$

One readily sees that the roots are then the elements on the diagonal of X. □

As a consequence, if we can find a Schur factorization of a matrix A we can find the eigenvalues of a matrix. In order to make this precise idea of a *similarity transformation*. This is just another word for changing the basis that one uses to represent the matrix.

DEFINITION 1.4. We say two matrices $A, B \in \mathbb{C}^{m\times m}$ are *similar* if there is a non-singular matrix X so that the matrix $B = X^{-1}AX$.

As we've seen B is the matrix A in the basis given by the columns of X. An important fact which follows from properties of the determinant is that if A and B are similar matrices then $P_A(z) = P_B(z)$, that is A and B have the same characteristic polynomial. In particular, A and B have the same eigenvalues with

the same algebraic multiplicities. In fact, as A and B are similar there is a non-singular $X \in \mathbb{C}^{m\times m}$ so that $B = X^{-1}AX$. Clearly, if $\mathbf{v}$ is an eigenvector of A corresponding to $\lambda \in \Lambda(A) = \Lambda(B)$, then $X^{-1}\mathbf{v}$ is an eigenvector of B corresponding to λ. In particular, the geometric multiplicity of λ with respect to A and B are the same. As mentioned, the proof uses from properties of the determinant. We refer to Theorem 24.3 of Trefethen and Bau.

A consequence of the Schur factorization is that any matrix $A \in \mathbb{C}^{m\times m}$ is similar to an upper triangular matrix $T \in \mathbb{C}^{m\times m}$ and hence the eigenvalues of A can be determined from the diagonal of T.

We can also use the Schur factorization to prove things:

THEOREM 1.5. *Let $A \in \mathbb{C}^{m\times m}$ be hermitian. Then A has real eigenvalues, is non-defective and there is an orthonormal set of eigenvectors of A.*

PROOF. Let $A = QTQ^*$ be a Schur factorization of A. One has $A^* = A$ so $QT^*Q^* = QTQ^*$ that is $T^* = T$. Since T is upper triangular this means that T is diagonal and all the entries on the diagonal are real. This implies that the eigenvalues of A are all real as desired. Finally, as T is diagonal, the columns of Q are the eigenvectors of A. □

REMARK 1.6. Another way to say this is that when A is hermitan it is *unitarily diagonizable*.

2. Applications of Eigenvalues

Once one knows the eigenvalues and eigenvectors of a matrix A one can tell a number of useful facts about A right away. However, one can also tell many useful things about iterates of A that is matrices of the form A^n (i..e the matrix obtained by multiplying A by itself n-times). In particular, it is relatively painless to compute A^n given such information. Indeed, suppose A is non-defective and so is diagonizable, i.e. we can write

$$A = X\Lambda X^{-1}$$

for some non-singular X and diagonal Λ here

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \\ 0 & \cdots & 0 & \lambda_m \end{bmatrix}$$

Then it is simple to see that

$$A^2 = X\Lambda X^{-1}X\Lambda X^{-1} = X\Lambda^2 X^{-1}$$

and so by induction

$$A^n = X\Lambda X^{-1}X\Lambda X^{-1} = X\Lambda^n X^{-1}$$

Notice that knowing the SVD does not allow for such a nice formula. In practice, the SVD gives alot of information about the matrix A, but tells one little about iterates of A.

Another thing we can do is take square-roots of (some) matrices. Consider first a diagonal matix

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \\ \vdots & & & \\ 0 & \cdots & 0 & a_m \end{bmatrix}$$

We want to find a B so that $B^2 = A$. For simplicity, we assume $a_i \geq 0$. Then we can take

$$B = \begin{bmatrix} \sqrt{a_1} & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \\ \vdots & & & \\ 0 & \cdots & 0 & \sqrt{a_m} \end{bmatrix}$$

and $B^2 = A$ and so we write $B = \sqrt{A}$ in this case. More generally, we say a hermitian matrix A is *positive semi-definite* if all the eigenvalues of A are non-negative. This is equivalent to $\langle A\mathbf{x}, \mathbf{x}\rangle \geq 0$ for all $\mathbf{x} \in \mathbb{C}^m$ (A still hermitian). Then we can check that there is a hermitian matrix B with $B^2 = A$. Indeed, as A is hermitan it is (unitarily) diagonizable, so

$$A = Q\Lambda Q^*$$

As all the eigenvalues of A are non-negative all the entries of Λ are non-negative, so we just set

$$B = Q\sqrt{\Lambda}Q^*$$

CHAPTER 20

Twenty-Seventh Lecture

We discuss here iterative methods of determining eigenvalues and eigenvectors.

As previously mentioned finding eigenvalues is not an easy procedure. This is because finding the roots of the characteristic polynomial is a non-linear problem, and is computationally difficult. Thus, we will instead discuss some other approaches. The methods we discuss are not the ones used in practice, but are related and will give some insight into how one would numerically find eigenvalues.

It turns out to be the case that the discussion is vastly simplified if we restrict attention to real symmetric matrices. I.e. $A \in \mathbb{R}^{m\times m}$ and $A^* = A^\top = A$. Notice, the eigenvalues (and hence also eigenvectors) are real. To fix notation for this lecture we let $\lambda_1, \dots, \lambda_m$ be the eigenvectors of A and $\mathbf{q}_1, \dots, \mathbf{q}_m$ the associated eigenvalues normalize so $||\mathbf{q}_j||_2 = 1$ (so they form an orthonormal basis of $\mathbb{R}^m$. We also order the eigenvalus so that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_m|$.

1. Rayleigh Quotient

One way to think about an eigenvalues is as follows: Fix a vector $\mathbf{x} \in \mathbb{R}^m$. We seek the scalar $\alpha \in \mathbb{R}$ that makes $\mathbf{x}$ an as close as possible to beign an eigenvector. I.e. we want to minimize

$$||A\mathbf{x} - \alpha\mathbf{x}||_2$$

Of course if $\mathbf{x}$ is actually an eigenvector this is minimized when α is actually the associated eigenvalue as the value is zero.

We can re-formulate this question as follows: We are trying to solve the overdetermined system of equations :

$$\mathbf{x}\alpha = A\mathbf{x}$$

in the one unknown α in the sense of least squares. To make this easier to parse, let us think of $\mathbf{x}$ as an $m \times 1$ matrix and write it as X. Then we are solving

$$X\left[\alpha\right] = A\mathbf{x}$$

in the sense of least squares.

To do this, we need to find P_X the projector onto $R(X) = span(\mathbf{x})$. This is given by

$$P = \frac{\mathbf{x}\mathbf{x}^*}{||\mathbf{x}||_2^2} = \frac{\mathbf{x}\mathbf{x}^\top}{||\mathbf{x}||_2^2}$$

Hence we may take

$$\alpha\mathbf{x} = PA\mathbf{x} = \frac{\mathbf{x}^\top A\mathbf{x}}{||\mathbf{x}||_2}\mathbf{x} = \frac{\langle A\mathbf{x}, \mathbf{x}\rangle}{||\mathbf{x}||_2^2}\mathbf{x}$$

This value α is called the *Rayleigh Quotient* and denote it by $r(\mathbf{x})$. So

$$r(\mathbf{x}) = \frac{\langle A\mathbf{x}, \mathbf{x}\rangle}{||\mathbf{x}||_2^2}$$

Notice that when $\mathbf{x} = \mathbf{q}_j$ is an eigenvector, $r(\mathbf{x}) = \lambda_j$ the associated eigenvalue.

We may think of

$$r : \mathbb{R}^m \to \mathbb{R}$$

as a function of several variables. It is not hard to see that away from $\mathbf{x} = 0$ this is a smooth function (i.e. all partial derivatives exist and are continuous). A straight forward computation gives

$$\nabla r(\mathbf{x}) = \frac{2}{||\mathbf{x}||_2^2} (A\mathbf{x} - r(\mathbf{x})\mathbf{x})$$

(here we have taken the gradient of r). In particular, the *critical values* of r are precisely the eigenvalues of A. While the *critical points* are the eigenvectors. To make this clearer one usually restricts r to the sphere $||\mathbf{x}||_2 = 1$.

How does this help us? Well we always know that the maximum of r and the minimum of r on the sphere $||\mathbf{x}||_2 = 1$ are critical points of r. In particular, these give us eigenvalues. This should remind you of the SVD. This approach (i.e. looking for the maximum and minimum) won't help us directly as it is not very computational. However, it does give us some useful information. Namely, notice that when $\mathbf{x}$ is an eigenvector, $r(\mathbf{x})$ gives the associated eigenvalue. Moreover, if we have reason to believe that $\mathbf{x}$ is near to an eigenvector $\mathbf{q}_j$ then $r(\mathbf{x})$ is near λ_j. Indeed, we Taylor's theorem gives that the following estimate holds:

$$r(\mathbf{x}) - r(\mathbf{q}_j) = O(||\mathbf{x} - \mathbf{q}_j||_2^2), \mathbf{x} \to \mathbf{q}_j \tag{1.1}$$

Here we used that ∇r vanishes at $\mathbf{x} = \mathbf{q}_j$. Notice that for $\epsilon > 0$ small that ϵ^2 is much smaller. In other words, if $\mathbf{x}$ is close to $\mathbf{q}_j$ then $r(\mathbf{x})$ is very close to λ_j.

2. Power Iteration

We introduce now a method called *power iteration* that finds the largest eigenvalue and eigenvector of matrices A under certain conditions on A. The basic idea is that repeated multplication by A tends to amplify the eigenvector corresponding to the largest eigenvalue more than the other eigenvectors. That is if we start with an appropriate $\mathbf{v}$ then consider $\mathbf{v}^{(k)} = A^k\mathbf{v}$, if one expresses $\mathbf{v}$ in the basis of eigenvectors $\{\mathbf{q}_1, \ldots, \mathbf{q}_m\}$ the coefficient in front of $\mathbf{q}_1$ should be much larger than all the other coefficients.

More precisely, start with a (randomly choosen) vector $\mathbf{v}^{(0)}$ with $||\mathbf{v}^{(0)}||_2 = 1$. And consider the iterative construction:

$$\mathbf{v}^{(k+1)} = \frac{A\mathbf{v}^{(k)}}{||A\mathbf{v}^{(k)}||_2}, \lambda^{(k+1)} = r(\mathbf{v}^{(k+1)})$$

Then in good circumstances one has that $\mathbf{v}^{(k)} \to \mathbf{q}_1$ and $\lambda^{(k)} \to \lambda_1$.

THEOREM 2.1. *Suppose that* $|\lambda_1| > |\lambda_2| \geq \ldots |\lambda_m| \geq 0$ *and* $\langle \mathbf{q}_1, \mathbf{v}^{(0)} \rangle \neq 0$ *then the iterates above satisfy*

$$||\mathbf{v}^{(k)} - (\pm\mathbf{q}_1)||_2 = O(\left|\frac{\lambda_2}{\lambda_1}\right|^k), |\lambda^{(k)} - \lambda_1| = O(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k})$$

REMARK 2.2. The signs in front of the $\mathbf{q}_1$ are an unfortunate technical annoyance. If $\lambda_1 > 0$ they may always be taken to be positive, while if $\lambda_1 < 0$ they alternate in k.

PROOF. We note that as the $\mathbf{q}_i$ form an orthonormal basis we can write $\mathbf{v}^{(0)}$ as

$$\mathbf{v}^{(0)} = a_1\mathbf{q}_1 + a_2\mathbf{q}_2 + \cdots + a_m\mathbf{q}_m.$$

Notice that $a_1 = \langle \mathbf{v}^{(0)}, \mathbf{q}_1 \rangle \neq 0$. Then (here c_k is a normalizing term):

$$\begin{aligned}\mathbf{v}^{(k)} &= c_k A^k \mathbf{v}^(0) \\ &= c_k(a_1\lambda_1^k\mathbf{q}_1 + \ldots + a_m\lambda_m^k\mathbf{q}_m) \\ &= c_k\lambda_1^k\left(a_1\mathbf{q}_1 + a_2(\lambda_2/\lambda_1)^k\mathbf{q}_2 + \ldots + a_m(\lambda_m/\lambda_1)^k\mathbf{q}_m\right)\end{aligned}$$

Tthe first estimate follows from this by noting that for $j > 1$, $\left(\frac{\lambda_j}{\lambda_1}\right)^k \to 0$ as $k \to \infty$ at the desired rate. The second follows from this and the quadratic estimate (1.1). When $\lambda_1 > 0$ the signs are all positive if $\lambda_1 < 0$ they alternate. □

Notice that as long as the largest (in magnituded) two eigenvalues have distinct magnitudes (not something one know a priori–a serious drawback) then the power iterates converge to the largest eigenvalue at a rate determined by the ratio between the two eigenvalues. This illustrates some of the drawbacks of this method.

Again the main idea of the method is that successive multiplications by A tends to amplify the part of $\mathbf{v}^{(0)}$ that corresponds the eigenvector $\mathbf{q}_1$ (i.e. the eigenvector associated to λ_1) much more than an other part of $\mathbf{v}^{(0)}$. In particular, after many iterations "most" of $A^k\mathbf{v}^{(0)}$ is in the direction of $\mathbf{q}_1$.

3. Inverse Iteration:NIC

As we saw above there are two major drawbacks to power iteration. First it only finds the largest eigenvalue. Second if there is not a large amount of separation between the largest two eigenvalues the convergence is slow. A way to overcome the first issue is to consider the matrix

$$A - \mu I$$

for some $\mu \in \mathbb{R}$ to be specified. The important point is that $\Lambda(A - \mu I) = \Lambda(A) - \mu$. I.e. the eigenvalues are $\lambda_i - \mu$. If μ is not an eigenvalue, then it is straightforward to see that $A - \mu I$ is invertible and the eigenvalues of

$$(A - \mu I)^{-1}$$

are $(\lambda_i - \mu)^{-1}$. That is if μ is near to the eigenvalue λ_{i_0} then $(A - \mu I)^{-1}$ has a very large eigenvalue given by $(\lambda_{i_0} - \mu)^{-1}$. If we then use this with power iteration it converges to λ_{i_0} and the associated eigenvector $\mathbf{q}_{i_0}$. That is we can find all the eigenvalues, at least as long as we start near enougth.

This procedure is known as inverse iteration. The basic idea is the start with a vector $\mathbf{v}^{(0)}$ with $||\mathbf{v}^{(0)}||_2 = 1$. Now iteratively do the following procedure:

(1) Solve for $(A - \mu I)\mathbf{w} = \mathbf{v}^{(k)}$.
(2) Set $\mathbf{v}^{(k+1)} = \frac{\mathbf{w}}{||\mathbf{w}||_2}$
(3) Set $\lambda^{(k+1)} = r(\mathbf{v}^{(k+1)})$.

In ideal situations one then gets convergence to an eigenvalue.

THEOREM 3.1. *Suppose that λ_{i_0} is the closes eigenvalue of A to μ and λ_{i_1} is the second closest and that $|\mu - \lambda_{i_0}| < |\mu - \lambda_{i_1}| \leq |\mu - \lambda_j|$ for $j \neq i_0$. Further suppose that $\langle \mathbf{v}^{(0)}, \mathbf{q}_{i_0} \rangle \neq 0$. Then the iterates of inverse iteration satisfy*

$$||\mathbf{v}^{(k)} - (\pm \mathbf{q}_{i_0})||_2 = O\left(\left|\frac{\mu - \lambda_{i_0}}{\mu - \lambda_{i_1}}\right|^k\right), |\lambda^{(k)} - \lambda_{i_0}| = O\left(\left|\frac{\mu - \lambda_{i_0}}{\mu - \lambda_{i_1}}\right|^{2k}\right)$$

as $k \to \infty$.

Notice this is essentially the same rate of growth as before, but it does allow one to find different eigenvalues. There is still an issue when one has nearby eigenvalues.

One way to think of the inverse iteration is that it is a way to transform an eigenvalue estimate to an eigenvector estimate. I.e. if we you have a pretty good idea of what one of the eigenvalues is, application of inverse iteration gives you a much better idea, as well as giving an associated eigenvector. This contrasts with (1.1) where having a vector $\mathbf{x}$ that is near to an eigenvector means that $r(\mathbf{x})$ is quite close to an eigenvalue. By a clever combination of inverse iteration with the Rayleigh quotient one obtains a algorithm that converges quite rapidly to an eigenvalue, provided one starts near enough.

CHAPTER 21

Twenty-Eighth Lecture

In the last two lectures we will talk about some more advanced topics relating to methods in numerical linear algebra. I will focus on algorithm called *Conjugate Gradients*. This is covered in Lec. 32 and 38 of T-B.

1. Iterative Methods

We discussed last time a (primitive) method of finding the largest eigenvalue of certain types of matrices via iteration. This week we will see an iterative method for solving systems of equations. You might ask why bother as we have a perfectly good algorithm in Gaussian elimination. The most fundamental reason is speed. For an $m \times m$ matrix Gaussian elimination takes $O(m^3)$ steps. This is too slow in practice once m gets large.

In practice it is often the case that such large matrices are very "sparse" that is they have a lot of zero entries. In this case the structure of the matrix suggests that one should not need to do so much work. Gaussian elimination cannot take advantage of the sparseness while iterative methods such as conjugate gradients can.

The key is that an iterative method just needs a "black-box" that given $\mathbf{x}$ as an input, outputs $A\mathbf{x}$. The point is it may be possible to program such a black-box to be faster than expected (if the matrix is sparse for instance). In contrast Gaussian elimination requires one to work directly with A where no such speed up is possible.

2. Krylov Spaces and solutions to linear systems

We introduce some notation in order to discuss these iterative algorithm. Fix $A \in \mathbb{C}^{m\times m}$ a square matrix and $\mathbf{b} \in \mathbb{C}^m$ a vector. We will consider the *Krylov sequence*, which is

$$\mathbf{b}, A\mathbf{b}, A^2\mathbf{b}, \ldots, A^k\mathbf{b}, \ldots$$

and the *Krylov subspaces*

$$\mathcal{K}_n = span(\mathbf{b}, A\mathbf{b}, A^2\mathbf{b}, \ldots, A^{n-1}\mathbf{b}).$$

That is the span of the first n vectors in the Krylov sequence. We note the easy to check fact that:

$$\mathcal{K}_n \subset \mathcal{K}_{n+1}$$

We also introduce the *Krylov Matrices*

$$K_n = \begin{bmatrix} \mathbf{b} | & A\mathbf{b} | & \cdots & | A^{n-1}\mathbf{b} \end{bmatrix} \in \mathbb{C}^{m\times n}$$

So that the column space of K_n is $\mathcal{K}_n$. We emphasize that our "Black-box" can be used to find the Krylov sequence (and hence K_n and $\mathcal{K}_n$ as $A^2\mathbf{b} = A(A(\mathbf{b})), \ldots,$.

We now discuss a general framework using the Krylov spaces for trying to solve systems of equations iteratively. The exposition is significantly simplified if we assume the system always has a unique solution. That is are trying to then solve:

$$A\mathbf{x} = \mathbf{b}$$

with A non-singular. In otherwords, we are seeking a way to compute:

$$\mathbf{x}_* = A^{-1}\mathbf{b}.$$

We emphasize that theoretically we know $\mathbf{x}_*$ exists, the issue is to numerically compute it in an efficient manner.

In order to motivate our approach we note the following fact:

LEMMA 2.1. *For A and b as above and*

$$\mathcal{K}_n = span(\mathbf{b}, A\mathbf{b}, \ldots, A^{n-1}\mathbf{b}) \subset \mathbb{C}^m$$

the associated Krylov subspaces, if for some k, $dim\mathcal{K}_k < k$ then $\mathbf{x}_* \in \mathcal{K}_k$

PROOF. We note if $\mathbf{b} = 0$ then $\mathbf{x}_* = 0$ is in every Krylov subspace so we may assume $\mathbf{b} \neq 0$ and hence (as A is non-singular) $A^i\mathbf{b} \neq 0$ for all i. If $dim\mathcal{K}_k < k$ then there is a linear dependence amongst the $A^i\mathbf{b}$ that is

$$\sum_{i=1}^{l} c_i A^{j_i}\mathbf{b} = 0$$

where here $l \leq k$ and we take $c_i \neq 0$ and order the terms so that $0 \leq j_1 < j_2 < \cdots < j_l < k$. But then we have

$$\mathbf{x}_* = A^{-1}\mathbf{b} = \frac{1}{c_1}\sum_{i=2}^{l} c_i A^{j_i - j_1 - 1}\mathbf{b} \in \mathcal{K}_k$$

as $k > j_i - j_1 - 1 \geq 0$. □

As a consequence if we look in a big enough Krylov subspace we will find $\mathbf{x}_*$

COROLLARY 2.2. *For A and b as above and*

$$\mathcal{K}_n = span(\mathbf{b}, A\mathbf{b}, \ldots, A^{n-1}\mathbf{b}) \subset \mathbb{C}^m$$

the associated Krylov subspaces, $\mathbf{x}_* \in \mathcal{K}_m$.

PROOF. If $dim\mathcal{K}_m < m$ then the result follows by the preceding Lemma. If $dim\mathcal{K}_m \geq m$ then $\mathcal{K}_m = \mathbb{C}^m$ and so $\mathbf{x}_*$ must be in this space. □

Our method for computing $\mathbf{x}_*$ method will be as follows: we try and find a sequence $\mathbf{x}_n \in \mathcal{K}_n$ so that the error term $\mathbf{E}_n = \mathbf{x}_* - \mathbf{x}_n$ are as small as possible. By the above we see that $\mathbf{E}_m = 0$, but the hope is that for $n << m$ one has $\mathbf{E}_i$ very small and so in only a few steps we've computed a vector very close to the true solution. Surprisingly, in practice this is often the case.

Notice we didn't specify with what norm we were measuring the size of the error term, nor have we indicated how to find the $\mathbf{x}_n$. This differs in different algorithms and is somewhat of a subtle point. One naive approach would be to try and minimize the 2-norm of $\mathbf{E}_n$ and then find $\mathbf{x}_n$ by orthogonal projection (of $\mathbf{x}_*$) onto $\mathcal{K}_n$. However, since we don't know the value of $\mathbf{x}_*$ (it is what we are trying

to compute) this method seems doomed to fail. Indeed, in practice it is better to look at the following "residual term"

$$\mathbf{r}_n = A\mathbf{E}_n = \mathbf{b} - A\mathbf{x}_n$$

as this does not depend on knowing $\mathbf{x}_*$.

CHAPTER 22

Twenth-Ninth Lecture

We now discuss a very neat way of minimizing the error term, the *Method of Conjugate Gradients*.

1. Conjugate Gradients

The method of Conjugate Gradients requires that we try and solve systems

$$A\mathbf{x} = \mathbf{b}$$

that are of a somewhat special form. Namely, we assume that $A \in \mathbb{R}^{m\times m}$ and $A = A^\top$ is real and symmetric. We must also assume that A is *positive definite*. That is, when $\mathbf{v} \neq 0$ then

$$\langle \mathbf{v}, A\mathbf{v}\rangle > 0$$

This corresponds to A have strictly positive eigenvalues. That is, $\lambda \in \Lambda(A)$ implies $\lambda > 0$. Notice a positive matrix is automatically non-singular.

The positivity (and symmetry) of A allows us to define a norm:

$$||\mathbf{x}||_A = \sqrt{\langle \mathbf{x}, A\mathbf{x}\rangle}$$

Which is a norm on $\mathbb{R}^m$. Notice that since A has positive eigenvalues, there is a symmetric B so that $B^2 = A$. In particular, we have that

$$\langle \mathbf{x}, A\mathbf{x}\rangle = \langle \mathbf{x}, B^2\mathbf{x}\rangle = \langle \mathbf{x}, B^\top B\mathbf{x}\rangle = \langle B\mathbf{x}, B\mathbf{x}\rangle = ||B\mathbf{x}||_2^2$$

.

The method of conjugate gradients does the following: It constructs iteratively the vector $\mathbf{x}_n \in \mathcal{K}_n$ with the property that if $\mathbf{E}_n = \mathbf{x}_* - \mathbf{x}_n$ then $||\mathbf{E}_n||_A$ is minimum. In other words, it finds the vector $\mathbf{x}_n$ in the nth Krylov subspace that is closest to the actual answer (in the norm $||\cdot||_A$). As discussed above, we use the $||\cdot||_A$ norm so that the minimizer can be found with out knowing the value of $\mathbf{x}_*$.

Let us first give the algorithm and then discuss its properties: We start by setting $\mathbf{x}_0 = 0, \mathbf{r}_0 = \mathbf{b}, \mathbf{v}_0 = \mathbf{r}_0$. Now do the following for $n = 1, 2, 3, \ldots$:

$$\begin{aligned}
\alpha_n &:= \frac{||\mathbf{r}_{n-1}||_2^2}{||\mathbf{v}_{n-1}||_A^2} \\
\mathbf{x}_n &:= \mathbf{x}_{n-1} + \alpha_n \mathbf{v}_{n-1} \\
\mathbf{r}_n &:= \mathbf{r}_{n-1} - \alpha_n A\mathbf{v}_{n-1} \\
\beta_n &:= \frac{||\mathbf{r}_n||_2^2}{||\mathbf{r}_{n-1}||_2^2} \\
\mathbf{v}_n &:= \mathbf{r}_n + \beta_n \mathbf{v}_{n-1}
\end{aligned}$$

The algorithm is terminated if $\mathbf{r}_n = 0$ as at that step we will have $\mathbf{x}_n = \mathbf{x}_*$.

Here $\mathbf{x}_n$ is the approximate solution and $\mathbf{r}_n$ is the residual term. We refer to the numbers α_n as the "step length" , the vectors $\mathbf{v}_n$ as the "search direction" and β_n as the "improvement". Notice that one avoids dividing by zero as long as $\mathbf{r}_{n-1} \neq 0$. But $\mathbf{r}_{n-1} = 0$ means we have found the solution at the $n-1$ step so the algorithm would have to terminate.

We claim (and will prove) that for this algorithm one always has that $\mathbf{x}_n \in \mathcal{K}_n$ and that $||\mathbf{E}_n||_A < ||\mathbf{x}_* - \mathbf{y}||_A$ for any $\mathbf{y} \in \mathcal{K}_n$ with $\mathbf{y} \neq \mathbf{x}_n$. In other words, this simple procedure allows us to find the minimizerer (in $||\cdot||_A$ norm) of the errors inside each Krylov subspace.

Let us now begin to justify why the Conjugate gradient method works we will do so by proving some useful facts:

LEMMA 1.1. *As long as $\mathbf{r}_{n-1} \neq 0$ the vectors in the Conjugate Gradient algorithm satisfy*

$$\begin{aligned}\mathcal{K}_n &:= span(\mathbf{b}, A\mathbf{b}, \ldots, A^{n-1}\mathbf{b}) \\ &= span(\mathbf{x}_1, \ldots, \mathbf{x}_n) \\ &= span(\mathbf{r}_0, \ldots, \mathbf{r}_{n-1}) \\ &= span(\mathbf{v}_0, \ldots, \mathbf{v}_{n-1})\end{aligned}$$

In addition the residuals are orthogonal

$$\langle \mathbf{r}_n, \mathbf{r}_j \rangle = 0 \quad (j < n)$$

and the search directions are "A-orthogonal"

$$\langle \mathbf{v}_n, A\mathbf{v}_j \rangle = 0 \quad (j < n)$$

PROOF. The proof is by induction on n. For $n = 1$ the results are all self-evident so we treat only the case that $n - 1 \to n$. Let us first show the spanning properties. We first note that as $\mathbf{x}_n := \mathbf{x}_{n-1} + \alpha_n \mathbf{v}_{n-1}$ we have by the induction hyptothesis that

$$\mathbf{x}_n \in span(\mathbf{v}_0, \ldots, \mathbf{v}_{n-1})$$

and

$$\mathbf{v}_{n-1} \in span(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

so

$$span(\mathbf{x}_1, \ldots, \mathbf{x}_n) = span(\mathbf{v}_0, \ldots, \mathbf{v}_{n-1})$$

In a similar manner, from $\mathbf{v}_{n-1} := \mathbf{r}_{n-1} + \beta_{n-1}\mathbf{v}_{n-2}$ and the induction hypothesis see that

$$span(\mathbf{v}_0, \ldots, \mathbf{v}_{n-1}) = span(\mathbf{r}_1, \ldots, \mathbf{r}_{n-1})$$

Finally, from $\mathbf{r}_{n-1} = \mathbf{r}_{n-2} - \alpha_{n-1} A\mathbf{v}_{n-2}$ and the induction hypothesis we see that

$$span(\mathbf{r}_1, \ldots, \mathbf{r}_{n-1}) = span(\mathbf{b}, \ldots, A^{n-1}\mathbf{b})$$

Notice in all cases we have used that α_n and β_n are non-zero.

To check the orthogonality condition we use that $\mathbf{r}_n = \mathbf{r}_{n-1} - \alpha_n A\mathbf{v}_{n-1}$ So

$$\begin{aligned}\langle \mathbf{r}_n, \mathbf{r}_j \rangle &= \langle \mathbf{r}_{n-1} - \alpha_n A\mathbf{v}_{n-1}, \mathbf{r}_j \rangle \\ &= \langle \mathbf{r}_{n-1}, \mathbf{r}_j \rangle - \alpha_n \langle \mathbf{v}_{n-1}, A\mathbf{r}_j \rangle \\ &= \langle \mathbf{r}_{n-1}, \mathbf{r}_j \rangle - \alpha_n \langle \mathbf{v}_{n-1}, A\mathbf{v}_j - \beta_j A\mathbf{v}_{j-1} \rangle \\ &= \langle \mathbf{r}_{n-1}, \mathbf{r}_j \rangle - \alpha_n \langle \mathbf{v}_{n-1}, A\mathbf{v}_j \rangle - \alpha_n \beta_j \langle \mathbf{r}_{n-1}, A\mathbf{v}_{j-1} \rangle\end{aligned}$$

If $j < n-1$ then the right hand side is zero by induction. If $j = n-1$ then the last term is zero by induction and

$$\langle \mathbf{r}_{n-1}, \mathbf{r}_j \rangle - \alpha_n \langle \mathbf{v}_{n-1}, A\mathbf{v}_j \rangle = ||\mathbf{r}_{n-1}||_2^2 - \frac{||\mathbf{r}_{n-1}||_2^2}{||\mathbf{v}_{n-1}||_A^2} * \langle \mathbf{v}_{n-1}, A\mathbf{v}_{n-1} \rangle = 0$$

To prove the final "A-orthogonality". We use that $\mathbf{v}_n = \mathbf{r}_n + \beta_n \mathbf{v}_{n-1}$ to see that

$$\begin{aligned} \langle \mathbf{v}_n, A\mathbf{v}_j \rangle &= \langle \mathbf{r}_n, A\mathbf{v}_j \rangle + \beta_n \langle \mathbf{v}_{n-1}, A\mathbf{v}_j \rangle \\ &= \frac{1}{\alpha_{j+1}} \langle \mathbf{r}_n, \mathbf{r}_j \rangle - \frac{1}{\alpha_{j+1}} \langle \mathbf{r}_n, \mathbf{r}_{j+1} \rangle + \beta_n \langle \mathbf{v}_{n-1}, A\mathbf{v}_j \rangle \end{aligned}$$

If $j < n-1$ the right hand side is zero by induction (and by the above). If $j = n-1$ then the first term is zero and one computes from the values of α_n and β_n that the second two terms cancel. □

REMARK 1.2. The orthogonality and the spanning properties imply that $\mathbf{r}_n$ is orthogonal ot $\mathcal{K}_n$.

LEMMA 1.3. *In the CG iteration* $A\mathbf{E}_n = \mathbf{r}_n$.

PROOF. We prove by induction. $\mathbf{E}_0 = \mathbf{x}_* - \mathbf{x}_0 = \mathbf{x}_*$ and so $A\mathbf{E}_0 = \mathbf{b} = \mathbf{r}_0$. More generally,

$$\begin{aligned} \mathbf{r}_n &= \mathbf{r}_{n-1} - \alpha_n A\mathbf{v}_{n-1} \\ &= A\mathbf{E}_{n-1}\mathbf{x}_* - \alpha_n A\mathbf{v}_{n-1} \\ &= A(\mathbf{x}_* - \mathbf{x}_{n-1} - \alpha_n A\mathbf{v}_{n-1}) \\ &= A(\mathbf{x}_* - (\mathbf{x}_{n-1} + \alpha_n A\mathbf{v}_{n-1})) \\ &= A(\mathbf{x}_* - \mathbf{x}_n \\ &= A\mathbf{E}_n \end{aligned}$$

□

2. Optimality

We can now show the claim that conjugate gradients minimizes the required norms:

THEOREM 2.1. *Let the CG iteration be applied to a symmetric positive definite real matrix problem* $A\mathbf{x} = \mathbf{b}$. *If the interation hasn't converged, ie.* $\mathbf{r}_{n-1} \neq 0$, *then* $\mathbf{x}_n$ *is the unique point in* $\mathcal{K}_n$ *so that*

$$||\mathbf{E}_n||_A = ||\mathbf{x}_* - \mathbf{x}_n||_A < ||\mathbf{x}_* - \mathbf{y}||_A$$

for any $\mathbf{y} \in \mathcal{K}_n$ *with* $\mathbf{y} \neq \mathbf{x}_n$ *Moreover, the convergence is monotone in that*

$$||\mathbf{E}_n||_A \leq ||E_{n-1}||_A.$$

PROOF. By the previous Lemma one has $\mathbf{x}_n \in \mathcal{K}_n$. As we saw above $A\mathbf{E}_n = \mathbf{r}_n$. Hence for any point $\mathbf{y} \in \mathcal{K}_n$ we can write $\mathbf{y} = \mathbf{x}_n - \Delta\mathbf{x} \in \mathcal{K}_n$ so the error of $\mathbf{y}$ is $\mathbf{E} = \mathbf{x}_* - \mathbf{y} = \mathbf{E}_n + \Delta\mathbf{x}$. We compute

$$\begin{aligned} ||\mathbf{E}||_A^2 &= \langle \mathbf{E}_n + \Delta\mathbf{x}, A(\mathbf{E}_n + \Delta\mathbf{x}) \\ &= \langle \mathbf{E}_n, A\mathbf{E}_n \rangle + \langle \Delta\mathbf{x}, A(\Delta\mathbf{x}) \rangle + 2\langle \mathbf{E}_n, A(\Delta\mathbf{x} \rangle. \end{aligned}$$

By symmetry one has $\langle \mathbf{E}_n, A(\Delta \mathbf{x}\rangle = \langle A\mathbf{E}_n, \Delta \mathbf{x}\rangle = \langle \mathbf{r}_n, \Delta \mathbf{x}\rangle = 0$. By the positive definite property of A, one has $\langle \Delta \mathbf{x}, A(\Delta \mathbf{x})\rangle \geq 0$ with equality when only when $\Delta \mathbf{x} = 0$. This implies that

$$||\mathbf{E}||_A \geq ||\mathbf{E}_n||_A$$

with equality only when $\Delta \mathbf{x} = 0$. The monotonicity follows from this fact and the fact that $\mathcal{K}_{n-1} \subset \mathcal{K}_n$. □

Where does Conjugate Gradients come from? The answer comes from a minimization problem. Recall we to find $\mathbf{x}_n$ we are minimizing

$$||\mathbf{E}||_A^2 = ||\mathbf{x}_* - \mathbf{x}||_A^2$$

with $\mathbf{x} \in \mathcal{K}_n$. The problem is that computing this function depends on knowing $\mathbf{x}_*$! The insight one needs to have is that if we take the function

$$\begin{aligned} \phi : \mathbb{R}^m &\to \mathbb{R} \\ \mathbf{x} &\mapsto \frac{1}{2}\langle \mathbf{x}, A\mathbf{x}\rangle - \langle \mathbf{x}, \mathbf{b}\rangle \end{aligned}$$

then $||\mathbf{E}||_A^2 = ||\mathbf{x}_* - \mathbf{x}||_A^2 = 2\phi(\mathbf{x}) + \langle \mathbf{x}_*, \mathbf{b}\rangle$ Notice that the second term is constant (i.e. independent of $\mathbf{x}$) so won't effect the minimization. I.e. the vector for which ϕ is minimized is the same vector for which $||\mathbf{E}||_A$ is minimized.

That there is such a nice iteration scheme for finding the minimum of ϕ is a non-obvious fact.