Math 104 Fall 2008
Class notes

Laurent Demanet

Draft December 1, 2008

Preface

These class notes complement the class textbook *Numerical Linear Algebra* by Trefethen and Bau. They do not suffice in themselves as study material; it is only in combination with designated portions of the textbook that you have a complete picture of the material covered. The notes also show the order in which the subjects are taught in class.

Chapter 1

Preliminaries

This chapter contains an expanded discussion of selected concepts introduced in Lecture 1 of the textbook.

A vector x of dimension n is an n-tuple of numbers x_j with $1 \leq j \leq n$. The x_j are called the components of x. When the x_j are real, one can think of the vector x as either a point in the abstract space $\mathbb{R}^n$, or equivalently the arrow pointing from the origin to that point, and the geometrical intuition you have about the case $n = 3$ is perfectly adequate. Note that sometimes people (e.g. physicists) prefer to be more careful about the definition of vector, and say that it should be defined independently of a basis, but in this course we will simply identify a vector with its components in a basis. (The notion of basis will be explained later.)

Am m-by-n matrix A is an array of numbers A_{ij}, where the subscript i indexes the rows ($1 \leq i \leq m$) and j indexes the columns ($1 \leq j \leq n$). The result of the matrix-vector product A times x is the vector b whose components b_i are given by the sum

$$b_i = \sum_{j=1}^{n} A_{ij} x_j.$$

In short, $b = Ax$. In matrix notation, this also reads

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Notice that the components of x and b have been put in a column: when matrices are around we will almost always view a vector as being a column vector, or in other words as a n-by-1 matrix. We need to transpose it in order to obtain a row vector: x^T is defined as the 1-by-n matrix

$$(x_1, \ldots, x_n).$$

You should become familiar with switching between the index notations and the matrix notations. For instance, notice the role of the *free* index i and the *dummy* (or summation) index j. Dummy indices can be renamed. Any free index in the left-hand side of an expression also needs to appear as a free index in the right-hand side.

Many problems in science engineering can be cast as *linear systems* of equations, or perhaps simplified in order to be cast as linear systems. In turn, one can view such systems in matrix notation as $Ax = b$ and the problem is to find x given the matrix A and the vector b.

The components x_j of a vector, or the entries A_{ij} of a matrix are usually real-valued, but things are not much different when they are complex-valued, so the Trefethen-Bau textbook takes the general viewpoint of considering complex numbers. Even if the entries of a matrix are real-valued, complex numbers often come back through the back door when we calculate its eigenvalues (discussed later), so it's a good idea to be familiar with complex numbers. We reserve the word "scalar" for either a real number or a complex number. If you are confident about complex numbers, skip the next section.

1.1 A short review of complex numbers

Square roots of negative numbers are forbidden in the usual calculus of real numbers, so they have long puzzled and posed a problem to mathematicians. It was only in the mid 1800s that it was recognized that such things could no longer be avoided, for instance in computing the roots of polynomials. The consensus that emerged was to denote by i the imaginary unit, work with the abstract law

$$i^2 = -1,$$

and define the two square roots of -1 as $\pm i$. Like for negative numbers, familiarity would do the rest to make this concept concrete.

The choice of letter i is commonplace in math and physics, and I apologize in advance to the engineers who use the letter j instead.

Complex numbers are then defined as linear combinations of real and imaginary numbers, like $a + ib$ where a and b are real. The number a is called the real part and b is called the imaginary part. In the field of complex numbers $\mathbb{C}$, any polynomial of degree n has exactly n roots, possibly multiple. This result is the fundamental theorem of algebra.

Graphically, it is standard to think of a complex number $z = a + ib$ as a vector with two components a and b, and represent it in the plane $\mathbb{R}^2$. Thus the x axis is the real axis, and the y axis is the imaginary axis. Addition or subtraction of complex numbers is defined in an obvious manner through addition or subtraction of the real and imaginary part respectively; this corresponds to the standard rules of addition and subtraction of vectors in vector calculus. The peculiarity of complex numbers is the nontrivial multiplication law

$$(a + ib)(c + id) = ac + ibc + iad + i^2bd = ac - bd + i(bc + ad)$$

that has no equivalent in standard vector calculus. We will come back later to the notion of dot product for real vectors, which is a very different operation:

$$\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} c \\ d \end{pmatrix} = ac + bd.$$

Since complex numbers are identified to vectors, we can talk about the *modulus* ρ as being the length of that vector,

$$\rho = \sqrt{a^2 + b^2} \qquad \text{(Pythagorean theorem)},$$

and the *argument* θ as the angle made with the positive real axis,

$$\tan\theta = \frac{b}{a}.$$

The modulus is also denoted as $\rho = |z|$. The couple (ρ, θ) is the polar decomposition of the complex number $a + ib$.

The *complex conjugate* of a number $z = a + ib$ where a and b are real, is simply $a - ib$ and is denoted as $\overline{z}$. It is easy to check (exercise) that $|z|^2 = z\overline{z}$. It can also be checked (exercise) that when $z, z' \in \mathbb{C}$, then

$$\overline{zz'} = \overline{z}\,\overline{z'}.$$

Functions $f(z)$ of complex numbers $z = a + ib$ are defined by extending the Taylor expansion of $f(x)$ to the complex numbers, when this Taylor expansion exists. So if around a point x_0 we have

$$f(x) = \sum_{n \geq 0} c_n (x - x_0)^n,$$

for x in some neighborhood of x_0, then we simply write

$$f(z) = \sum_{n \geq 0} c_n (z - x_0)^n,$$

and then start worrying about convergence. For instance, it can be checked (exercise) that the exponential, sine and cosine functions are all well defined on complex numbers with everywhere-convergent Taylor expansions, and for $z = i\theta$, satisfy

$$e^{i\theta} = \cos\theta + i \sin\theta.$$

A neat interpretation of this quantity is that $e^{i\theta}$ is identified as the vector with components $\cos\theta$ and $\sin\theta$ in the plane; it points in a direction that makes an angle θ with the positive real axis, and has length one since

$$|e^{ib}| = \sqrt{\cos^2 b + \sin^2 b} = 1.$$

So we see that a complex number z is linked to its polar decomposition (ρ, θ) through

$$z = \rho e^{i\theta}.$$

This characterization also leads to the curious result that $e^{i\pi} = -1$.

1.2 Linear combinations, linear maps

Let us return to the interpretation of the matrix-vector product Ax. As explained on page 4 of the textbook, we can view $b = Ax$ as a *linear combination of columns of A*. If the columns of A are denoted a_j, then a linear combination of $a_1, \ldots, a_n$ with coefficients x_j is simply

$$b = \sum_{j=1}^{n} x_j a_j.$$

This is saying exactly the same thing as $b = Ax$. We say that x_j are the components of b in the collection of vectors a_j. (I'm holding off from talking about a basis of vectors instead of a collection, but that will come later.) For reference, adding vectors means adding the components of the first vector to the corresponding ones of the second vector. Multiplying a vector by a scalar means multiplying each of its components by that scalar.

The space generated by all possible linear combinations of some vectors a_j has a special name.

Definition 1.1. *We call* linear span *of the vectors* a_j*, or space generated by the vectors* a_j*, the space of all linear combinations* $\sum_j c_j a_j$ *of these vectors, where the coefficients* c_j *are scalars. It is denoted span*$\{a_j\}$*.*

Linear spans can be visualized graphically in the case when the scalars appearing in the linear combination are real:

- The linear span of a vector a_1 consists of all the multiples αa_1 for $\alpha \in \mathbb{R}$, hence if a_1 is not the zero vector, the linear span is a line through the origin and the point a_1.
- The linear span of two (nonzero) vectors a_1, a_2 consists of all the combinations $\alpha a_1 + \beta a_2$, hence it is a plane through the origin and the points a_1, a_2.
- Etc. in higher dimensions, we may talk about hyperplanes.

Example 1.1. *In* $\mathbb{R}^3$ *(*$n = 3$*), the vectors* $a_1 = (1, 0, 0)$ *and* $a_2 = (0, 1, 0)$ *span the xy-plane, i.e., the plane perpendicular to the z-axis and passing through the origin. If, to this collection, we add another vector whose third component is nonzero, then it will not lie in the xy-plane, and the linear span of* $(1, 0, 0)$, $(0, 1, 0)$ *and this new vector will be the whole space.*

The xy-plane is equivalently generated by $b_1 = (1, 1, 0)$ *and* $b_2 = (1, -1, 0)$*, among other choices. In order to (rigorously) prove this fact, consider any linear combination of* b_1 *and* b_2*:*

$$\alpha \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

By the usual rule of addition it can be rewritten as

$$(\alpha + \beta) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (\alpha - \beta) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

hence it is also a linear combination of a_1 *and* a_2*. Conversely, we also have to show that any linear combination of* a_1 *and* a_2 *is also a combination of* b_1 *and* b_2*, which we leave as an exercise to the reader.*

In view of the interpretation of $b = Ax$ as a linear combination of columns of A, a particularly important example of linear span is the span of all the columns of a matrix.

Definition 1.2. *The span of all the columns of a matrix is called the* range space *of the matrix, or simply range, or column space, and denoted range*(A)*. In other words, it is the set of all* b *that can be expressed as* Ax *for some* x*.*

The word "linear" has a deep significance in all these concepts, and in this course in general. A *vector space* is by definition a set of vectors that have the *linearity* property, i.e., when we take two vectors inside the set, then any linear combination of these vectors will also belong to the set. In other words, E is a vector space when, for any scalars α and β,

$$x, y \in E \quad \Rightarrow \quad \alpha x + \beta y \in E.$$

For instance, any linear span is manifestly a vector space (why?).

The other important instance of the word linear in this course is in the notion of *linear map*. A map is synonym to function from one set to another: a map M from a vector space E to another vector space F acts on $x \in E$ to output $M(x) \in F$. This map is linear when, for all $x, y \in E$, for all scalars α, β,

$$M(\alpha x + \beta y) = \alpha M(x) + \beta M(y).$$

For instance, any m-by-n matrix A defines a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ through matrix-vector multiplication. Indeed, it is easy to check from the definition that

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay.$$

This example is generic; matrices are the most general way of writing linear maps from one linear space to another. This fact is why matrix multiplication is defined the way it is!

1.3 Linear independence, dimension, rank

The number of vectors that generate a linear span is sometimes representative of the actual dimensionality of that linear space, but not always. For instance, we can continue the discussion of example 1.1 by considering not two, but three vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -1/2 \\ \sqrt{2} \\ 0 \end{pmatrix}.$$

These three vectors are coplanar, i.e., they lie in the same plane $z = 0$, because their third component is zero. As a result the linear span of these three vectors is still the xy-plane, which is two-dimensional. The discrepancy between the number of vectors and the dimensionality of their span lies in the fact that either vector among the three above is a linear combination of the other two, for instance

$$\begin{pmatrix} -1/2 \\ \sqrt{2} \\ 0 \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \sqrt{2}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

We could therefore eliminate either one of the three vectors from the collection without changing the linear span a bit.

The right notion of dimensionality stems when one considers vectors that are *linearly independent*, i.e., when no vector in the collection can be expressed as a linear combination of the others. The proper way to formalize this idea (without favoring any particular vector in the process) is the following definition.

Definition 1.3. *The vectors $a_1, \ldots, a_n$ are said to be linearly independent if, in case scalars $c_1, \ldots, c_n$ can be found such that*

$$c_1 a_1 + \ldots + c_n a_n = 0,$$

then necessarily $c_1 = \ldots = c_n = 0$.

Let us check that if vectors are not linearly independent in the sense of this definition, then they are linearly dependent. Indeed if there exists a nonzero solution to $c_1 a_1 + \ldots + c_n a_n = 0$, say with $c_1 \neq 0$ (without loss of generality), then we can isolate a_1 and write

$$a_1 = -\frac{1}{c_1}(c_2 a_2 + \ldots + c_n a_n),$$

which would mean that a_1 is a linear combination of the other vectors.

When considering the space spanned by a collection of linearly independent vectors, and only then, the number of vectors is a correct indicator of the "type" of space generated (line, plane, hyperplane): we simply refer to this number as the *dimension* of the space.

Definition 1.4. *Consider a vector space E spanned by n linearly independent vectors. Then n is called the dimension of that space. It is denoted dim(E).*

For example, a line has dimension 1, a plane has dimension 2, the whole space has dimension 3, and the concept extends to arbitrary linear subspaces of $\mathbb{R}^n$ now that we are armed with the notion of linear independence.

When a collection of vectors is linearly independent, then we say it is a *basis* for its linear span. More generally, we talk of a basis for a vector space when the following to conditions are met.

Definition 1.5. *A collection of vectors* $\{a_j\}$ *is called a* basis for a vector space E *when*

- *they are linearly independent; and*
- *they generate* E *(i.e.,* E *is their linear span).*

For instance

$$\{(1,0,0),(0,1,0)\}$$

is a basis for the xy-plane;

$$\{(1,0,0),(0,1,0),(0,0,1)\}$$

is a basis for the whole space $\mathbb{R}^3$, but

$$\{(1,0,0),(0,1,0),(3,4,0)\}$$

is not a basis of any space since the vectors are linearly dependent. The concept of basis is important and will come back later.

It is important to point out that, given a vector space, there may be different bases for that space, but the *number* of such basis vectors does not change. In other words the dimension is an intrinsic number, and does not depend on the choice of basis. We prove this result in Section 1.5. (And, strictly speaking we have been cavalier in writing the above definition of dimension without having previously established that it is independent of choice of basis.)

If we are in presence of some vectors that depend linearly on the others, then we may remove those vectors from the collection and still the linear span would remain the same. I prove this fact in footnote[1]. The proof is quite simple and may seem verbose to some, but I hope that it will also serve as one of many examples of proof writing for those of you unfamiliar with proofs.

We can apply these notions to the columns of a matrix. They need not always be linearly independent as the following example shows.

[1]By assumption, one vector depends linearly on the others. Without loss of generality,

Example 1.2. *Consider the matrix*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Its columns are linearly dependent, because

$$2 \begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix} - \begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix} - \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix} = 0.$$

We may therefore remove $(2, 5, 8)$ *from the set of columns without changing the linear span. The remaining two columns are obviously not collinear, so we conclude that the dimension of the range space of* A *is two.*

Definition 1.6. *The dimension of the range space of a matrix is called the* rank *of the matrix. It is also the number of linearly independent columns.*

If a matrix of size m-by-n has rank n, then we say it has full rank. If on the other hand the rank is $< n$ then we say the matrix is rank-deficient.

We will see in details in the following chapters at least two different ways of obtaining the rank of a matrix, and a basis for the range space, without

call this vector a_1 (otherwise relabel the vectors). For some scalars $c_2, \ldots, c_n$ we have

$$a_1 = c_2 a_2 + \ldots + c_n a_n.$$

We have to show that span $\{a_1, \ldots, a_n\}$ is the same as span $\{a_2, \ldots, a_n\}$. Two sets are equal when they are included in each other, so let us show the double inclusion.

- If a vector v belongs to span $\{a_2, \ldots, a_n\}$ then it is written as a linear combination of $a_2, \ldots, a_n$, which is a fortiori also a linear combination of $a_1, \ldots, a_n$ (with the coefficient of a_1 equal to zero). Hence the vector also belongs to span $\{a_1, \ldots, a_n\}$.
- If a vector v belongs to span $\{a_1, \ldots, a_n\}$, then

$$\begin{aligned} v &= \alpha_1 a_1 + \ldots + \alpha_n a_n \\ &= \alpha_1 (c_2 a_2 + \ldots + c_n a_n) + \alpha_2 a_2 + \ldots + \alpha_n a_n \\ &= (\alpha_1 c_2 + \alpha_2) a_2 + \ldots + (\alpha_1 c_n + \alpha_n) a_n. \end{aligned}$$

  So v is a linear combination of $a_2, \ldots, a_n$ and therefore belongs to span $\{a_2, \ldots, a_n\}$. We are done.

having to play the game of trying to identify columns that depend linearly on each other. It becomes quickly impractical for large matrices. Algorithms are needed instead (and they are quite elegant algorithms.)

Anytime we can find a linear combination of the columns of a matrix that vanishes (equals zero), then we are in presence of linearly dependent columns, and this is the signature that the rank is smaller than the actual number of columns. Call x_j the nonzero coefficients of the linear combination of columns that vanishes: in matrix notation this means

$$Ax = 0, \qquad \text{(here 0 means the zero vector in } \mathbb{R}^m \text{ or } \mathbb{C}^m\text{)}$$

There is a special name for the linear space of such x.

Definition 1.7. *The set of all x such that $Ax = 0$ is called the* nullspace *of the matrix A, or kernel of the matrix, and is denoted null(A).*

The dimension of the nullspace is the number of linearly independent x such that $Ax = 0$; it shows the number of "essentially different" ways of forming linear combinations of the columns of A that vanish. For example, if $x = 0$ is the only way that $Ax = 0$, then the nullspace consists of $\{x = 0\}$ (is zero-dimensional) and all the columns of A are linearly independent (the matrix is full-rank).

More generally, the dimension of null(A) is exactly the discrepancy between the rank of a matrix and the number of columns. We formulate this observation as a theorem.

Theorem 1.1. *Let A by m-by-n. Then dim(range(A)) + dim(null(A)) = n.*

We postpone the justification of this result until Chapter 3.

Another property of ranks is the following: the rank of an m-by-n matrix is at most equal to $\min(m, n)$, the minimum of m and n. If for instance n is greater than m and the matrix is horizontal-looking (or, wide,) then there is no way one can find more than m linearly independent columns among the n columns, because it is impossible to find more than m linearly independent vectors in $\mathbb{R}^m$. (If you are not convinced of this fact, see Section 1.5, Lemma 1.2 below for a discussion and a pointer to the proof.)

1.4 Invertibility of a matrix

Let us return to the basic question of solving a system of linear equations written in matrix form as $Ax = b$. The problem is to find x; whether that is possible or not will depend on the properties of A and b. Let us first consider a case in which things go wrong.

Example 1.3. *Again, consider*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

whose nullspace is

$$null(A) = \{\alpha \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} : \alpha \in \mathbb{C}\}.$$

Try to solve $Ax = b$. Here are two choices for b:

- *If*

  $$b = \begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix},$$

  then one solution is

  $$x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

  but so is the sum of this x_1 with any element of the nullspace:

  $$x_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}.$$

  So we have non-uniqueness of the solution, an undesirable situation.

- *If on the other hand $b \notin range(A)$, then by definition one cannot express b as a linear combination of columns of A, i.e., we cannot express b as Ax. In that case there does not exist a solution; this is another undesirable situation.*

As we'll make precise below, the problem has to do with the nonzero nullspace, i.e, rank-deficiency of the matrix A.

We say that the matrix A is *invertible*, or *nonsingular*, when for each b the equation $Ax = b$ has one and only one solution x. In other words, to each b corresponds a single x and vice-versa. There is a precise concept for this in math: we say that the matrix A defines via matrix-vector product a linear map, and that this map is a *bijection*, or bijective map, from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Definition 1.8. *A linear map M between two vector spaces E and F is called bijective if it is*

- injective*, or one-to-one: this means that if $M(x) = M(y)$, then $x = y$. Or equivalently, if $x \neq y$, then $M(x) \neq M(y)$. Said yet another way: M maps no two distinct vectors to the same vector.*
- surjective*, or onto: this means that for every $y \in F$, there exists $x \in E$ such that $M(x) = y$.*

The following result makes precise the intuition we developed earlier that invertibility is linked to the rank of a matrix.

Theorem 1.2. *A square, full-rank matrix A is invertible.*

Proof. As a linear map,

- is A surjective? Or, for every b, does there exist a solution x to $Ax = b$? Yes, because the columns of A generate $\mathbb{C}^n$, hence each b has a linear expansion in the columns of A, and the coefficients of this expansion are precisely the components of x.
- is A injective? Or, does $Ax = b$ have a unique solution? Assume that there exist x_1, x_2 such that

  $$Ax_1 = b, \qquad \text{and} \qquad Ax_2 = b.$$

  Then we can subtract the two equations and obtain $Ax_1 - Ax_2 = 0$, or by linearity. $A(x_1 - x_2) = 0$. Because the columns of A are linearly independent, this implies $x_1 - x_2 = 0 \Rightarrow x_1 = x_2$. So the two solutions must be the same, which establishes uniqueness.

□

Theorem 1.3. *A square invertible matrix A is full-rank.*

Proof. Assume by contradiction that the matrix isn't full-rank. Then the columns of A cannot be linearly independent. (Strictly speaking, this is the contraposition of Lemma 1.3.) So there exists $c \neq 0$ such that $Ac = 0$. Fix any $b \in \mathbb{C}^n$ and consider the system $Ax = b$. Since A is invertible, there exists a solution x^* to this system. (Here the star notation has nothing to do with adjoints.) But $x^* + c$ is also a solution, since $A(x^* + c) = Ax^* + Ac = b + 0 = b$. This contradicts uniqueness and therefore the assumption of invertibility. Thus the matrix must be full-rank. □

The proof of Theorem 1.2 points to an interesting interpretation of the solution x to $Ax = b$: the components of x are the coefficients of the expansion of b in the basis of the columns of A.

So we may in fact view solving a linear system as a change of basis for b. To explain this concept, introduce the canonical basis $\{e_j\}$ of $\mathbb{R}^n$ (or $\mathbb{C}^n$) as the columns of the identity matrix. In other words,

$$e_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where the 1 is in j-th position. Then the components of b can be seen, quite tautologically, as the components of its own expansion in the canonical basis. On the other hand x, again, contains the coefficients of the expansion of b in the columns of A.

This observation is well-explained in the textbook. It is also argued that, in order to form the coefficients in the system $\{e_j\}$ (that is, b), it suffices to multiply the coefficients in the system $\{a_j\}$ by the matrix A (that is, it suffices to form Ax). The reverse operation, going from b to x, is done by multiplying by the *inverse* of A:

$$b = A^{-1}x.$$

This is also an explicit form of the solution of the system $Ax = b$. The following result establishes the existence of the inverse matrix A^{-1}. So, in particular, it establishes that the map from b to x is linear.

Theorem 1.4. *Assume $A \in \mathbb{C}^{n\times n}$ (or $\mathbb{R}^{n\times n}$) is square and invertible. Then there exists a unique $B \in \mathbb{C}^{n\times n}$ (or $\mathbb{R}^{n\times n}$, respectively), such that*

$$AB = BA = I,$$

where I is the n-by-n identity matrix. [B is called the inverse matrix of A, and denoted $B = A^{-1}$.]

Proof. In three steps, let us show 1) the existence of B such that $AB = I$, 2) the existence of C such that $CA = I$, and 3) that $B = C$.

1. Consider $\{e_j\}$ the canonical basis of $\mathbb{C}^n$ as we did earlier. Since the columns of A form a basis by assumption (the matrix is invertible), we can expand

   $$e_j = \sum_{k=1}^{n} b_{kj} a_k,$$

   where a_k is the k-th column of A. In matrix notation, let b_{ij} be the (i, j) element of a matrix B. Then we have $I = AB$, as desired.

2. In order to show the existence of C such that $CA = I$, one would need to know that the rows of A are linearly independent, and proceed in a very similar manner. We postpone this question to Chapter 3.

3. Form CAB. On the one hand $C(AB) = CI = C$. On the other hand $(CA)B = IB = B$. So $B = C$.

□

There are many ways to check that a matrix is invertible, as in Theorem 1.3 on p.8 of the textbook. We need a few definitions before proceeding with it.

Definition 1.9. *(Eigenvalue, eigenvector) Let $A \in \mathbb{C}^{n\times n}$. The scalar $\lambda \in \mathbb{C}$ is called an eigenvalue of A, with eigenvector $v \in \mathbb{C}^n$, when*

$$Av = \lambda v.$$

There are a few different ways to introduce the determinant of a matrix, but the following definition is relatively simple in contrast to some others.

Definition 1.10. *(Determinant) This in an inductive definition on the size of the matrix: we'll assume we know the determinant for matrices of size n-by-n, and we'll prescribe what it is for matrices of size* $n+1$*-by-*$n+1$*. We need a base for this induction: if* $a \in \mathbb{C}$*, then* $det(a) = a$*. Now let* $A \in \mathbb{C}^{(n+1)\times(n+1)}$*.*

Consider B_{ij} *the submatrix of* A *where the* i*-th row and the* j*-th column have been removed. Let* A_{ij} *be the* (i,j) *element of* A*. Then, regardless of* i*, the following expression gives the determinant of* A*:*

$$det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij}\, det(B_{ij}).$$

(Strictly speaking, we'd have to show that indeed, the definition does not depend on the choice of row for calculating the determinant expansion, before stating this definition.)

Example 1.4.

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

Here is now a summary of what we know about invertibility of a matrix. Only the link to determinants has not been covered so far.

Theorem 1.5. *Let* $A \in \mathbb{C}^{n\times n}$ *(or* $\mathbb{R}^{n\times n}$*). The following are equivalent:*

- A *is invertible.*
- A *has an inverse* A^{-1} *such that* $AA^{-1} = A^{-1}A = I$*.*
- $rank(A) = n$*.*
- $range(A) = \mathbb{C}^n$*.*
- $null(A) = \{0\}$*.*
- 0 *is not an eigenvalue of* A*.*
- $det(A) \neq 0$*.*

We may also add "0 is not a singular value" to this list – we'll see later what singular values are.

1.5 Proofs: worked examples

You may find useful to get acquainted with the material in this section, in anticipation of the homework exercises that consist in doing math proofs.

First, a very general remark concerning the structure of a proof. It is never okay to write an argument by assuming that the result is true and deriving consequences from it. For instance, it would be like being asked to prove $A \Rightarrow B$ but you would instead show $B \Rightarrow A$. It could be occasionally okay to work your way backwards (and for instance find C such that $C \Rightarrow B$ and then $A \Rightarrow C$), but it is very important to keep track of the proper logic of the argument (direction of the implication.)

Now for some notations. "An element x belongs to a set E" is denoted $x \in E$. "A set E is included in a set F" is denoted $E \subset F$. Writing $E \subset F$ is equivalent to the statement "for all $x \in E$, then $x \in F$". The equality $E = F$ for sets means $E \subset F$ and $F \subset E$. An assertion A holds "if and only if" assertion B holds, means: $A \Leftrightarrow B$ (they are equivalent).

Some of the following results are very obvious, but being confident that they are obvious is rewarding. You may find it useful to call upon them when you solve an exercise. Their justifications contain common proof techniques, such as *induction*, and *contradiction*.

Lemma 1.1. *Every vector space has a basis.*

The above result is so foundational that it is in fact hard to prove! You need the "axiom of choice" for it. This is way out of scope for this course.

Lemma 1.2. *Let $A \in \mathbb{C}^{m\times n}$ with $m < n$ (the matrix is short and wide). Then there necessarily exists a nonzero solution to $Ax = 0$.*

Intuitively, if a matrix is short and wide, the corresponding system has more unknowns than equations, and it is natural to expect that there may be different ways of explaining the right-hand side identically equal to zero by means of different x's — or any right-hand side for that matter. A proof of this result involves reduction to a row echelon form and is not too difficult, but I won't reproduce it here. For reference, see the book "Introduction to linear algebra" by Gilbert Strang where row echelon forms are covered in great detail. (We may come back to this topic when we discuss solutions of linear systems.)

Lemma 1.3. *Let E be a vector space with $dim(E) = n$. If we can find n linearly independent vectors a_j in E, then they form a basis of E.*

Proof. We have to show that any element of E can be written as a linear combination of the a_j. Suppose instead that we cannot, i.e., that there exists a vector $v \in E$ that would be outside the span of the others, and let us show that this will lead to an absurdity. That's the principle of a proof by contradiction.

Let $n = \dim(E)$. If v is outside span $\{a_1, \ldots, a_n\}$, then it is linearly independent from the a_j, because if

$$c_1 a_1 + \ldots c_n a_n + dv = 0,$$

then $d = 0$ (v cannot be a linear combination of the others) and $c_1 = \ldots = c_n = 0$ (by linear independence of the a_j.) Then we are in presence of a collection $\{a_1, \ldots, a_n, v\}$ of $n+1$ linearly independent vectors in E. Create a n-by-$(n+1)$ matrix A with this collection of vectors as columns; the linear independence property is expressed as $Ax = 0 \Rightarrow x = 0$. This is in contradiction with Lemma 1.2, hence the desired contradiction. □

It is now a good point to prove a result mentioned earlier.

Lemma 1.4. *Any two bases of a vector space have the same number of elements.*

Proof. Let E be this vector space, and let $\{v_j | j = 1, \ldots, m\}$ and $\{w_j | j = 1, \ldots, n\}$ be two bases. By contradiction assume that $m < n$ (if $n < m$, swap v and w). Since $\{v_j\}$ is a basis, each w_k must be a combination of the v_j:

$$w_k = \sum_{j=1}^{m} c_{jk} v_j,$$

for some scalars c_{kj}. In matrix notations where vectors are placed in columns of matrices, we have

$$W = VC$$

Now C is a m-by-n matrix with $m < n$ (it is short and wide), with entries c_{jk}. By Lemma 1.2, there exists a nonzero solution x of $Cx = 0$. It is also a solution of $VCx = 0$, hence $Wx = 0$. This is to say that the w_k are linearly dependent, which contradicts our assumption that $\{w_k\}$ is a basis. This is the desired contradiction. □

A careful inspection of the proof of the previous result reveals that, if we are solely interested in the claim $m \geq n$, then the assumptions on v_j and w_k can be relaxed.

Corollary 1.1. *Let E be a vector space. Assume $\{v_j | 1 \leq j \leq m\} \subset E$ is generating E, and assume $\{w_k | 1 \leq k \leq n\} \subset E$ are linearly independent. Then $m \geq n$*

The proof of the following result is slightly more involved.

Lemma 1.5. *(The basis extension lemma) Let E be a vector space. If we can find $n < dim(E)$ linearly independent vectors a_j in E, then there exists a basis of E, which they are a part of. In other words, one can extend the a_j into a basis.*

Proof. Let $m = dim(E)$. Consider $\{b_j | 1 \leq j \leq m\}$ a basis of E; it exists by Lemma 1.1. Then the collection

$$C_m = \{a_1, \ldots, a_n, b_1, \ldots b_m\}$$

is clearly generating E, but there may be too many vectors. It suffices to show that, possibly after removing some vectors among the b's, we can obtain a basis. The proper way to proceed is to do a proof by induction.

Fix $k \geq 0$. Suppose that for this particular k, anytime we have a set

$$C_k = \{a_1, \ldots, a_n, v_1, \ldots v_k\}$$

of the vector space E, assumed to be a generating set, then it is possible to extend $a_1, \ldots, a_n$ into a basis by adjoining some v's among $v_1, \ldots, v_k$ to them. (I use the letter c instead of b because they may be different; it is a generic argument.) Then the proof by induction consists in showing that the same assertion is true if we replace k by $k + 1$. Of course we have to check first that the assertion is true for $k = 0$: that's the basis of the induction. When this is done, we will have a guarantee by "bootstrapping" that the assertion is true for any k, in particular $k = m$, which is what we want. We say that the induction is over the parameter k.

(Notice in passing that the induction assumption that "the assertion is valid for k" looks a bit like assuming that the result itself is true — something highly forbidden — but of course it is not the same thing at all.)

We start with the start of the induction: $k = 0$. In this case we are in presence of the set $\{a_1, \ldots, a_n\}$, which is linearly independent by assumption, and generating by the induction assumption. Hence it is a basis, trivially.

Let us now make the induction assumption that if we are in presence of a generating set with k extra vectors, a basis can be constructed by taking the a's and some well-chosen c's. Consider then the set

$$C_{k+1} = \{a_1, \ldots, a_n, v_1, \ldots v_{k+1}\}.$$

In case $a_1, \ldots, a_n$ would already generate E on their own, hence would be a basis, there would be nothing more to prove. So let us consider the case when $a_1, \ldots, a_n$ are not generating. Then at least one of the v_j does not belong to span $\{a_1, \ldots, a_n\}$, because otherwise span $\{a_1, \ldots, a_n\}$ = span $\{a_1, \ldots, a_n, v_1, \ldots, v_{k+1}\} = E$.

Let us call this vector v_1 without loss of generality — otherwise simply relabel the collection v_j. We claim that v_1 is linearly independent from the other vectors $a_1, \ldots, a_n$. Indeed, if

$$c_1 a_1 + \ldots + c_n a_n + d v_1 = 0,$$

then $d = 0$ (because v_1 cannot be a linear combination of the others) and $c_1 = \ldots = c_n = 0$ (because the a_j are linearly independent.) Append v_1 to the collection of a_j to form the set

$$\{a_1, \ldots, a_n, v_1\}.$$

It is a set of linearly independent vectors, as we have seen. There are precisely k remaining vectors $c_2, \ldots, c_{k+1}$, and together with the set $\{a_1, \ldots, a_n, c_1\}$, we are still in presence of C_{k+1}, a generating set. So we can apply the induction assumption and conclude that a basis can be extracted from this sequence C_{k+1}. This finishes the proof. □

Lemma 1.6. *Let E and F be two vector spaces. If $E \subset F$, then dim $E \leq$ dim F.*

Proof. Let $\{a_j\}$ be a basis for E. Then they are also linearly independent as elements of F. By the basis extension lemma, there exists a basis of F that contains the $\{a_j\}$. Hence dim (F) is at least greater than the cardinality of $\{a_j\}$. (The cardinality of a set is the number of elements in that set.) □

Lemma 1.7. *Let E and F be two vector spaces. If $E \subset F$ and dim $(E) =$ dim (F), then $E = F$.*

Proof. Let $\{a_j\}$ be a basis for E, with $j = 1, \ldots, \dim(E)$. They are also linearly independent as elements of F. Since $\dim(F) = \dim(E)$, we can invoke Lemma 1.3 and conclude that the a_j are generating F. Thus a_j is also a basis of F. A vector space is entirely specified by a basis that generates it, so E and F are equal. □

Here is a result complementary to Lemma 1.3.

Lemma 1.8. *Let E be a vector space with $dim(E) = n$. If we can find n generating vectors a_j in E, then they form a basis of E.*

Proof. It suffices to show that the a_j are linearly independent. By contradiction, assume they are not. Choose k such that

$$a_k \in \text{span}\{a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n\}.$$

As we saw earlier, removing a_k from E does not change the space spanned by the remaining vectors, hence

$$\text{span}\{a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n\} = E.$$

We are in presence of a sequence of $m = n - 1$ generating vectors in a vector space E which also contains, by the dimension assumption, a collection of n linearly independent vectors. This contradicts the result that $m \geq n$ from Corollary 1.1. The proof is complete. □

Lemma 1.9. *A matrix $A \in \mathbb{C}^{n \times n}$ is full-rank if and only if its columns form a basis of $\mathbb{C}^n$.*

Proof. Let us examine the claim full rank $\Rightarrow$ basis.

- Are the columns generating? The range space is a subset of $\mathbb{C}^n$, and its dimension is n, so we can apply Lemma 1.7 to conclude that $\text{range}(A) = \mathbb{C}^n$.

- Are the columns linearly indepedent? There are n of them and they are generating, hence they are necessarily linearly independent by Lemma 1.8.

Now let us examine basis $\Rightarrow$ full rank. Since we have a basis of the range space with n elements, it follows by the definition of dimension that the dimension of the range space is m, which means the matrix is full-rank. □

Chapter 2

Dot product, orthogonality

The material in this Chapter is partly covered in Lecture 2 of the textbook, so I will not repeat that part here. I will only give miscellaneous remarks and additions.

Here is a useful definition.

Definition 2.1. *The Kronecker delta, or Kronecker symbol, is defined as*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

In other words, δ_{ij} is the (i, j) element of the identity matrix.

With the intent of introducing the concept of proof by induction, let us solve exercise 2.2.

Example 2.1. . *Let $\{x_j : 1 \leq j \leq m\}$ be a set of m orthogonal vectors in some vector space. Let us show the Pythagorean law*

$$\| \sum_{j=1}^{m} x_j \|^2 = \sum_{j=1}^{m} \|x_j\|^2.$$

First, we can show the case $m = 2$ by hand:

$$\|x_1+x_2\|^2 = (x_1+x_2)^*(x_1+x_2) = x_1^*x_1+x_1^*x_2+x_2^*x_1+x_2^*x_2 = x_1^*x_1+x_2^*x_2 = \|x_1\|^2+\|x_2\|^2,$$

where we have used orthogonality of x_1 and x_2. The principle of a proof by induction is to assume that the result has already been established for

$m = n$—that's called the induction assumption—and then establish the result for $m = n + 1$. This works if we manage to also prove it for the lowest interesting value of m, here $m = 2$—that's the base of the induction. Then, when that is done, the result will be true for $m = 3, 4, 5$, etc.

So consider $m = n + 1$. We decompose

$$\begin{aligned}
\|\sum_{i=1}^{n+1} x_i\|^2 &= \left(\sum_{i=1}^{n} x_i \ + \ x_{n+1}\right)^* \left(\sum_{i=1}^{n} x_i \ + \ x_{n+1}\right) \\
&= \left((\sum_{i=1}^{n} x_i)^* \ + \ x_{n+1}^*\right) \left(\sum_{i=1}^{n} x_i \ + \ x_{n+1}\right) \\
&= (\sum_{i=1}^{n} x_i)^* (\sum_{i=1}^{n} x_i) + (\sum_{i=1}^{n} x_i)^* x_{n+1} + x_{n+1}^* (\sum_{i=1}^{n} x_i) + x_{n+1}^* x_{n+1} \\
&= (\sum_{i=1}^{n} x_i)^* (\sum_{i=1}^{n} x_i) + x_{n+1}^* x_{n+1} \qquad \textit{by orthogonality} \\
&= \sum_{i=1}^{n} \|x_i\|^2 + \|x_{n+1}\|^2 \qquad \textit{by the induction assumption.}
\end{aligned}$$

This proves the result for $m = n + 1$, and hence for all $m \geq 2$.

2.1 Orthogonal projections

Here is the definition of orthogonal projection.

Definition 2.2. *Let $\{u_1, \ldots, u_m\} \subset \mathbb{C}^n$ be an orthonormal set, with $m \leq n$. Then the orthogonal projection of $v \in \mathbb{C}^n$ onto $\{u_1, \ldots, u_m\}$ is defined as*

$$w = \sum_{i=1}^{m} (u_i^* v) u_i.$$

This definition only works if the u_i are orthonormal!

Proposition 2.1. *If $m = n$, then an orthonormal set $\{u_1, \ldots, u_m\} \subset \mathbb{C}^n$ in an orthonormal basis, and the orthogonal projection recovers the original vector v:*

$$v = \sum_{i=1}^{n} (u_i^* v) u_i.$$

We say that this formula is the orthogonal expansion of v in the basis $u_1, \ldots, u_m$.

Proof. Since $\{u_1, \ldots, u_n\}$ is orthonormal, Theorem 2.1 in the book asserts that the vectors are linearly independent. A collection of n linearly independent vectors in $\mathbb{C}^n$ is a basis, as we saw in Chapter 1. So for any $v \in \mathbb{C}$, there exists coefficients c_i such that

$$v = \sum_{i=1}^{n} c_i u_i.$$

Dot this equality with u_j:

$$u_j^* v = \sum_{i=1}^{n} c_i u_j^* u_i.$$

This is where the Kronecker symbol is useful; since $\{u_i\}$ is an orthobasis, we have $u_j * u_i = \delta_{ij}$, so the above relation simplifies to

$$u_j^* v = \sum_{i=1}^{n} c_i \delta_{ij} = c_j.$$

This is an explicit formula for the coefficients c_j that we can plug back in the expansion of v:

$$v = \sum_{i=1}^{n} (u_i^* v) u_i.$$

□

Here is probably the most useful characterization of orthogonal projections.

Proposition 2.2. *Let $\{u_1, \ldots, u_m\} \subset \mathbb{C}^n$ be an orthonormal set, with $m \leq n$. Then*

$$u_i^*(v - w) = 0, \qquad \forall i = 1, \ldots, m.$$

In other words, $v - w$ is orthogonal to span $\{u_1, \ldots, u_m\}$.

Proof.

$$\begin{aligned} u_i^*(v - w) &= u_i^*(v - \sum_{j=1}^{m} (u_j^* v) u_j) \\ &= u_i^* v - \sum_{j=1}^{m} (u_j^* v) u_i^* u_j \\ &= u_i^* v - u_i^* v = 0. \end{aligned}$$

□

Remark 2.1. *Interestingly, the result above is also true if the* u_i *are simply a basis (not necessarily orthonormal). We'll come back to this later.*

We saw that a unitary matrix is defined as having the vectors of an orthonormal basis as columns. We can then view the statement

$$u_i^* u_j = \delta_{ij}$$

in matrix notation as

$$U^* U = I,$$

or in other words $U^* = U^{-1}$. Since the inverse is unique and the same on the left and on the right, we also necessarily have $UU^* = I$.

Chapter 3

The four fundamental spaces of a matrix

This Chapter is not covered in the textbook, for the most part.

Let us return to a question that was left unanswered in Chapter 1: do the rank of a matrix and the dimension of its nullspace sum up to n, the total number of columns? That was Theorem 1.1.

The answer to this question will naturally bring about the four fundamental subspaces of a matrix. The upgrade from Chapter 1 is that we will now need to manipulate not just linearly independent columns, but linearly independent rows as well. The first thing to notice about rows is that if we pre-multiply a matrix A by a row vector x^T, then $x^T A$ has the interpretation of being a linear combination of rows.

Secondly, the rows of a matrix A are the columns of the transpose A^T (or adjoint A^*) of the matrix A. So our interpretation of mat-vec product as a linear combinations of either rows and columns is consistent if we notice that

$$x^T A = (A^T x)^T.$$

More generally, it is a good exercise involving manipulation of subscript notations to show that $B^T A^T = (AB)^T$, or $B^* A^* = (AB)^*$.

3.1 Column rank and row rank

We need to be a little careful with the notion of rank. We've seen that it is the number of linearly independent columns, but we haven't justified yet that it's also equal to the number of linearly independent rows (that's true – we'll come back to this later in this section.) So for the time being, let us denote the rank of A by c-rank(A)—for column rank—and denote the number of linearly independent rows by r-rank(A)—for row rank.

Using rows, we can also give another interpretation to elements in the nullspace of A: $Ax = 0$ means that x is orthogonal to all the rows of A.

Let us now move on to material that will be useful in the proof of Theorem 1.1. We will mostly manipulate real matrices in this chapter.

Lemma 3.1. *Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times n}$. If null($A$) = null($B$), then c-rank($A$) = c-rank($B$).*

(Notice that it would be false to claim that the two range spaces are the same, since they are subsets of different vectors spaces: range(A) $\subset \mathbb{R}^m$, while range(B) $\subset \mathbb{R}^p$. It is only the dimensions of these range spaces that are the same.)

Proof. Denote a = c-rank(A) and b = c-rank(B). Assume by contradiction that $a > b$. Let $\{i_1, \ldots, i_a\}$ be the indices (labels) of a subset of linearly independent columns of A. So if a_j denote the columns of A, the only way that

$$\sum_{j=1}^{a} c_j a_{i_j} = 0$$

is that all $c_j = 0$. Now, since the rank of B is strictly smaller than A, there must exist a linear combination of the columns $b_{i_1}, \ldots, b_{i_a}$ that vanishes:

$$\sum_{j=1}^{a} d_j b_{i_j} = 0, \qquad \text{not all } d_j = 0.$$

Form the vector $x \in \mathbb{R}^n$ in such a way that $x_{i_j} = d_{i_j}$ for $j = 1, \ldots, a$, and the other components are zero. Then the relation above concerning linear dependence of columns of B reads $Bx = 0$. Since by assumption the two nullspaces are equal, then for that particular x we must also have $Ax = 0$. Back in columns notations, this says

$$\sum_{j=1}^{a} d_j a_{i_j} = 0, \qquad \text{not all } d_j = 0,$$

which contradicts what we had said earlier concerning linear independence of those columns of A. So we have proved that $a \leq b$. The proof that $a \geq b$ is the same after interchanging the roles of A and B. □

Here is another preliminary result that we will need to invoke later.

Lemma 3.2. *Permuting rows of $A \in \mathbb{R}^{m \times n}$ changes neither r-rank(A), nor c-rank(A).*

Proof. That the row rank does not change is trivial, since the span of a collection of vectors does not depend on the order in which these vectors appear in the collection.

After permutation of the rows, the matrix A becomes B. Consider the nullspace of A; as we observed earlier, $Ax = 0$ means that x is orthogonal to all the rows of A. But the rows of B are also rows of A¡ albeit in a different order, so if $Ax = 0$ then $Bx = 0$ and vice versa. Thus null(A) = null (B). By the preceding lemma, this shows that c-rank(A) = c-rank(B) and we are done. □

Let us now prove that r-ank and c-rank are equal.

Theorem 3.1. *Let $A \in \mathbb{R}^{m \times n}$. Then r-rank$(A)$ = c-rank(A).*

Proof. Call r = r-rank(A) and c = c-rank(A). Permute rows of A such that its first r rows are linearly independent, where r is the row rank. Since permuting rows changes neither the row rank nor the column rank, we might as well (without loss of generality) assume that the matrix A comes in this form in the first place, i.e., with its first r rows linearly independent. So let us write

$$A = \begin{pmatrix} B \\ C \end{pmatrix},$$

where $B \in \mathbb{R}^{r \times n}$ and $C \in \mathbb{R}^{(m-r) \times n}$, and the rows of B are linearly independent. Since the rows of C must depend linearly on the rows of B, there exists a matrix T such that

$$C = TB$$

(To see why that's the case, consider the interpretation of $x^T B$ as a linear combination of rows of B and generalize this to a matrix product TB where each row of T generates a linear combination of rows of B.)

Then

$$A = \begin{pmatrix} B \\ TB \end{pmatrix}.$$

From this formula, we readily see that $Ax = 0 \Leftrightarrow Bx = 0$. Hence $\text{null}(A) = \text{null}(B)$. By Lemma 3.1, this implies

$$c = \text{c-rank}(A) = \text{c-rank}(B).$$

Now B is a r-by-n matrix, so the number of linearly independent columns cannot exceed the length (number of components) of each column, so c-rank$(B) \leq r$. This proves that $c \leq r$. The proof of $r \leq c$ is very analogous and involves repeating the same argument on A^T in place of A. □

With this result in the bank, we can now revisit two of the proofs that we had "left for Chapter 3".

- One theorem said that if a matrix is invertible then there exists a unique inverse matrix, and it is the same on the left and on the right. With the equivalence of a matrix being full-column-rank and full-row-rank, we can now fully justify the theorem. (Exercise)

- The other theorem is what we set out to justify in the beginning of this Chapter, namely

$$\text{rank}(A) + \dim(\text{null}(A)) = n.$$

3.2 The sum of two vector spaces

But we still need a few more notions before we can address the latter point.

Definition 3.1. *The sum of two vector spaces is defined as*

$$V + W = \{v + w : v \in V, w \in W\}.$$

Example 3.1. *Consider*

$$V = span\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \},$$

and

$$W = span\{\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}\}.$$

Then

$$\begin{aligned} V + W &= span\{\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}\} \\ &= \{\alpha \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}; \alpha, \beta \in \mathbb{R}\} \\ &= \{\begin{pmatrix} -\alpha \\ -\beta \\ \alpha + \beta \end{pmatrix} : \alpha, \beta \in \mathbb{R}\}. \end{aligned}$$

It is the subset of vectors in $\mathbb{R}^3$ for which the third component is minus the first, minus the second one, or in other words $z = -x - y$, i.e., the plane $x + y + z = 0$.

Definition 3.2. *Let V, W be two vector spaces. W is said to be orthogonal to V (or vice-versa) if*

$$\forall v \in V, \quad \forall w \in W, \qquad v \cdot w = 0.$$

Lemma 3.3. *Let V and W be two orthogonal vector spaces. Then their intersection is $V \cap W = \{0\}$.*

Proof. Let $x \in V \cap W$. Since on the one hand $x \in V$, and on the other hand $x \in W$, we can take the dot product of x with itself and obtain

$$x \cdot x = 0.$$

This means $\|x\| = 0$, which implies $x = 0$. □

Lemma 3.4. *Let V, W be two orthogonal vector spaces. Then*

$$dim(V + W) = dim(V) + dim(W).$$

Proof. Let $\{v_j\}$ be a basis for V, and $\{w_k\}$ be a basis for W. It suffices to check that

$$\{v_j\} \cup \{w_k\}$$

is a basis for $V+W$, because then the dimension of $V+W$ will be the number of elements in that compound basis, which is the sum of the elements in each basis taken separately.

Hence we need to check two things:

- Together, are the vectors in $\{v_j\} \cup \{w_k\}$ generating $V+W$? Let $x \in V+W$. By definition of $V+W$, there exist $v \in V$, $w \in W$ such that $x = v + w$. Since $\{v_j\}$ is a basis for V, we can find some coefficients such that

  $$v = \sum c_j v_j,$$

  and similarly since $\{w_k\}$ is a basis for W,

  $$w = \sum d_k w_k.$$

  Hence $x = \sum c_j v_j + \sum d_k w_k$, which means $\{v_j\} \cup \{w_k\}$ is generating.

- Are the vectors in $\{v_j\} \cup \{w_k\}$ linearly independent? Assume that for some c_j, d_k,

  $$\sum c_j v_j + \sum d_k w_k = 0.$$

  Then we write

  $$\sum c_j v_j = -\sum d_k w_k = \sum (-d_k) w_k$$

  So we are in presence of an element $x = \sum c_j v_j$ of the space V, which is also an element of W since $x = \sum (-d_k) w_k$. So $x \in V \cup W$. We saw in the previous lemma that the only possible way of having an element in the intersection of two orthogonal spaces is that it be equal to zero:

  $$x = 0 \qquad \Rightarrow \qquad \sum c_j v_j = \sum (-d_k) w_k = 0.$$

  We can now consider each relation $\sum c_j v_j = 0$ and $\sum (-d_k) w_k = 0$ in isolation and use linear independence of $\{v_j\}$ and $\{w_k\}$ to obtain that

  $$\text{all } c_j = \text{all } d_k = 0.$$

This concludes the proof. □

Remark 3.1. *In the more general case when the two spaces need not be orthogonal, their intersection may be non-empty and we have*

$$dim(V + W) = dim(V) + dim(W) - dim(V \cap W).$$

(Exercise)

3.3 The orthogonal complement of a vector subspace

The largest orthogonal space to a given vector space is called its orthogonal complement.

Definition 3.3. *Let $V \subset \mathbb{R}^n$ be a vector space. Its orthogonal complement is defined as*

$$V^\perp = \{x \in \mathbb{R}^n : x \cdot v = 0 \quad \forall v \in V\}.$$

The two typical examples of orthogonal complements that one can easily visualize in $\mathbb{R}^3$ are

- Given a plane passing through the origin, its orthogonal complement is the line going through the origin and perpendicular to the plane;
- Given a line passing through the origin, its orthogonal complement is the plane going through the origin and perpendicular to the line.

Here is a obvious lemma, but its constructive proof is interesting. We will address that proof later in the course, but we need the lemma now.

Lemma 3.5. *Any vector space contains an orthonormal basis.*

Here is a useful fact about orthogonal complements.

Lemma 3.6. *Let $V \in \mathbb{R}^n$. then*

$$V + V^\perp = \mathbb{R}^n.$$

In other words, any element of $\mathbb{R}^n$ decomposes as the sum of an element of V, and an element of $V^\perp$. That's because $V^\perp$ is the largest vector space that is orthogonal to V.

Proof. The inclusion $V + V^\perp \subset \mathbb{R}^n$ is obvious, since vectors in V and in $V^\perp$ are all vectors of $\mathbb{R}^n$. For the reverse inclusion, let us proceed by contradiction. Assume that there is an element $x \in R^n$ such that

$$x \notin V + V^\perp.$$

Consider $\{v_j : j = 1, \ldots, m\}$ an orthonormal basis of V with $m \leq n$; we can do so by the preceding lemma. Consider the orthogonal projection $\tilde{x}$ of x onto V:

$$\tilde{x} = \sum_{i=1} (v_j^* x) v_j.$$

By Lemma 2.2, we have the orthogonality property

$$v_j^*(x - \tilde{x}) = 0, \qquad \forall j = 1, \ldots, m.$$

This is exactly saying that $x - \tilde{x} \in V^\perp$. But then we can decompose x into

$$x = \tilde{x} + (x - \tilde{x}).$$

The first term is an element of V, whereas the second term is an element of W. So $x \in V + V^\perp$, a contradiction. This finishes the proof.

□

Remark 3.2. *We didn't really need orthogonality of the v_j above, since the formula $\tilde{x} = \sum_{i=1} (v_j^* x) v_j$ was not used explicitly. But we haven't seen orthogonal projections with general bases yet.*

Corollary 3.1. *If $V \subset \mathbb{R}^n$ is a vector space, then*

$$dim(V) + dim(V^\perp) = n.$$

Proof. We only need to put together the result that $V + V^\perp = \mathbb{R}^n$ with the result that

$$\dim(V) + \dim(W) = \dim(V + W).$$

□

3.4 The four fundamental subspaces

We now come to the bottom line. The point of introducing orthogonal complements, for us, is that by construction, the nullspace of a matrix is orthogonal to the rows of that matrix A, and therefore is the orthogonal complement of the space spanned by the rows.

We call the space spanned by the rows the row space of a matrix. It is denoted

$$\mathrm{row}(A),$$

and if we interchangeably view vectors as either row vectors or column vectors, we have the characterization that

$$\mathrm{row}(A) = \mathrm{col}(A^T) = \mathrm{range}(A^T).$$

I will use col instead of range for the range space, in this discussion. Let us now consider what happens in $\mathbb{R}^n$ (target space of A as a linear map), and in $\mathbb{R}^m$ respectively (domain space of the linear map).

- As a subspace of $\mathbb{R}^n$, the nullspace of A is

  $$\mathrm{null}(A) = \{x \in \mathbb{R}^n : Ax = 0\},$$

  and manifestly, we have

  $$\mathrm{null}(A) = (\mathrm{row}(A))^{\perp}.$$

  Notice that if $A \in \mathbb{R}^{m\times n}$, $\mathrm{row}(A)$ and $\mathrm{null}(A)$ are both subspaces of $\mathbb{R}^n$. It follows from the theory we have developed so far that

  $$\mathrm{row}(A) + \mathrm{null}(A) = \mathbb{R}^n,$$

  and in terms of dimensions, that

  $$\dim(\mathrm{row}(A)) + \dim(\mathrm{null}(A)) = n.$$

- Another useful space is the *left-nullspace*, defined as

  $$\text{l-null}(A) = \{x \in \mathbb{R}^m : x^T A = 0\},$$

  which is the same thing as saying that $\text{l-null}(A) = \mathrm{null}(A^T)$. Manifestly,

  $$\text{l-null}(A) = (\mathrm{col}(A))^{\perp}.$$

Notice that if $A \in \mathbb{R}^{m\times n}$, $\text{col}(A)$ and $\text{l-null}(A)$ are both subspaces of $\mathbb{R}^m$. It follows from the theory we have developed so far that

$$\text{col}(A) + \text{l-null}(A) = \mathbb{R}^m,$$

and that

$$\dim(\text{col}(A)) + \dim(\text{l-null}(A)) = m.$$

The assertions in the two bullets above are sometimes called the Fundamental Theorem of Linear Algebra, and we have already provided its proof since every statement follows from some other lemma seen earlier in the Chapter.

The four fundamental subspaces of a matrix are precisely those that we have introduced:

$$\text{row}(A), \quad \text{col}(A), \quad \text{null}(A), \quad \text{and l-null}(A).$$

We can now go back to the claim concerning dimensions of range and nullspace:

$$\dim(\text{col}(A)) + \dim(\text{null}(A)) = n.$$

As such, it does not immediately follow from the theory of orthogonal subspaces since $\text{col}(A)$ and $\text{null}(A)$ are not orthogonal. If the matrix is rectangular $m \neq n$, these subspaces are not even included in the same $\mathbb{R}^n$. But we have seen that

$$\dim(\text{row}(A)) + \dim(\text{null}(A)) = n.$$

On the other hand, we know from Lemma 3.1 that

$$(\text{r-rank}(A) =) \quad \dim(\text{row}(A)) = \dim(\text{col}(A)) \quad (= \text{c-rank}(A)),$$

so we can substitute column rank for row rank and conclude that, indeed, $\dim(\text{col}(A)) + \dim(\text{null}(A)) = n$.

Let us finish this chapter with one last important result concerning orthogonal subspaces.

Proposition 3.1. *Let $V \subset \mathbb{R}^n$ be a vector space. Then*

$$(V^{\perp})^{\perp} = V.$$

Proof. The inclusion $V \subset (V^{\perp})^{\perp}$ is obvious: any vector in V is orthogonal to any vector in $V^{\perp}$, by definition of orthogonal complement. For the reverse inclusion, denote $m = \dim(V)$. Since

$$V + V^{\perp} = \mathbb{R}^n,$$

we have

$$\dim(V) + \dim(V^{\perp}) = n,$$

so $\dim(V^{\perp}) = n - m$. Now, we also have

$$V^{\perp} + (V^{\perp})^{\perp} = \mathbb{R}^n,$$

so

$$\dim(V^{\perp}) + \dim((V^{\perp})^{\perp}) = n,$$

and $\dim(V^{\perp}) = n-(n-m) = m$. Hence we are in presence of V, a subset of $(V^{\perp})^{\perp}$, yet both spaces have the same dimension, m. By Lemma 1.7, these spaces must be the same.

□

Chapter 4

Change of basis and change of coordinates

We saw earlier that solving $Ax = b$ for x has an interpretation in terms of "change of basis". Namely if the vector b is seen as consisting of its own components in the canonical basis e_j, then x can be interpreted as the components of b in the basis of the columns of A.

It can be surprising to read that the matrix for updating the components of b is in fact the inverse of A, whereas the way of changing the basis vectors e_j into a_j involves no inverse whatsoever. To understand this phenomenon properly, we need to differentiate between *change of basis* and *change of coordinates*. The following exposition is typically also presented in Physics classes, and turns out to be the proper axiomatic approach to subjects such as general relativity.

4.1 Change of basis and change of coordinates for vectors

In this chapter, and only in this chapter, we will be very careful to discriminate between a vector $\mathbf{x}$ (we underlined those as $\underline{x}$ on the black board, instead of using boldface), and its components $x = (x_1, \ldots, x_n)$ (ordered in a row or column—that does not matter here.) The components of a vector depend on a choice of basis, but a vector is an invariant object, like an arrow in $\mathbb{R}^2$ or $\mathbb{R}^3$, and does not depend on the choice of basis.

Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the canonical basis of $\mathbb{R}^n$. A change of basis means making a new choice $\{\mathbf{e}'_1, \ldots, \mathbf{e}'_n\}$ of the basis vectors. The coordinates of a vector $\mathbf{x}$ in the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ are denoted $x_1, \ldots, x_n$, while the coordinates of the same vector in the basis $\{\mathbf{e}'_1, \ldots, \mathbf{e}'_n\}$ are denoted $x'_1, \ldots x'_n$. That the same vector $\mathbf{x}$ has different components (coordinates) in different bases means that

$$\mathbf{x} = x_1\mathbf{e}_1 + \ldots + x_n\mathbf{e}_n = x'_1\mathbf{e}'_1 + \ldots + x'_n\mathbf{e}'_n$$

This is called an invariance principle, and expresses the idea that the notion of vector does not depend on any particular choice of basis, (and ultimately, that the laws of Physics should be formulated in an invariant manner as well).

A linear change of coordinates is the relationship

$$x' = Px \qquad \Leftrightarrow \qquad x'_i = \sum_j P_{ij} x_j$$

between coordinates in two different bases. It is a matrix-vector product involving a matrix P.

A linear change of basis is the expression

$$\text{“}\mathbf{e}' = Q\mathbf{e}\text{”} \qquad \Leftrightarrow \qquad \mathbf{e}'_i = \sum_j Q_{ij}\mathbf{e}_j.$$

We are in presence of a linear relationship between old basis vectors $\mathbf{e}_j$ and new basis vectors $\mathbf{e}'_j$ involving a matrix Q, but it is *not a matrix vector product*. Rather, the last expression involves vectors indexed by i or j instead of involving scalars as it did earlier for coordinates. Don't let the simple form of $\mathbf{e}' = Q\mathbf{e}$ fool you—this is why I wrote it in quotes.

We may now ask what is the relationship between P and Q. Using the invariance principle,

$$\begin{aligned}
\mathbf{x} = \sum_i x_i \mathbf{e}_i &= \sum_i x'_i \mathbf{e}'_i \\
&= \sum_i (\sum_j P_{ij} x_j)(\sum_k Q_{ik} \mathbf{e}_k) \\
&= \sum_j \sum_k (\sum_i P_{ij} Q_{ik}) x_j \mathbf{e}_k \\
&= \sum_j \sum_k (\sum_i (P^T)_{ji} Q_{ik}) x_j \mathbf{e}_k \\
&= \sum_j \sum_k (P^T Q)_{jk} x_j \mathbf{e}_k.
\end{aligned}$$

But this quantity also equals $\sum x_j \mathbf{e}_j$. Since x_j cannot (generically) be the component of any other basis vector than $\mathbf{e}_k$ for $k = j$, and the component of $\mathbf{e}_j$ is not just any multiple of x_j, but x_j itself, then necessarily

$$(P^T Q)_{jk} = \delta_{jk}.$$

In matrix notations,

$$P^T Q = I,$$

or

$$Q = (P^T)^{-1}.$$

It is a good, short exercise to prove that $(P^T)^{-1} = (P^{-1})^T$, so we simply denote the inverse transpose by P^{-T}.

Example 4.1. *Let us see how these considerations relate to our discussion of systems $Ax = b$ of linear equations. In the notations used in this chapter, b_j are given the meaning of components of a vector $\mathbf{b}$ through $\mathbf{b} = \sum_j b_j \mathbf{e}_j$. The system $Ax = b$ means that x_j are the components of $\mathbf{b}$ in the basis $\mathbf{a}_j$ of the columns of A:*

$$\mathbf{b} = \sum_j x_j \mathbf{a}_j.$$

So we may write $\mathbf{e}'_j = \mathbf{a}_j$ for the new basis vectors, and $b'_j = x_j$ for the new coordinates.

Let us check that the matrices P of change of coordinates, and Q of change of basis, are in fact inverse transpose of each other:

- *For the coordinate change,* $b'_j = x_j = (A^{-1}b)_j$*, so we have* $P = A^{-1}$*.*
- *For the change of basis, we seek a matrix* Q_{ij} *such that*

$$\mathbf{a}_i = \sum_j Q_{ij}\mathbf{e}_j.$$

In words, we are in presence of the decomposition of $\mathbf{a}_i$ *as a super-spoition of the canonical basis vectors* $\mathbf{e}_j$ *with coefficients* Q_{ij}*. So* $\mathbf{e}_j$ *is responsible for giving* $\mathbf{a}_i$ *its* j*-th component, and setting it equal to* Q_{ij}*. The* j*-th component of* $\mathbf{a}_i$ *is the element at row* j *and column* i *in the original matrix* A*, which is* A_{ji}*. So* $Q_{ij} = A_{ji}$*, i.e.,* $Q = A^T$*.*

We indeed have $A^{-1} = P = Q^{-T}$*.*

Example 4.2. *Let us now consider the special case of rotations in the plane. We can rotate* $\{\mathbf{e}_1, \mathbf{e}_2\}$ *by an angle* θ *in the counter-clockwise direction. A picture reveals that*

$$\mathbf{e}'_1 = \cos\theta\,\mathbf{e}_1 + \sin\theta\,\mathbf{e}_2,$$

$$\mathbf{e}'_2 = -\sin\theta\,\mathbf{e}_1 + \cos\theta\,\mathbf{e}_2,$$

from which the matrix Q *of change of basis is*

$$Q = R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

To which change of coordinates does this transformation correpsond? Since R_θ *is unitary, we have*

$$P = Q^{-T} = R_{-\theta}^{-1} = R_\theta,$$

so that the change of coordinates is written as

$$x'_1 = \cos\theta\,x_1 + \sin\theta\,x_2,$$

$$x'_2 = -\sin\theta\,x_1 + \cos\theta\,x_2.$$

This could also be obviously obtained form a picture, but our approach of looking for matrices P *and* Q *such that* $P^{-1} = Q^T$ *has the advantage of working for arbitrary invertible matrices.*

A transformation according to a change of basis matrix Q is called a "covariant" transformation by physicists, while a transformation according to a change of coordinates matrix $P = Q^{-T}$ is called "contravariant". These distinctions are the basis of tensor algebra, as well as analysis on manifolds, and also form the notational foundations of general relativity. In other application fields such as continuum mechanics, the decision is sometimes made to consider only unitary change of basis, for which $P = Q$, in which case the distinction between covariant and contravariant objects is not present in the notations.

Note that everything that we have explained above works for complex matrices if we substitute adjoints A^* for transposes A^T.

4.2 Change of coordinates for matrices

The notion of change of coordinates, or change of components, also makes sense for a matrix. It follows from the notion that a matrix is a linear transformation between vectors. People also speak of "change of basis" for matrices, but it would take us too far into tensor calculus to try and make this precise. So for matrices we'll stick with change of components, or coordinates.

If under some linear invertible transformation P, vectors have components that change according to

$$x' = Px,$$

then any vector also has its components change according to the same matrix. In particular, if we let $y = Ax$, then under the same linear transformation we necessarily have

$$y' = Py = PAx.$$

Naturally, we let A' denote the matrix that helps pass from x' to y':

$$y' = A'x'.$$

Combining the above expressions, we have

$$y' = PAx = A'Px$$

Since P is invertible, and the above holds for all x, then

$$A' = PAP^{-1}.$$

This is the relation that defines changes of components of a matrix A under a linear transformation P; it is called a *similarity transformation*. It does not just involve the matrix P on the left, like for vectors, but also the inverse matrix P^{-1} on the right.

Chapter 5

Matrix norms

The reference for matrix norms is **Lecture 3 in the textbook.**

Matrix norms are useful because they give a notion of distance between matrices:

$$d(A, B) = \|A - B\|.$$

Each notion of matrix norm gives a different way of measuring how good an approximation is. This will be the most important application for us.

One result is quoted but not proved in the textbook: the all-important Cauchy-Schwarz inequality. (It's Hermann Schwarz here, not Laurent Schwartz.)

Theorem 5.1. *(Cauchy-Schwarz inequality) Let $x, y \in \mathbb{C}^n$. Then*

$$|x^* y| \leq \|x\| \, \|y\|,$$

where the norm used is the usual Euclidean, 2-norm.

Proof. Assume that $y \neq 0$, otherwise the inequality is trivial.

Consider some $\lambda \in \mathbb{C}$, for which we have yet to make a choice. Then

$$\begin{aligned} 0 &\leq \|x - \lambda y\|^2 \\ &= (x - \lambda y)^* (x - \lambda y) \\ &= x^* x - \overline{\lambda} y^* x - \lambda x^* y + \lambda \overline{\lambda} y^* y \\ &= x^* x - \overline{\lambda} \overline{x^* y} - \lambda x^* y + |\lambda|^2 y^* y. \end{aligned}$$

We have used $x^* y = \overline{y^* x}$. Now we make a choice for λ so that this quantity is as small as possible, hence given the first inequality, as informative as possible. Consider

$$\lambda = \frac{y^* x}{y^* y}.$$

The denominator does not vanish by assumption. We substitute this value above to obtain

$$0 \geq x^*x - \frac{|x^*y|^2}{y^*y} - \frac{|x^*y|^2}{y^*y} - \frac{|x^*y|^2}{(y^*y)^2} y^*y$$

The last two terms cancel out, hence

$$0 \geq x^*x - \frac{|x^*y|^2}{y^*y}.$$

After rearranging, this is

$$\|x\|^2 \|y\|^2 \geq |x^*y|^2,$$

After taking a square root, it is what we wanted to show. □

Chapter 6

Orthogonal projectors

The reference for orthogonal projectors is **Lecture 6 in the textbook.**

There is one proof that we have not touched upon, and which is not part of the material. It is the one that concerns the implication "if a projector is *orthogonal*, then $P = P^*$". (It is the part of the proof that uses the SVD.) We only proved the converse in class, namely if $P = P^*$, we have $\text{range}(P) \perp \text{range}(I - P)$, which means that P is an orthogonal projector.

The topic of an isometry as a unitary matrix restricted to a certain subset of columns is not covered in the textbook, so let us cover it here.

For a square matrix Q, the relation $Q^*Q = I$ is the definition of Q being unitary, and implies that $Q^{-1} = Q^*$, hence implies the other relation $QQ^* = I$.

For tall-and-thin rectangular matrices, the situation is more complicated. A matrix that satisfies $Q^*Q = I$ is called an isometry, because it preserves lengths:

$$\|Qx\|_2^2 = x^*Q^*Qx = x^*x = \|x\|_2^2.$$

But, unless Q is square, this does not imply $QQ^* = I$.

We can understand this phenomenon as follows. For unitary matrices,

- $Q^*Q = I$ is a statement of orthogonality of the columns q_i of Q:

$$(Q^*Q)_{ij} = q_i^* q_j = \delta_{ij}.$$

- $QQ^* = I$ is a statement of resolution of the identity, or

$$\sum_i q_i q_i^* = I.$$

Now if a matrix $Q \in \mathbb{C}^{m\times n}$ with $m \geq n$ is only an isometry,

- we have by definition $Q^*Q = I$, which means that the n columns of Q are still orthonormal,
- but $QQ^* \neq I$, and $QQ^* \in \mathbb{C}^{m\times m}$, because there are only n vectors to draw from. Together, they cannot span the whole of $\mathbb{C}^m$. Instead, we let

  $$P = QQ^* = \sum_{i=1}^{n} q_i q_i^*.$$

  It is easy to see that P is the orthogonal projector onto range$(Q) =$ span$\{q_1, \ldots, q_n\}$, because $P^2 = QQ^*QQ^* = QQ^* = P$ and $P^* = (QQ^*)^* = QQ^* = P$.

Example 6.1. *Any m-by-n submatrix of a m-by-m unitary matrix, when $m \geq n$, is an isometry. For instance,*

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \end{pmatrix}$$

are all isometries.

Chapter 7

Least-squares, SVD, and QR

The reference for least-squares is **Lecture 11**. The reference for the SVD is **Lectures 4 and 5**. The reference for QR is **Lectures 7 and 8**, although we only covered Gram-Schmidt orthogonalization in Lecture 8. In class, we made back-and-forths between Lecture 11 and the other lectures, in order to pick up SVD and QR before applying them to least-squares.

The two applications presented in class are the following.

- The story of blood pressure as a function of age, for journalists vs. university professors, is taken from the book *Statistics with applications in biology and geology,* available for casual browsing on google books.

- The story of Gauss's prediction of Ceres's orbit, and his invention of least-squares, can be read from wikipedia and other Internet sources.

In both examples, least-squares are used to solve a polynomial fitting problem. How matrices arise from these problems is well explained in the textbook.

For the SVD, the textbook contains all the relevant material, except for the discussion of the relationship with the four fundamental subspaces. It goes as follows.

For $A \in \mathbb{C}^{m \times n}$, let $A = U \Sigma V^*$, where U, V are unitary and Σ is diagonal with positive entries along the diagonal, sorted in decreasing order. Denote the left singular vectors — the columns of U — by $u_i, 1 \leq i \leq m$, and the right singular vectors — the columns of V — by $v_i, 1 \leq i \leq n$. Spot the location of the last nonzero singular values, and call it σ_r; we have

$$\sigma_1 \geq \ldots \geq \sigma_r > 0, \qquad \sigma_{r+1} = \ldots = \sigma_{\max(m,n)} = 0.$$

The four fundamental subspaces can be read off from the SVD of A.

- By orthogonality, $V^* v_j = \mathbf{e}_j$, so $\Sigma V^* v_j = \Sigma \mathbf{e}_j = \sigma_j \mathbf{e}_j$. Therefore $A v_j = 0$ as soon as $j \geq r+1$, because of the zero singular values for $j \geq r+1$. So
$$\{v_j : r+1 \leq j \leq n\} \subset \text{ null } (A).$$
- By a similar argument, $u_j^* U \Sigma = \sigma_j \mathbf{e}_j^*$ is zero as soon as $j \geq r+1$, so
$$\{u_j : r+1 \leq j \leq m\} \subset \text{ l-null } (A).$$
- Now any vector u_1 through u_r can be reached by letting A act on an appropriate vector, namely $u_j = A \frac{v_j}{\sigma_j}$. This works because $\sigma_j > 0$ for $j \leq r$. So
$$\{u_j : 1 \leq j \leq r\} \subset \text{ range } (A).$$
- Similarly, any vector v_1^* through v_r^* can be obtained by letting A act on an appropriate vector on the left — which corresponds to forming linear combinations of rows. Namely, $v_j^* = \frac{u_j^*}{\sigma_j} A$. So
$$\{v_j : 1 \leq j \leq r\} \subset \text{ row } (A).$$

Since we know that the dimensions of row(A) and null(A) add up to n, and the dimensions of range(A) and l-null(A) add up to m, these subspaces could not possibly contain more vectors than the linear combinations of the collections specified above, hence the inclusions are in fact equalities. The number r emerges as the rank of A.

Notice also that the relations of orthogonality between the relevant subspaces are built in the SVD.