Lecture Notes Refresher Course: Linear Algebra

Cindy Catherine Orozco Bohorquez

Fall 2017

Abstract

These notes are designed for the refresher course of Linear Algebra for the incoming ICME students during Fall 2017. The idea of this extremely short course is to have a warm up in basic Linear Algebra concepts that will be fundamental for the future courses taken in ICME and other related fields, such as computer science, engineering and statistics. Therefore we assume that the students have already taken an undergraduate course on Linear Algebra, but probably they have not used the concepts during the last years.

This material is based on previous versions of the course, especially the tasteful ones developed by Karianne Bergen, Victor Minden and Ron Estrin (thank you guys).

My first Linear Algebra book to return to the basics:

- B. Fraleigh, R. Beauregard. *Linear Algebra.* Addison-Wesley Publishing Company, 1995.

The reference books of CME 302: Numerical Linear algebra, for more advance/fun/-computational stuff (The first one is the encyclopedia for all the algorithms, the second one has a softer approach):

- G. Golub, C. Van Loan, *Matrix Computations*, John Hopkins University Press, 2013
- L. N. Trefethen, D. Bau, *Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997.

If you have any questions, comments or suggestions, you can find me at orozcocc@stanford.edu

Enjoy the material!

PS (A bit about me): I am a third year PhD Student at ICME working on tensors with Professor Lexing Ying. When I arrived two years ago, I took this course with Danielle Maddix and Lan Huong Nguyen, and it was extremely useful and fun. I am from Bogotá, Colombia and I love painting.

Contents

Chapter 1

Introduction: Why Linear Algebra?

I took my first Linear Algebra course during the first year of my engineering studies, and although it was fun, it was a bit confusing why I had to learn about vector spaces, eigenvectors, linear transformations and decompositions in order to construct buildings and roads. With time (and sometimes pain) I learned that all these concepts are used to solve **linear systems** which are important because:

- They are **easy to solve**: We just need to "invert" the matrix, and we know if there exist solutions, and which are those
- They are **easy to compute**: The algorithms to solve them can be decomposed into basic sums and products = easy to program in a computer.
- They have a **strong theoretical background**, at the point that they serve as inspiration of other branches of mathematics, such as Functional Analysis and Abstract Algebra.
- They are widely used to **solve differential equations**, the basic language of physics and engineering problems.
- Lately, they have been rediscovered to **extract deterministic properties** of a subject set. Think about random walks and Principal Component Analysis.

Arguably, the core of all the concepts in Linear Algebra is the **matrix** (the mathematical object, not the movie - lol), which can be seen as:

- A purely **algebraic object** with a sum, multiplication and product by scalar
- An element of a **vector space** with an inner product and a trace.
- An **operator** in vector spaces representing a **linear transformation**, sometimes with very unique properties such us orthogonality or positive definiteness, and also with canonical representations encoded in eigenvalues and singular values

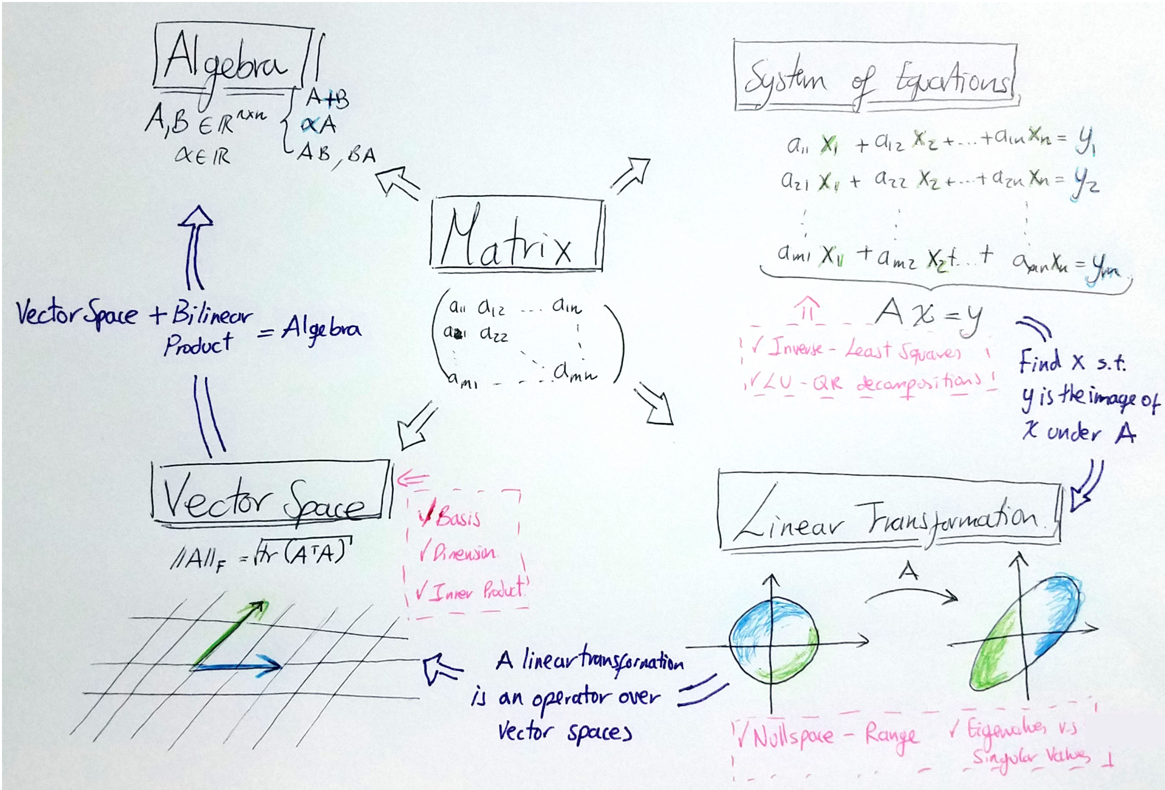


Figure 1.1: Different approaches to matrices in Linear Algebra

- A representation of a **linear system of equations**, which theoretically is solved finding its inverse (or more generally minimizing a least squares problem), but computationally uses multiple decompositions, such as LU and QR.

Therefore, during this course we will divide the lectures into these four chapters, to go from the pure mathematical object to the efficient solution of engineering problems.

Chapter 2

Matrices as an Algebra: Basic game rules

Learning how to operate matrices is the simplest subject of all four interpretations of a matrix because it only requires sums and products. Nevertheless, the study of algebras as a field of mathematics has inspired some of the most theoretical contributions.

We will restrict ourselves to the operational side of the algebra: how sum, product by scalar and multiplication is defined.[1]

2.1 Definition of Matrix

A ***matrix*** over a field $\mathbb{F}$ (For example $\mathbb{R}$ =real or $\mathbb{C}$ =complex) is a two-dimensional array of elements of the field.

For example

$$A = \begin{bmatrix} 2 & 3i & -\pi \\ 0 & -2.1 & 0.5 \\ -1 & 0 & 1/3 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{C}^{4\times 3}$$

is a complex matrix with 4 ***rows*** and 3 ***columns***

In particular, a matrix $x \in \mathbb{F}^{1\times n}$ is called ***row vector*** and a matrix $y \in \mathbb{F}^{m\times 1}$ is called ***column vector*** and a matrix $\alpha \in \mathbb{F}^{1\times 1} = \mathbb{F}$ is called a ***scalar***.[2]

[1] For curiosity, an ***algebra*** over a field is a vector space with a bilinear product, that in this case is matrix multiplication

[2] Notice that for scalars we use *Greek* letters, for vectors we use *lowercase Latin* letters, and for the rest of matrices *uppercase Latin* letters

Therefore we can see A as a set of row vectors or as a set of column vectors:

$$A = \begin{bmatrix} [2 & 3i & -\pi] \\ [0 & -2.1 & 0.5] \\ [-1 & 0 & 1/3] \\ [0 & 0 & 0] \end{bmatrix} \qquad A = \begin{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -1 \\ 0 \end{bmatrix} & \begin{bmatrix} 3i \\ -2.1 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} -\pi \\ 0.5 \\ 1/3 \\ 0 \end{bmatrix} \end{bmatrix}$$

Symbolically we will denote that matrix with m rows and n columns belongs to $\mathbb{F}^{m\times n}$. The element of the row i and column j is $a_{i,j}$ and therefore[3]:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots \vdots & \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} = \begin{bmatrix} --a_1^\top-- \\ --a_2^\top-- \\ \vdots \\ --a_m^\top-- \end{bmatrix} = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_n \\ | & | & & | \\ | & | & & | \end{bmatrix} \tag{2.1}$$

Therefore the third row vector and the second column vector of A are

$$a_3^\top = \begin{bmatrix} -1 & 0 & 1/3 \end{bmatrix} \qquad a_2 = \begin{bmatrix} 3i \\ -2.1 \\ 0 \\ 0 \end{bmatrix}$$

2.2 Sum/Product by scalar/Multiplication

Given any scalar $\alpha \in \mathbb{F}$ and any matrix $B \in \mathbb{F}^{m\times n}$ for any number of m rows and n columns, we define the ***product by scalar*** element by element as:

$$(\alpha B)_{i,j} = \alpha * b_{i,j} \tag{2.2}$$

For example

$$(-2i)\begin{bmatrix} 2 & 3i & -\pi \\ 0 & -2.1 & 0.5 \\ -1 & 0 & 1/3 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -4i & 6 & 2\pi i \\ 0 & 4.2i & -1i \\ 2i & 0 & -2/3i \\ 0 & 0 & 0 \end{bmatrix}$$

Now, if we have two matrices of the same size $A, B \in \mathbb{F}^{m\times n}$ then the ***sum*** is defined element by element as:

$$(A + B)_{i,j} = a_{i,j} + b_{i,j} \tag{2.3}$$

[3] I agree there is a bit of abuse of notation here because $a_1^\top \neq (a_1)^\top$ (the row vector vs. the transpose of the column vector), but in most of the cases it will be clear from the context

For example

$$\begin{bmatrix} 2 & 3i & -\pi \\ 0 & -2.1 & 0.5 \\ -1 & 0 & 1/3 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1+3i & -\pi \\ 1 & -2.1 & 0.5 \\ -2 & 0 & 1/3 \\ 1 & 0 & 0 \end{bmatrix}$$

Up to know, we have described the basic operations of a Vector Space (as we will explain in Chapter 3). But the definition of a multiplication is what it makes it an algebra. As we can expect, the multiplication is closely related with the idea of Linear Transformation (in Chapter 4), and that is one of the reasons why it is not simply defined as the product element by element.

Given two matrices $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{p\times q}$, the ***multiplication*** $AB \in \mathbb{F}^{m\times q}$ is defined iff $n = p$ such that:

$$(AB)_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j} \quad \text{for} \quad 1 \le i \le m \ , \ 1 \le j \le q \tag{2.4}$$

For example

$$\begin{bmatrix} 2 & 3i & -\pi \\ 0 & -2.1 & 0.5 \\ -1 & 0 & 1/3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ i & -1 \\ 100 & 0 \end{bmatrix} = \begin{bmatrix} (3i)(i) + (-\pi)(100) & (2)(1) + (3i)(-1) \\ (-2.1)(i) + (0.5)(100) & (2.1)(-1) \\ (1/3)(100) & (-1)(1) \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -3 - 100\pi & 2 - 3i \\ 50 - 2.1i & -2.1 \\ 100/3 & -1 \\ 0 & 0 \end{bmatrix}$$

With this example is clear to see that the multiplication is **not commutative**, i.e. $AB \neq BA$. In the previous example, even BA is not defined because $A \in \mathbb{C}^{4\times 3}$ and $B \in \mathbb{C}^{3\times 2}$ and $4 \neq 2$.[4]

Equation (2.4) is really easy to understand computationally but it gives little intuition from where it comes from. Let's think in the row vectors of A and in the column vectors of B, then we can see that

$$\begin{bmatrix} - - a_1^\top - - \\ - - a_2^\top - - \\ \vdots \\ - - a_m^\top - - \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \dots & b_q \\ | & | & & | \\ | & | & & | \end{bmatrix} = \begin{bmatrix} (a_1^\top)(b_1) & (a_1^\top)(b_2) & \dots & (a_1^\top)(b_q) \\ (a_2^\top)(b_1) & (a_2^\top)(b_2) & \dots & (a_2^\top)(b_q) \\ \vdots & \vdots & \ddots \vdots & \\ (a_m^\top)(b_1) & (a_m^\top)(b_2) & \dots & (a_m^\top)(b_q) \end{bmatrix}$$

[4]For the rest, the sum is commutative and associative, and it is distributive under both product by scalar and matrix multiplication. Multiplication is also associative.

where $(a_i^\top)(b_j)$ is the matrix multiplication between the row vector of A: $a_i^\top \in \mathbb{F}^{1\times n}$, and the column vector of B: $b_j \in \mathbb{F}^{n\times 1}$. In Chapter 3 we will see that this special multiplication corresponds to the dot product (inner product) between vectors.

Instead if we think on the column vectors of A and the row vectors of B we have:

$$\begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_n \\ | & | & & | \\ | & | & & | \end{bmatrix} \begin{bmatrix} - - b_1^\top - - \\ - - b_2^\top - - \\ \vdots \\ - - b_n^\top - - \end{bmatrix} = \sum_{k=1}^{n} \begin{bmatrix} | \\ a_k \\ | \\ | \end{bmatrix} \begin{bmatrix} - - b_k^\top - - \end{bmatrix}$$

here we have the matrix multiplication between column vector of A: $a_i \in \mathbb{F}^{m\times 1}$, and row vector of B: $b_j^\top \in \mathbb{F}^{1\times q}$. In Chapter 3 we will see that this special multiplication corresponds to outer product between vectors.[5]

Finally, we can also write equation (2.4) combining matrices and vectors, as right and left actions:

$$\begin{bmatrix} - - a_1^\top - - \\ - - a_2^\top - - \\ \vdots \\ - - a_m^\top - - \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \dots & b_q \\ | & | & & | \\ | & | & & | \end{bmatrix} = \begin{bmatrix} - - a_1^\top B - - \\ - - a_2^\top B - - \\ \vdots \\ - - a_m^\top B - - \end{bmatrix}$$
$$= \begin{bmatrix} | & | & & | \\ A\,b_1 & A\,b_2 & \dots & A\,b_q \\ | & | & & | \\ | & | & & | \end{bmatrix}$$

Comment: Depending on the application, it is useful to think in a matrix of matrices or ***block matrix***. Sum and Matrix product remain the same, as long as the conditions are satisfied in both the submatrices and the block matrix. For example

$$\begin{bmatrix} L_{1,1} & 0 \\ L_{2,1} & L_{2,2} \end{bmatrix} \begin{bmatrix} U_{1,1} & U_{1,2} \\ 0 & U_{2,2} \end{bmatrix} = \begin{bmatrix} L_{1,1}U_{1,1} & L_{1,1}U_{1,2} \\ L_{2,1}U_{1,1} & L_{2,1}U_{1,2} + L_{2,2}U_{2,2} \end{bmatrix}$$

where $L_{1,1}, L_{2,1}, L_{2,2}, U_{1,1}, U_{1,2}, U_{2,2}$ are matrices with an appropriate size

[5] Although theoretically the result is the same regardless of the decomposition, computationally it makes a big difference in the algorithms regarding memory allocation and parallelization.

Chapter 3

Matrices and Vector Spaces: Universal Language

In Chapter 2 we talked about row and column vectors as a particular case of matrices. Nevertheless this implies that vectors with different orientations are different (such as it is the case in MATLAB). For both cases, we can think in a simpler way, looking to a vector as a one-dimensional array belonging to $\mathbb{R}^n$ (as it is the case in Python).

Vector Spaces are algebraic structures that behave as $\mathbb{R}^n$, and even more, in finite dimension, we can show that $\mathbb{R}^n$ are the only vector spaces that exists modulo coordinatization. On the other hand, infinite dimension vector spaces (such as the set of real functions) have its own branch in mathematics called Functional Analysis. Although the basic rules are the same, having infinite dimensions changes their behavior.

3.1 Vector Space: Coordinates

A ***vector space*** over a field $\mathbb{F}$ is a set V with defined (let $u, v, w \in V, \alpha, \beta \in \mathbb{F}$):

- **addition** $+ : V \times V \to V$
  - associative $u + (v + w) = (u + v) + w$
  - commutative $u + v = v + u$
  - there exists identity $0 : \forall u \ 0 + u = u$
  - $\forall u$ there exists inverse $-u : u + (-u) = 0$
- **product by scalar** $* : \mathbb{F} \times V \to V$
  - Distributivity $(\alpha + \beta)u = \alpha u + \beta u$ and $\alpha(u + v) = \alpha u + \alpha v$
  - identity $1 \in \mathbb{F} : 1u = u$

Its objects are called ***vectors***. Subsets closed under addition and product by scalar are called ***subspaces***

For example:

- $\mathbb{C}^n = \{[a_1, .., a_n] | a_i \in \mathbb{C}\}$ is a vector space
- $\mathbb{C}^{m\times n} = \left\{ \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix} \middle| a_{i,j} \in \mathbb{C} \right\}$ is a vector space
- $P_n(x) = \{\sum_{i=0}^n a_i x^i | a_i \in \mathbb{R}\}$ is a vector space
- $\mathcal{C}[0,1] = \{f : [0,1] \to \mathbb{R}|\ f \text{ continuous}\}$ is a vector space that contains $\{f \in C[0,1] | f[0] = f[1] = 0\}$ as subspace

3.1.1 Span and Dimension

The set of all linear combinations of a subset of vectors is a subspace called ***span***

$$\mathrm{span}(\{v_1, ..., v_k\}) = \left\{ \sum_{i=1}^k \alpha_i v_i \middle| \alpha_i \in \mathbb{F} \right\} \tag{3.1}$$

For example

$$\mathrm{span}(\{x, \sin(x)\}) = \{\alpha_1 x + \alpha_2 \sin(x) |\ \alpha_i \in \mathbb{F}\}$$

Nevertheless, it may happen that there is **redundancy** in the subset of vectors, i.e. not all of them are necessary to express the same span. For example

$$\mathrm{span}(\{x, \sin(x)\}) = \mathrm{span}(\{x, 3x + 2\sin(x), \sin(x)\}) = \mathrm{span}(\{3x + 2\sin(x), \sin(x)\})$$

To identify redundacy we say that a set of vectors $\{v_i\}_{i=1}^k$ is ***linearly dependent*** if there exist $\alpha_i \in \mathbb{F}$ not all zero, such that

$$\sum_{i=1}^k \alpha_i v_i = 0 \tag{3.2}$$

implying that we can express one of them as a linear combination of the others. Otherwise the set is ***linearly independent***.

Therefore a ***basis*** of a vector space V is a linearly independent set of vectors $\{v_i\}_{i=1}^k$ such that

$$V = \mathrm{span}\left(\{v_i\}_{i=1}^k\right) = \left\{ v = \sum_{i=1}^k \alpha_i v_i \middle| \alpha_i \in \mathbb{F} \right\}$$

The number of vectors in a basis of vector space V is called ***dimension***. Moreover, given a basis, there is a unique set of coefficients in $\mathbb{F}^k$ for each element of the vector space. Therefore it is equivalent to work with the elements of V or with their counterpart in $\mathbb{F}^k$. This is called ***coordinatization*** and make vector spaces universal. Notice that k can be infinite, but for now we will consider only $k < \infty$

3.2 Normed Space: Length

To make things more interesting, we introduce ***Normed Spaces*** which are Vector Spaces with a norm defined. Notice that we can define different norms over the same vector space, and we will have different normed spaces.
A ***norm*** over a vector space V is a function $||\cdot|| : V \to \mathbb{R}^+ \cup \{0\}$ such that (for $u, v \in V, \alpha \in \mathbb{F}$)

- Non degenerate $||v|| = 0 \iff v = 0$
- Absolute homogeneity $||\alpha v|| = |\alpha| \; ||v||$
- Subadditivity: ***Triangular inequality*** $||u + v|| \leq ||u|| + ||v||$

From this properties we can conclude the ***reverse triangular inequality***

$$| \; ||u|| - ||v|| \; | \leq ||u - v||$$

For example

- $(\mathbb{R}^n, ||\cdot||_\infty = \max_{i=1,..,n} |a_i|)$
- $(\mathcal{C}[0,1], ||\cdot||_\infty = \sup\{|f(x)| : x \in [0,1]\})$

3.2.1 p-norms

A set of famous norms are the ***p-norms*** in $\mathbb{C}^n$ define for $p \geq 1$ real number as:

$$||v||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \tag{3.3}$$

taking the limit as $p \to \infty$ we get ∞-norm

$$||v||_\infty = \max_{i=1,\dots,n} |x_i| \tag{3.4}$$

For example:

- Let $v \in \mathbb{C}^n$ then $||v||_1 = \sum_{i=1}^n |v_i|$
- Let $A \in \mathbb{C}^{m\times n}$, we can identify it with $v(A) \in \mathbb{C}^{mn}$. Then $||v(A)||_2 = \sqrt{\sum_{k=1}^{mn} v(A)_k^2} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j}^2} = ||A||_f$ the ***Frobenius Norm***

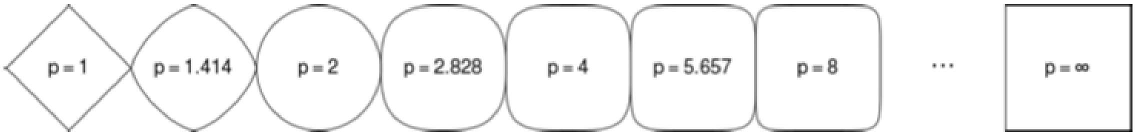


Figure 3.1: Unit ball for different values of p. Courtesy of Victor Minden

3.3 Inner Product Space: Angle

In $\mathbb{R}^2$ to relate two vectors, we usually use angles to know if they are collinear or orthogonal. We can generalize this idea including an inner product. An ***inner product space*** is a vector space V with a map $< \cdot, \cdot >: V \times V \to \mathbb{F}$ such that (for $u, v, w \in V$ and $\alpha \in \mathbb{F}$):

- Conjugate symmetry: $\langle u, v\rangle = \overline{\langle v, u\rangle}$
- Linearity in first argument: $\langle \alpha u + v, w\rangle = \alpha\langle u, w\rangle + \langle v, w\rangle$
- Positive Definiteness: $\langle u, u\rangle \geq 0$. $\langle u, u\rangle = 0 \iff u = 0$

From these axioms we can prove[1] ***Cauchy Schwartz inequality***, for any $u, v \in V$

$$|\langle u, v\rangle|^2 \leq \langle u, u\rangle\langle v, v\rangle \tag{3.5}$$

Once we have equation (3.5), we can prove[2] that any inner product induce a norm

$$||u|| = \sqrt{\langle u, u\rangle} \tag{3.6}$$

Therefore all inner product spaces are normed spaces, but not the converse. For example of all the p-norms in $\mathbb{C}^n$ only the 2-norm is induced by the standard inner product in $\mathbb{C}^n$:

$$\langle u, v\rangle = \sum_{i=1}^{n} u_i \overline{v_i}$$

In particular the Frobenius norm of matrices come from the inner product in $\mathbb{R}^{m\times n}$ as [3]

$$\langle A, B\rangle = \mathrm{tr}(A^* B)$$

We define the angle θ between 2 vectors such that

$$\cos\theta = \frac{|\langle u, v\rangle|}{||u||\ ||v||} \tag{3.7}$$

because of Cauchy Schwartz we know that $\cos(\theta)$ is between -1 and 1, and we say the vectors are ***collinear*** if $\theta = 0, \pi$ and ***orthogonal*** if $|\theta| = \pi/2$

[1] To prove if $v = 0$ it is clear, for $v \neq 0$ consider $z = u - \frac{\langle u,v\rangle}{\langle v,v\rangle} v$ and massage the inequality

[2] Checking the axioms

[3] $tr(A) = trace(A) = \sum_i a_{ii}$ and $(A^*)_{i,j} = \overline{a_{j,i}}$

3.3.1 Orthogonal basis

In particular we say that a basis is a ***orthogonal basis*** if all the vectors are orthogonal pairwise.

We say that an orthogonal basis is an ***orthonormal basis*** if all the vectors have norm one. For example the canonical basis of $\mathbb{R}^n$: $(e_i)_j = \delta_{i,j}$

3.3.2 Vector Projection

We define the ***vector projection*** of u into a vector v as

$$proj_v(u) = \frac{\langle u, v\rangle}{\langle v, v\rangle} v = (||u|| \cos\theta)\hat{v} \tag{3.8}$$

where $\hat{v} = v/||v||$. We can show that $proj_v(u)$ is the collinear vector to v that minimizes $||u - proj_v(u)||$, and moreover $u - proj_v(u)$ is orthogonal to v.

Thanks to this property, we can generate an orthonormal basis starting from any basis following the ***Gram-Schmidt*** process as [4]

- Choose $u_1 = v_1/||v_1||$ from
- For $j = 2, .., n$ take

$$u_j = \frac{\tilde{v}_j}{||\tilde{v}_j||} \text{ where } \tilde{v}_j = v_j - \sum_{i=1}^{j-1} proj_{u_i}(v_j)$$

On the other hand, notice that if we have an orthogonal basis $\{v\}_{i=1}^n$ of V, we can easily compute the corresponding coordinates for any vector $u \in V$ because:

$$u = \sum_{i=1}^n \mathrm{proj}_{v_i}(u) = \sum_{i=1}^n \frac{\langle u, v_i\rangle}{\langle v_i, v_i\rangle} v_i = \sum_{i=1}^n \alpha_i v_i$$

Comment: Usually, the notion of orthogonality in $\mathbb{C}$ is related with the standard inner product. Nevertheless, orthogonality is defined for any inner product. This is really useful when we consider change of coordinates. For example: the energy norm in physics.

[4] In computational algorithms, we change the order of the sum of projections to have numerical stability, having 2 variants of the ***Modified Gram-Schmidt*** method

Chapter 4

Matrices as Linear Transformations: Defining Interactions

Linear Algebra is not about matrices. It is about linear transformations. In the first 2 chapters we talked how matrices are a vector space, and even an algebra. But their main job in linear algebra is to enable computations related with any finite dimensional vector space. So instead of looking at them as objects of $\mathbb{F}^{m \times n}$, we will see them as maps $\mathbb{F}^n \to \mathbb{F}^m$.

A **linear transformation** from a vector space V to a vector space U is a map $T : V \to U$ such that for $v, w \in V$, $\alpha, \beta \in \mathbb{F}$:

$$T(\alpha v + \beta w) = \alpha T(v) + \beta T(w) \tag{4.1}$$

For example

- Identity map $T(v) = v$, Zero map $T(v) = 0$
- Differentiation of polynomials $\partial : P_n(x) \to P_{n-1}(x)$ such that $p(x) \rightarrowtail p'(x)$

This linear map define 2 important sets:

- ***Range*** is the Image of V under the map T: $\mathcal{R}(T) = \{u = T(v) | v \in V\} \subset U$. The dimension of $\mathcal{R}$ is called ***rank***.
- ***Nullspace*** or kernel is the preimage of $0 \in U$ under T: $\mathcal{N}(T) = \{v | T(v) = 0\} \subset V$.

From here we get the ***Rank-nullity*** theorem[1]:

$$dim(V) = dim(\mathcal{R}(T)) + dim(\mathcal{N}(T)) \tag{4.2}$$

[1] To prove construct a basis $\{v_i^1\}$ the for nullspace, and then extend it to a basis for V $\{v_i^1\} \cup \{v_j^2\}$, show that $\forall v \in V : T(v) = \sum_j \beta_j T(v_j^2)$ and then show $\{T(v^2)_j\}$ is a basis for the range

4.1 From Linear Transformations to Matrices

In section 3.1.1 we introduce **coordinatization**, which is the process to identify a vector with the coefficients of the linear transformation that represents it in a specific basis. This process represents a tradeoff between generality and computability: the coefficients are different for each basis, but having them enables to remove the abstraction of a Vector Space and just work with $\mathbb{F}^n$: either $\mathbb{C}^n$ or $\mathbb{R}^n$.

To get a matrix form of a linear transformation $T : V \to U$ we:

1. Fix a basis for V: $\{v_j\}_{i=j}^n$ and a basis for U: $\{u_i\}_{i=1}^m$
2. Write the image of the basis of V into the basis of U
$$T(v_j) = \sum_{i=1}^{m} \alpha_{i,j} u_i$$
3. Now for any $v = \sum_{j=1}^n \beta_j v_j \in V$ we have that:
$$\begin{aligned} T(v) &= \sum_{j}^{n} \beta_j T(v_j) \\ &= \sum_{j}^{n} \beta_j \sum_{i=1}^{m} \alpha_{i,j} u_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_{i,j} \beta_j \right) u_i \\ &= \sum_{i=1}^{m} (Ab)_i u_i \end{aligned}$$
4. Construct $A \in \mathbb{F}^{m \times n}$ such that $(A)_{i,j} = \alpha_{i,j} =$ coefficient of $T(v_j)$ w.r.t. u_i

So from now on, we will fix $V = \mathbb{R}^n$ and the canonical basis $(e_i)_j = \delta_{i,j}$. But keep in mind that all of the following apply for **any** finite dimensional vector space and a given basis.

So if we recall from Chapter 1, we can see matrix multiplication as: Let $A \in \mathbb{F}^{m \times n}$ and $b \in \mathbb{F}^n$
$$Ab = \begin{bmatrix} | & | & | \\ a_1 & \dots & a_n \\ | & | & | \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \beta_1 a_1 + \dots + \beta_n a_n$$
It is simply a linear combination of the set $\{a_i\}_{i=1}^n$, and therefore given $T : \mathbb{F}^n \to \mathbb{F}^m$ such that $T(b) = Ab$, then $\text{span}(\{a_i\}_{i=1}^n) = \mathcal{R}(T)$, the column space of A spans the range of T.

4.2 Some useful properties

4.2.1 Symmetry

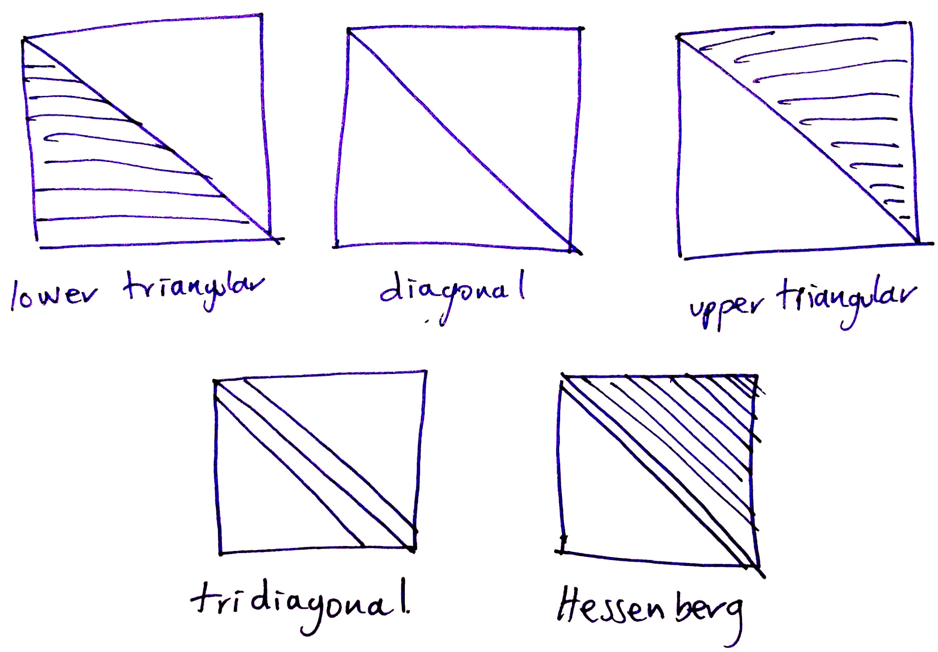
Given a matrix $A \in \mathbb{F}^{m \times n}$, we define:

- the ***transpose*** $A^\top \in \mathbb{F}^{m \times n}$ such that $(A^\top)_{i,j} = a_{j,i}$

- the ***conjugate transpose*** $A^* \in \mathbb{F}^{m \times n}$ such that $(A^\top)_{i,j} = \overline{a_{j,i}}$

- a matrix is ***symmetric*** if $A = A^\top$
- a matrix is ***skew-symmetric*** if $A = -A^\top$
- a matrix is ***hermitian*** or self-adjoint if $A = A^*$
- a matrix is ***skew-hermitian*** or self-adjoint if $A = -A^*$

4.2.2 Shape



- A ***Permutation*** matrix has for each row and column exactly one 1, and the rest of the entries are zero. The product PA permutes the rows of A and AP the columns of A.

4.2.3 Coming from a transformation

- ***Identity*** matrix $I=$ diagonal matrix with ones in the diagonal representing $T(v) = v$

- ***Zero*** matrix $0 =$ all zeros matrix representing $T(v) = 0$

- ***Unitary*** matrix Q if $Q^*Q = I$, it implies that its columns vectors form an orthonormal basis for $\mathcal{R}(A)$ (in the real they are called ***Orthogonal*** matrix).

  - Isometries: The standard inner product is invariant under unitary matrices $< Qu, Qv >= u^*Q^*Qv =< u, v >$
  - For example:***Rotation*** matrix in $\mathbb{R}^{2\times2}$:

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

- ***Projection*** matrix P such that $P^2 = P$ (i.e. $P(P(x)) = P(x)$) If P is hermitian then it is called ***Orthogonal Projection***[2]

- ***Singular*** matrix if A is square and $\mathcal{N}(A) \neq \{0\}$. Otherwise, the matrix is called ***invertible***, i.e. $\exists A^{-1} : A^{-1}A = AA^{-1} = I$

4.3 Fundamental Theorem of Linear Algebra

The four fundamental subspaces of a matrix $A \in \mathbb{F}^{m\times n}$ are :

$$\mathcal{R}(A) \subset \mathbb{F}^m \ , \ \mathcal{N}(A) \subset \mathbb{F}^n \ , \ \mathcal{R}(A^*) \subset \mathbb{F}^n \ , \ \mathcal{N}(A^*) \subset \mathbb{F}^m$$

The ***fundamental theorem of linear algebra*** states that they satisfy[3]:

$$\mathcal{R}(A^*) \perp \mathcal{N}(A) \qquad \text{and} \qquad \mathcal{R}(A^*) \cup \mathcal{N}(A) = \mathbb{F}^n \tag{4.3}$$

$$\mathcal{R}(A) \perp \mathcal{N}(A^*) \qquad \text{and} \qquad \mathcal{R}(A) \cup \mathcal{N}(A^*) = \mathbb{F}^m \tag{4.4}$$

4.4 Operator Norms

In Chapter 2, we saw as an example the Frobenius norm, that identifies a matrix as a vector in $\mathbb{F}^{nm}$. Nevertheless, we can have a different approach if we consider the linear transformation underneath.

---

[2]This does not mean P is an orthogonal matrix!

[3]given $A, B \subset V$ we say $A \perp B$ if $\forall a \in A,\ b \in B :\ \langle a, b\rangle = 0$

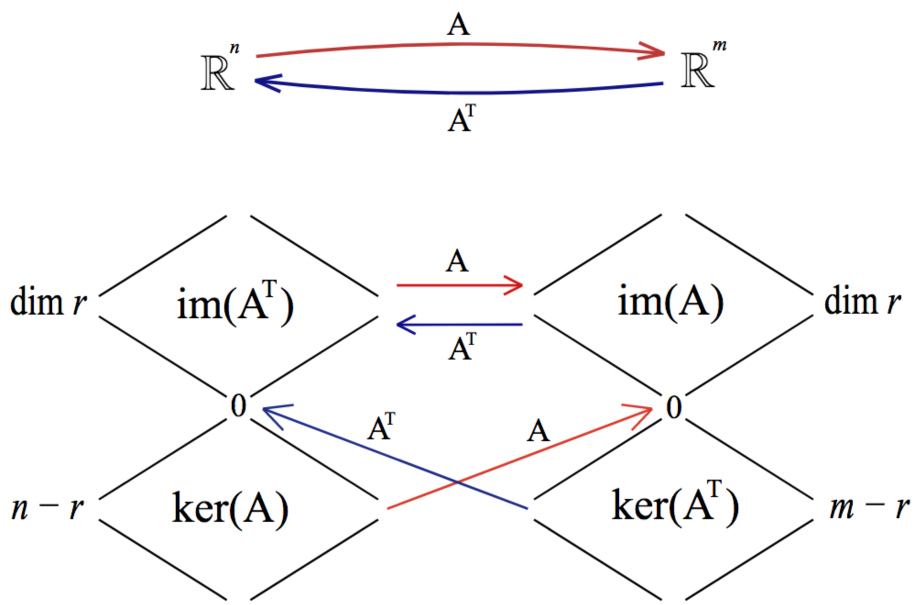


Figure 4.1: The four subspaces" by Cronholm144 - Own work. Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons.

Let V, W finite dimensional normed spaces, then the operator norm for $T : V \to W$ is defined as:

$$||T|| = \sup_{v \neq 0} \frac{||T(v)||_W}{||v||_V} = \sup_{||v||_V = 1} ||T(v)||_W \tag{4.5}$$

We can extend this notion to matrices $A \in \mathbb{F}^{m \times n}$ for any norm in $\mathbb{F}^m$ and $\mathbb{F}^n$

$$||A||_{(m,n)} = \sup_{v \neq 0} \frac{||Av||_{(m)}}{||v||_{(n)}} = \sup_{||v||_{(n)} = 1} ||Av||_{(m)} \tag{4.6}$$

For example for $A \in \mathbb{F}^{m \times n}$:

$$||A||_1 = \sup_{||v||_a = 1} ||Av||_1 = \max_{j=1,\dots,n} \sum_{i=1}^{m} |a_{ij}| = \max_{j=1,\dots,n} ||a_j||_1$$
$$||A||_\infty = \sup_{||v||_a = 1} ||Av||_\infty = \max_{i=1,\dots,m} \sum_{j=1}^{n} |a_{ij}| = \max_{i=1,\dots,m} ||a_i^\top||_1$$

All the induced norms (operators norms) are ***sub multiplicative***:

$$||AB|| \leq ||A|| \, ||B||$$

the Frobenius norm is also submultiplicative. On the other hand $||\cdot||_f$ and $||\cdot||_2$ are invariant under orthogonal transformations: $||QA|| = ||A||$

4.5 Eigenvalues: Matrix as Endomorphism

The invariants of a linear transformation $T : V \to V$ reveal the behavior of T^n as $n \to \infty$. This is important for example to find resonance modes of buildings under an earthquake.

Given $A \in \mathbb{C}^{n \times n}$ square matrix we say that λ is an ***eigenvalue*** of A if $\exists\ v \neq 0$ (***eigenvector***):

$$Av = \lambda v \tag{4.7}$$

Each eigenvalue has a related ***eigenspace***: $E_\lambda = \mathcal{N}(A - \lambda I)$.
The set of all eigenvalues is called ***spectrum***: $\sigma(A) = \{\lambda | \lambda I - A \text{ is singular}\}$
The magnitude of the largest eigenvalue is called ***spectral radius*** $\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|$

Let us define the ***resolvent*** of a matrix $A \in \mathbb{C}^{n \times n}$ the map $\text{res} : \mathbb{C} \to \mathbb{C}^{n \times n}$ such that

$$\text{res}(\alpha) = (\alpha I - A)^{-1}$$

Notice that the poles of the resolvent are exactly the eigenvalues of the matrix A

It is easy to see that A is singular iff 0 is an eigenvalue of A. Also notice that for a triangular matrix, the eigenvalues are the values in the diagonal.
On the other hand notice that if A is hermitian then [4]:

- All the eigenvalues are real
- All the eigenspaces are orthogonal

4.5.1 Useful Theorems

Theorem 1 (Neumann Series)**.** *Let $||A||$ be a square matrix with $||A|| < 1$. Then the matrix $I - A$ is non-singular , moreover*

$$(I - A)^{-1} = \sum_{j=0}^{\infty} A^j$$
$$||(I - A)^{-1}|| \leq \frac{1}{1 - ||A||}$$

This implies that for any matrix A and $\beta \in \mathbb{C}$ such that $|z| > ||A||$ the resolvent is finite:

$$||res(\beta)|| = ||(\beta I - A)^{-1}|| \leq \frac{1}{|\beta| - ||A||} < \infty$$

[4] Check the product $\lambda x^* y$ where $x \in E_\lambda$ and y is an eigenvector

Therefore $\rho(A) \le ||A||$

Theorem 2 (Gershgorin's Disc)**.** *The i-th Gershogorin disc is a ball of radius* $r_i = \sum_{j\ne i} |a_{ij}|$ *centered at* a_{ii} *in the complex plane:*

$$\mathcal{D}_i = \left\{ z \in \mathcal{C} \,\middle|\, |z - a_{ii}| \le \sum_{j\ne i} |a_{ij}| \right\}$$

Every eigenvalue of A sits in a Gershgorin disc

4.5.2 Useful Relations

Two matrices A, B are ***similar*** if there exists X non singular such that $B = X^{-1}AX$. It is easy to see that this implies that A, B represent the same linear transformation under a change of basis X. Therefore they have the **same eigenvalues**, and if v is an eigenvector of A with λ then X^{-1} is an eigenvector of B with λ

On the other hand we say that A, B are ***congruent*** if there exists X non singular such that $B = X^\top AX$. Similarly they are ***Hermitian congruent*** if $B = X^*AX$.

Given a **hermitian** matrix, we define ***inertia*** of a matrix as the triple (n_+, n_-, n_0) equals to number of positive , negative and zero eigenvalues. The ***Sylvester's law of inertia*** says that the inertia of an hermitian matrix is preserved under congruence. For example

$$\begin{bmatrix} -1 & -1 & 8 \\ -1 & -1 & 8 \\ 8 & 8 & -10 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & 2/\sqrt{6} & 0 \end{bmatrix} \begin{bmatrix} 6 & & \\ & -18 & \\ & & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & 2/\sqrt{6} & 0 \end{bmatrix}^\top$$

has inertia $(1, 1, 0)$. Now

$$\begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} -1 & -1 & 8 \\ -1 & -1 & 8 \\ 8 & 8 & -10 \end{bmatrix} \begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 4 \end{bmatrix}^\top$$

$$= \begin{bmatrix} -16 & 28 & 124 \\ 28 & 5 & -1 \\ 124 & -1 & -97 \end{bmatrix}$$

$$= \begin{bmatrix} 0.591 & -0.179 & 0.787 \\ -0.090 & 0.955 & 0.285 \\ -0.802 & -0.239 & 0.548 \end{bmatrix} \begin{bmatrix} -188.5 & & \\ & 0 & \\ & & 88.5 \end{bmatrix} \begin{bmatrix} 0.591 & -0.179 & 0.787 \\ -0.090 & 0.955 & 0.285 \\ -0.802 & -0.239 & 0.548 \end{bmatrix}^\top$$

That has the same inertia.

4.5.3 Useful Invariants

We define ***trace*** of a square matrix $A \in \mathbb{F}^{n \times n}$ as

$$tr(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$$

Properties

- $tr(A + B) = tr(A) + tr(B)$
- $tr(cA) = ctr(A)$
- $tr(ABC) = tr(BCA) = tr(CAB)$ Therefore $tr(X^{-1}AX) = tr(A)$

We define the ***determinant*** of a square matrix $A \in \mathbb{F}^{n \times n}$ as

$$det(A) = |A| = \sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} = \prod_{i=1}^{n} \lambda_i$$

Properties

- $det(I_n) = 1$
- $det(A^\top) = det(A)$
- $det(A^{-1}) = \frac{1}{det(A)}$
- A is singular iff $det(A) = 0$
- A triangular then $det(A) = \prod_{i=1}^{n} a_{ii}$

4.5.4 Algebraic v.s. Geometric multiplicity

Given that A is singular iff $det(A) = 0$, we can find all eigenvalues if we solve the equation $det(\lambda I - A) = 0$. Let the ***Characteristic Polynomial*** be

$$p_A(x) = det(xI - A)$$

For the fundamental theorem of algebra, (if we work in $\mathbb{C}$) the characteristic polynomial has n complex roots that correspond to the eigenvalues of A:

$$p_A(x) = \prod_{i=1}^{n} (x - \lambda_i) = \prod_{k=1}^{|\sigma|} (x - \lambda_k)^{m_k}$$

Notice that $p_A(A) = 0$, and the coefficient of x^0 is the determinant and the coefficient of x^{n-1} is the trace.

For each different eigenvalue λ_k, we define m_k as the ***algebraic multiplicity*** of λ_k. On the other hand the ***geometric multiplicity*** of λ_k is $\dim(E_{\lambda_k}) \geq 1$ (dimension of eigenspace). Notice that the algebraic multiplicity $\geq$ geometric multiplicity. If they are not equal, we say the matrix A is ***defective***, otherwise A is ***diagonalizable***.

For example

$$A = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$$

Has $\lambda = 0$ as eigenvalue, with geometric multiplicity $= 1$ ($E_0 = span([1,0])$) and algebraic =2. Then A is defective.

If A is diagonalizable, there exists an invertible matrix X and a diagonal matrix D such that:

$$A = XDX^{-1}$$

such as D contains the eigenvalues of A and the corresponding columns of X, form a basis for the eigenspace of each eigenvalue. Notice that this decomposition makes it easier to compute A^k and therefore polynomials, power series and therefore analytic functions over matrices.

We say that a matrix is ***Normal*** if $AA^* = A^*A$. The ***spectral theorem*** says that a matrix is normal iff it is ***unitary diagonalizable*** i.e. $A = QDQ^T$ for Q unitary. Notice that in particular, hermitian matrices are unitary diagonalizable. In particular this means that the eigenspace of different eigenvalues are orthogonal.

4.5.5 Useful decompositions

For **any** matrix $A \in \mathbb{C}^{n \times n}$ with eigenvalues $\lambda_1, ...\lambda_n$:

- ***Schur decomposition*** $\exists$ T triangular matrix and Q unitary matrix such that $A = QTQ^*$ (Embedded sequence of invariant orthogonal subspaces). Since A and T are similar, then the diagonal of T contains the eigenvalues of A If A is normal, then it is the sames as the unitary diagonalization. .

- ***Jordan form*** $\exists$ X non singular and J block diagonal matrix with entries J_i such that $A = XJX^{-1}$ where:

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

  The number of Jordan blocks of the same eigenvalue is the geometric multiplicity. The sum of sizes of all the Jordan blocks of the same eigenvalue is the algebraic multiplicity. Then for A diagonalizable, it is the same as the diagonalization.

Note: Notice that there is not a explicit formula to find the roots of polynomials of degree greater than 4. Therefore, computing any of this decompositions is extremely expensive, (and in many cases it can be inaccurate.)

4.6 Singular Values: matrix as Morphism

Most of the times, we consider transformations between **different** vector spaces. This means that we have different bases in each of them. For this cases, computing eigenvalues does not make sense. Therefore we introduce **singular values**

Given a matrix $A \in \mathbb{F}^{m\times n}$ with $rank(A) = r$, there exists $U \in \mathbb{F}^{m\times r}$, $V \in \mathbb{F}^{r\times n}$ unitary and $\Sigma \in \mathbb{R}^{r\times r}$ such that

$$A = U\Sigma V^* = \sum_{i=1}^{r} \sigma_i u_i v_i^*$$

such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_i > 0$. This entries are called ***singular values***. We define the first one $\sigma_1 = ||A||_2$, to know it exists, use induction in this fact (Check Golub for details). We define $\sigma_k = 0$ for $r < k \leq n$. The columns of U are the left singular vectors and the columns of V are the right singular vectors.

It is easy to see that the columns vectors of U are the eigenvectors of AA^* and the columns of V are the eigenvectors of A^*A. The singular values are the square root of the non-zero eigenvalues of AA^* which are the same as A^*A

Properties:

- $span(\{u\}_i) = range(A)$, $span(\{v\}_i) = range(A^*)$
- $||A||_2 = \sigma_1$, $||A^{-1}||_2 = 1/\sigma_n$, $||A||_F = (\sum_{i=1}^{r} \sigma_i^2)^{1/2}$
- We can define the pseudoinverse as $A^{\perp} = V\sigma^{-1}U*$
- We can obtain a k-rank approximations of A choosing the largest k singular values and the minimizes error $||A - A_k||_2 = \sigma_{k+1}$.

4.6.1 Geometric Interpretation.

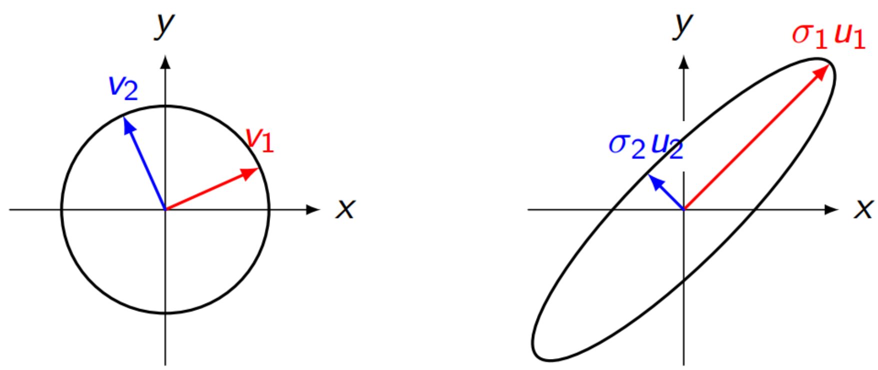


Figure 4.2: Courtesy of Victor Minden

If we look A as a linear transformation for V to U, it is clear that $A = U\Sigma V^*$ defines that $Av_i = \sigma_i u_i$. Therefore to understand the transformation, we can look at the n-th unitary ball in V, identify the vectors v_i inside it, and then map each of them to $\sigma_i u_i$ in U.

4.7 Symmetric Positive definite matrices

If $\mathbb{F} = \mathbb{R}$ We say that a matrix is ***symmetric positive definite*** if $\forall x \neq 0$ $x^\top Ax > 0$. Similarly if $\mathbb{F} = \mathbb{C}$ we say a matrix is ***hermitian positive definite*** if $\forall x \neq 0$ $x^*Ax > 0$. This is equivalent to require that all the eigenvectors of A are positive. Notice that for this case, since A is hermitian, then it is unitary diagonalizable, and since the eigenvalues are positive, then it is clear that this is the only case when eigenvalues and singular values are the same.
We say that it is ***semidefinite*** if we require non-negativity instead of positivity ($x^*Ax \geq 0$)

Properties:

- For any matrix A, $AA*$ and $A*A$ are hermitian positive semidefinite(PSD) $A \succ 0$. If A is full column rank then A^*A is hermitian positive definite (HPD)
- For A HPD, then $\langle x, y\rangle_A = x^*Ay$ defines an inner product.
- For A HPD, there exists $A^{1/2}$ matrix such that $A = A^{1/2}A^{1/2}$
- if A HPD, then $det(A) > 0$, it is non-singular

The related inner product has a different orthogonal basis and measurements than the canonical one.

Chapter 5

Matrices as Linear Systems of Equations

In engineering problems, we are interested in solving linear systems of equations of the form:

$$\begin{aligned}\alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n &= y_1 \\ \ldots &= \ldots \\ \alpha_{m1}x_1 + \alpha_{m2}x_2 + \cdots + \alpha_{mn}x_n &= y_m\end{aligned}$$

This is equivalent to find a vector $x \in \mathbb{F}^n$ such that $Ax = y$, where $(A)_{i,j} = \alpha_{i,j}$ Therefore in order to solve the system, we just need to find the inverse of the linear transformation related with A. But this process is not straightforward.

5.1 Non-singular matrices

- If A is diagonal, it is non-singular only if all its diagonal entries are non-zero, and A^{-1} is diagonal such that $(A^{-1})_{ii} = 1/a_{ii}$
- If A is unitary and square, by definition we know that $A^* = A^{-1}$
- If A is upper triangular we can use ***Backward substitution*** i.e. we can solve the value of the unknown x_i in terms of the following ones:

  $$x_i = \frac{1}{a_{ii}}\left(y_i - \sum_{j=i+1}^{n} a_{ij}x_j\right)$$

  Notice that during this process, it was not necessary to find A^{-1}, moreover we would need to repeat the same procedure for $y_i = e_i$ (n-times more). If A is lower triangular then we can use ***Forward substitution***. Notice that in both cases, we need the elements of the diagonal to be non-zero to have an invertible matrix

- If A satisfy none of the previous properties, then the approach is to find a factorization into known matrix types:
  - ***Gauss Elimination/LU factorization*** Find L lower triangular and U upper triangular such that $LU = A$, then solve using forward and backward substitution. To get U, transform A using linear combinations of its rows. To get L, we keep track of the linear transformations in matrix form:

$$M_i \begin{bmatrix} a_{1j} \\ \vdots \\ a_{jj} \\ a_{j+1,j} \\ \vdots \\ a_{m,j} \end{bmatrix} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{jj} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ where } M_j = I - we_j^\top = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & -\frac{a_{j+1,j}}{a_{j,j}} & 1 & & \\ & & \vdots & & \ddots & \\ & & -\frac{a_{m,j}}{a_{j,j}} & & & 1 \end{bmatrix}$$

    Notice that $M_{n-1}...M_2M_1A = U$, and M_i has nice properties
    * If $M_i = I - we_j^\top$ defined as before, then $M_i^{-1} = I + we_j^\top$
    * If $i \leq j$ then if $M_i = I - we_j^\top$ and $M_j = I - \tilde{w}e_j^\top$ then

$$M_iM_j = I - we_i^\top - \tilde{w}e_j^\top$$

    Then we can show that $L = M_1^{-1} \dots M_{n-1}^{-1}$

    Sadly, this decompositon not always this exists. For example:

$$\begin{bmatrix} 0 & 2 \\ 3 & 1 \end{bmatrix}$$

    Therefore we need to consider permutation of rows and columns. Then for any $A \in \mathbb{C}^{n\times n}$, there exists L lower triangular, U upper triangular, and P permutation matrix, such that:

$$PA = LU$$

    Notice that if A is sparse (for example for finite elements), with this factorization we can obtain L, U sparse too (huge memory and computation savings).

    If A is **HPD**, then there exists L lower triangular with real entries such that $A = LL^*$, and this is called ***Cholesky Factorization***
  - ***Gram-Schmidt/Householder/Givens/QR factorization*** Find Q unitary and R upper triangular such that $A = QR$. Therefore to solve $Ax = b$ is equivalent to solve using backward substitution $Rx = Q^*b$.
    This factorization is not unique, therefore there are 2 approaches

* Construct R applying orthogonal transformations (Householder/Givens). Q is given by keeping track of the transformations.
* Construct Q orthogonalizing the columns space of A (Gram-Schmidt) R is given by the projection of the columns vectors in the new orthonormal basis.

Although not necessarily you can preserve sparsity of the linear system, you have a better control in the norm of the error $||Ax - b||_2 = ||Rx - Q^*b||_2$

Notice that up to know, there is an explicit algorithm to compute the factorization. We can also use either the SVD or the Schur decomposition if we know it. The problem with those is that we can not compute them exactly in polynomial time.

5.2 General Matrices

For a general system $Ax = b$ for $A \in \mathbb{F}^{n \times m}$ we can have the following options:

- $b \notin \mathcal{R}(A)$ then there is no solution. Otherwise there is a solution. When there is no solution, we can choose x such that minimizes the error $||Ax - b||_2$ i.e. **Least squares solution**. In this case, these function is differentiable and the solution is given by the normal equations

$$A^*Ax = A^*b$$

  We can use other norms too, but they are not differentiable, so we need to use linear programing.

- If $\mathcal{N}(A) = \{0\}$ the solution is unique. And just use one of the previous factorizations.

- If none of the 2 conditions are satisfied, then we have infinite number of solutions. Given a solution v, then $v + \mathcal{N}(A)$ is also a solution. A typical choice is to take x with minimal norm $||x||_2$ such that $Ax = b$. The solution is given by $x_{mn} = A^*(AA^*)^{-1}b$

Notice that A^*A in the normal equations is symmetric positive definite (if A was full column rank) and therefore we can use Cholesky factorization. Nevertheless in practice we prefer to use QR factorization to solve least squares, using that the 2-norm is invariant under orthogonal transformations.

If A is not full column rank, QR factorization could fail. then we can use SVD factorization to solve the problem.