Lecture Notes for Linear Algebra Two

Yilong Yang

Updated on May 10, 2019

Contents

Some Words of Warning:

1. As the terminology of various textbooks and of the class may differ in language, it is HIGHLY advised that you preview ahead of each class. Every class I shall outline the sections of the textbook that you should read.

2. I would ESPECIALLY encourage you to come to my office hours or write me emails for questions. As a class taught NOT in your mother language, you are BOUND to misunderstand and confuse things. I had plenty of first hand experience myself. Above all, ask questions, do problems, and be honest with yourself.

3. When viewing or reviewing these notes, you should try to keep some pens and scratch paper ready. You should try to draw stuff or do minor calculations on the side as the content progress.

0.1 Prerequisite

You should have mastered the following materials and skills:

1. Linear Combination and Linear Dependency.

2. Gausssian Elimination.

3. Row and Column Operations to Find Determinants.

4. Matrix Inversion and Multiplication.

5. The Fundamental Theorem of Linear Algebra (Rank-nullity theorem and orthogonality between the four fundamental subspaces of a matrix.)

6. Gram-Schmidt Orthogonalization and QR-Decomposition.

7. Eigenvalues and Eigenvectors.

8. Criteria for Diagonalizability.

9. Spectral Theorem for Real Symmetric Matrices.

The following are also taught in my Linear Algebra I class. They are not required for this class, but you might know them or want to learn about them anyway.

1. LU decomposition.

2. Meaning of row and column operations. (Interpeting matrices as parallelotopes, system of linear equations, linear maps or as completing the square for quadratic forms.)

3. Change of Basis and orthogonal change of basis.

4. Abstract Vector Space and Abstract Inner Product Space (do things without a basis or matrix representation).

5. Projections and Orthogonal Projections.

6. Spectral Theorem for Normal Matrices.

7. Singular Value Decomposition and Applications.

8. Applications to Linear system of ODE.

9. Applications of Linear Algebra in Complex Numbers, Graph Theory, Dynamics and CT scans.

In the stuff listed above, I want to specifically stress that, even though we do NOT need singular value decomposition in this class, it is probably HIGHLY IMPORTANT that you know it. It has tons of applications and will likely show up in your future. I have already taught it in my Linear Algebra I class, while the Chinese Linear Algebra class usually do it in Linear Algebra II.

If you do not know about singular value decompositions, here are my advice: you may come to my office hour for a quick walk-through, or you can read Gilbert Strang's introduction to linear algebra, chapter seven. (There are also accompanying online videos from MIT open course if you like.)

0.2 Class materials and Textbooks

I plan to cover the following materials with the following textbooks. The textbooks are the following: Introduction to Linear Algebra by Gilbert Strang (GS), Linear Algebra Done Wrong by Sergei Treil (LADW), Functions of Matrices: Theory and Computation by Nicholas J Higham (NH), Projective Geometry: An Introduction by Rey Casse (RC), Chinese Textbook by Zhengguang Yu etc. (CT).

1. Complex Matrices (GS Ch 9 and Lecture notes, CT ch 10)
2. Jordan Normal Form (LADW Ch 9, GS Ch 8, CT ch 9)
3. Matrix Analysis (NH ch 1, CT ch 11)
4. Dual and Tensor (LADW Ch 8 and Lecture notes)
5. (Optional) Linear Algebra over Finite Field (Lecture notes)
6. (Optional) Projective Geometry (RC ch 2-4)
7. (Optional) Hyperbolic Geometry (Lecture notes)
8. (Optional) Functional Analysis (Lecture notes)

We will NOT be able to cover all the optional topics. We will probably only have time for one or two optional topics here.

We will NOT cover the following materials:

1. Polynomial Theory (CT ch 8). We simply don't need this in our class, and it is largely disconnected with everything else in this class.
2. Singular Value Decompositon (CT ch 10.2). I have already done this in my Linear Algebra I class.

0.3 Grading

We will have a research project (10%), weekly homework (30%), takehome midterm (30%) and an open book final (30%).

For the research project, you are asked to read Chapter 2 of NH and do all 7 problems of that chapter. Alternatively, you can pick one of the following topics and write a introduction to these things:

1. Stress tensor, Strain tensor or Elasticity tensor field in continuum mechanics.
2. Electromagnetic tensor.
3. Riemann curvature tensor used in general relativity.
4. Other topics related to class. (Ask me to be sure that your topic is viable.)

If you choose to write about a topic, you need to explain the concepts clearly, give examples, and hopefully play with these concepts a littlbe bit in some original manner. As long as you work on it in good faith, you will be graded generously.

1 Complex Structure

In case you have not seen these before, here are the standard definition of a vector space.

Definition 1.1. A field $(F, +, \times)$ is a set F together with two operations called addition $+ : F \times F \to F$ and multipliction $\times : F \times F \to F$, that obeys the following rules:

1. Addition $+$ is commutative and associative, has identity $0 \in F$ and each element $a \in F$ has an additive inverse $-a \in F$.
2. Multiplication $\times$ is commutative and associative, has identity $1 \in F$, and each NON-ZERO element $a \in F$ has a multiplicative inverse $a^{-1} \in F$.
3. Multiplication is distributive over addition, i.e., $a(b+c) = ab+ac$. You may also say that multiplication respect addition.

Definition 1.2. A vector space $(V, +, \times)$ over a field F is a set V with two operations called vector addition $+ : V \times V \to V$ and scalar multiplication $\times : F \times V \to V$, that obeys the following rules:

1. Vector addition $+$ is commutative and associative, has identity, i.e., the zero vector $0 \in V$, and each element $v \in V$ has an additive inverse $-v \in V$.
2. Scalar multiplication $\times$ respect the field operations and vector addition. I.e., for any $a, b \in F$ and $v, w \in V$, we have $1v = v$, $0v = 0$, $(-1)v = -v$, $(ab)v = a(bv)$ and $(a+b)v = av + bv$ and $a(v+w) = av + aw$.

For exotic examples of weird vector fields and how to think of them, see my lecture notes for Linear Algebra One. See example 16.1 and 16.2.

Recall the central philosophy here: matrices = linear maps under some specific basis.

1.1 Two views of the complex numbers

Example 1.3. *There is a mysterious connection between the complex number $a+bi$ and the matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. For example, consider $(1+2i)(3+4i) = -5+10i$. Amazingly, you can immediately check that $\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 3 & -4 \\ 4 & 3 \end{pmatrix} = \begin{pmatrix} -5 & -10 \\ 10 & -5 \end{pmatrix}$.*

What is a complex number? A cheap view is the following: it is nothing more than $\mathbb{R}^2$ where we call $\begin{pmatrix} a \\ b \end{pmatrix}$ as $a+bi$, and we define an exotic weird product structure, i.e., $(a+bi)(c+di) = ac - bd + (ad+bc)i$. You can verify that this can be simply computed by remembering $i^2 = -1$ and use the law of distributivity. You can verify that complex multiplication is associative, commutative, and distributive. More formally, we have

Definition 1.4. The complex numbers $\mathbb{C}$ is the two-dimenisonal real vector space $\mathbb{R}^2$, where addition in $\mathbb{C}$ is the vector addition in $\mathbb{R}^2$, and multiplication in $\mathbb{C}$ is defined as $\begin{pmatrix} a \\ b \end{pmatrix} \times \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} ac - bd \\ ad + bc \end{pmatrix}$. We call the vector $\begin{pmatrix} a \\ b \end{pmatrix}$ as the complex number $a + bi$.

Now we fix $z = a + bi$. For each complex number $v = \begin{pmatrix} c \\ d \end{pmatrix}$, the map m_z going from v to zv is linear! (We call this linear map "multiplication by z".) Indeed, $z(v+w) = zv + zw$ and for any real number k, we have $z(kv) = k(zv)$! So what is the matrix of this linear map? Well, $zv = \begin{pmatrix} ac - bd \\ ad + bc \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix}$. So $m_z = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ as a linear map.

So if you ever want to study a matrix like $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, you are essentially studying the process of multiplying by $a + bi$, i.e., the map m_{a+bi}.

But then for any two complex numbers z and w, you can easily see that $m_z m_w = m_{zw}$. Since composition of linear maps corresponds to multiplication of the corresponding matrices, we see that the matrix for m_{zw} is exactly the multiplication of the matrices m_z and m_w. In particular, if $(a + bi)(c + di) = e + fi$, then we immediately have $\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} e & -f \\ f & e \end{pmatrix}$. You can also see that addition is also immediatly satisfied.

Definition 1.5. The complex number $\mathbb{C}$ is the two-dimensional subspace of 2 by 2 real matrices M_2 of the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Addition in $\mathbb{C}$ is the same as addition of matrices, and multiplication of $\mathbb{C}$ is the same as multiplication of matrices.

In mathematics, we don't really care about specific objects. We only care about how objects interact. So when we say complex numbers, we do NOT mean that there are some mysterious objects somewhere in the universe called complex numbers. What we really interested in is the way these complex numbers add and the way they multiply, and such interactions is what we mean by complex numbers.

Recall the following definition of a field.

Definition 1.6. A field $(F, +, \times)$ is a set F together with two operations called addition $+ : F \times F \to F$ and multipliction $\times : F \times F \to F$, that obeys the following rules:

1. Addition $+$ is commutative and associative, has identity $0 \in F$ and each element $a \in F$ has an additive inverse $-a \in F$.
2. Multiplication $\times$ is commutative and associative, has identity $1 \in F$, and each NON-ZERO element $a \in F$ has a multiplicative inverse $a^{-1} \in F$.
3. Multiplication is distributive over addition, i.e., $a(b+c) = ab+ac$. You may also say that multiplication respect addition.

Proposition 1.7. *The two definitions of complex numbers above gives you the same field, i.e., the field of complex numbers, denoted as* $\mathbb{C}$.

1.2 Complex Structure

We now study complex vector spaces. First recall the following definition:

Definition 1.8. A vector space $(V, +, \times)$ over a field F is a set V with two operations called vector addition $+ : V \times V \to V$ and scalar multiplication $\times : F \times V \to V$, that obeys the following rules:

1. Vector addition $+$ is commutative and associative, has identity, i.e., the zero vector $0 \in V$, and each element $v \in V$ has an additive inverse $-v \in V$.
2. Scalar multiplication $\times$ respect the field operations and vector addition. I.e., for any $a, b \in F$ and $v, w \in V$, we have $1v = v$, $0v = 0$, $(-1)v = -v$, $(ab)v = a(bv)$ and $(a + b)v = av + bv$ and $a(v + w) = av + aw$.

Example 1.9. *An obvious example is* $\mathbb{C}^n$*, a list of* n *complex numbers. This is the standard* n *dimensional complex vector space.*

Now consider the complex numbers $\mathbb{C}$, which is obvious a one dimensional vector space over itself. Multiplying by the imaginary i is a also linear map m_i. So as a linear map, m_i is in fact the matrix $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. This is rotation by 90 degree counter-clockwise. You can also easily check that $J^2 = -I$. This turned out to be a very useful piece of information: Multiplying by i is a linear map whose square is negative identity.

Example 1.10. *Consider the following vector space $\mathbb{R}^2$ with a linear map $J = \begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix}$. It is already a real vector space, so vector addition is fine as it is. I can define a complex scalar multiplication as $(a+bi)v = av + bJv = (aI + bJ)v$. You can easily verify all rules, and see that I have defined a complex vector space structure on $\mathbb{R}^2$ in this manner. In particular, $(aI + bJ)(cI + dJ) = (ac - bd)I + (ad + bc)J$ as expected.*

So you can also verify that for any two vectors v, w, if v is non-zero, say $v = (1,0)^T$, then v and $Jv = (1,-1)^T$ are in fact linearly independent and form a basis. So w can be written as a unique linear combination of v and Jv, for example $w = av + bJv$. Then $w = (a+bi)v$ the two are linearly dependent. So all vectors in $\mathbb{R}^2$ are complex-colinear, this is a one-dimensional complex vector space.

So what is a complex vector space of dimension one? It is simply any $\mathbb{R}^2$ with any linear map J such that $J^2 = -I$. We can interpret this J as "multiplying by i".

Similarly, if V is any 4-dimensional real vector space, we can fix any linear map $J : V \to V$ such that $J^2 = -I$, and then defining $(a+bi)v = (aI + bJ)v$ gives a complex vector space structure on V. V could then be seen as a 2-dimensional complex vector space. The linear map J can be interpreted as "multiplying by i".

To sum up, there is nothing imaginary about complex numbers or complex vector spaces. They are really just real vector spaces equipped with a special linear map J with $J^2 = -I$. Any such J could define a complex structure on the real vector space.

Remark 1.11.

1. *For the vector space $\mathbb{C}$, consider $1, i \in \mathbb{C}$. Now on one hand, 1 and i are clearly linearly dependent over $\mathbb{C}$. On the other hand, they are linearly INDEPENDENT over $\mathbb{R}$. Similarly, in the vector space $\mathbb{C}^2$, $(1,0)^T, (i,0)^T, (0,1)^T, (0,i)^T$ are linearly INDEPENDENT over $\mathbb{R}$, but they are linearly dependent over $\mathbb{C}$.*

2. *You can also see that, for any $J \in M_k(\mathbb{R})$ with $J^2 = -I$, then the 2-dimensional real vector space spanned by I and J would naturally be another model for $\mathbb{C}$ with the usual matrix addition and matrix multiplication as field addition and field multiplication. We now have infinitely many models for $\mathbb{C}$.*

3. *Suppose $J^2 = -I$. Then squares of any eigenvalue of J would be -1! So all eigenvalues of J are i and $-i$. Furthermore, since J is a real matrix, i and $-i$ must appear in pairs.*

4. *You can see that, since J has NO real eigenvalue, it has NO real eigenvector. For any vector v, v and Jv are always linearly independent.*

5. *You can also see that, since the eigenvalues i and $-i$ come in pairs, for any $J \in M_k(\mathbb{R})$ with $J^2 = -I$, you can prove that k must be an even number! So any complex vector space must have even real dimension. Say the vector space is $\mathbb{R}^{2n}$, then any complex structure would turn it into $\mathbb{C}^n$.*

6. *For any real vector space $\mathbb{R}^{2n}$ and a comlpex structure given by J, you can find a complex basis in this way: first find any non-zero vector v_1. Then find a vector v_2 linearly independent from v_1 and Jv_1 over $\mathbb{R}$ (or equivalently, linearly independent from v_1 over $\mathbb{C}$). Then find a vector v_3 linearly independent from v_1, Jv_1, v_2 and Jv_2 over $\mathbb{R}$ (or equivalently, linearly independent from v_1 and v_2 over $\mathbb{C}$). And so on so forth. Then $v_1, ..., v_n$ form a complex basis, while $v_1, ..., v_n, Jv_1, ..., Jv_n$ form a real basis.*

Example 1.12. *(Quoted from the book from one to two to infinity.)*

A treasure is burried on an island. According to a treasure map, to find the treasure, you should start at a secret location X. You walk towards a tree A, and turn left, and walk the same distance as the distance from X to A, to reach a location S. Then you start from X again, walk towards a tree B, and turn right, and walk the same distance as the distance from X to B, to reach a location T. Then the midpoint of S and T is the treasure.

Unfortunately, on the treasure map, the location of X is blurry. What can you do?

Notice that "turn right", or "rotation counter-clockwise by 90 degree", is a linear map J whose square is $-I$, so it can be thought of as multiplying by i! Draw the complex coordinate chart, set $A = -1$ and $B = 1$. Suppose $X = a + bi$. Then the description says $S - A = i(A - X)$ and $T - B = -i(B - X)$. So $S + T = A + B + iA - iB = -2i$. So the treasure is located at $\frac{1}{2}(S + T) = -i$. You can draw the graph for some X to verify the truth of the location. So our treasure location is in fact independent from X.

Example 1.13. *The complex structure mentioned here appear naturally in physics. Suppose we have something of unit mass, moving vertically up and down, affected only by gravity. Its height is a function $h(t)$ with values in $\mathbb{R}$, with positive values pointing up, and its speed is a function $v(t)$ with values in $\mathbb{R}$, again with positive values pointing up. We are interested with the dynamics of the pair $\begin{pmatrix} h(t) \\ v(t) \end{pmatrix}$. This vector lives in $\mathbb{R}^2$.*

Now the total energy of this system is $E = gh + \frac{1}{2}v^2$. You can see that $\frac{\partial E}{\partial h} = g = -\frac{d}{dt}v$, while $\frac{\partial E}{\partial v} = v = \frac{d}{dt}h$. So we have a system $\begin{pmatrix} \frac{dh}{dt} \\ \frac{dv}{dt} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial E}{\partial h} \\ \frac{\partial E}{\partial v} \end{pmatrix}$. But hey, the matrix here can be squared to $-I$! So it is in fact the J for some complex structure. Under such a complex structure, we can in fact write $\frac{dr}{dt} = i(\nabla E)$, where $r(t) = \begin{pmatrix} h(t) \\ v(t) \end{pmatrix}$ is the status of your system, and $\nabla E = \begin{pmatrix} \frac{\partial E}{\partial h} \\ \frac{\partial E}{\partial v} \end{pmatrix}$ is the gradient of the engergy, or the "derivative" of the energy function.

So the evolution of your system is i times the derivatives of the energy. (In this very special case at least.)

1.3 Complex Linear Maps

Definition 1.14. For vector spaces V, W over a field F, a function $f : V \to W$ is F-linear if $f(kv) = kf(v)$ for any $k \in F$ and $f(v + w) = f(v) + f(w)$.

Suppose V, W are both complex vector spaces. Then functions $f : V \to W$ could be real linear, or complex linear, and the two concepts are very different.

Example 1.15. *Consider the following map $f : \mathbb{C} \to \mathbb{C}$ such that $f(a + bi) = a + b$. This is a real linear map, because f clearly respect vector addition, and $f(k(a + bi)) = kf(a + bi)$ for any REAL number k. If you pick 1 and i as a real basis for $\mathbb{C}$, then f can be written as a 2 by 2 real matrix $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$.*

However, f is NOT complex linear map. We have $f(i) = 1$. On the other hand, $f(i \times i) = f(-1) = -1$ while $if(i) = i \times 1 = i$. As you can see, they are different. f CANNOT be written as a 1 by 1 complex matrix.

The easy observation is the following.

Proposition 1.16. *If $L : V \to W$ is complex linear, then it is real linear.*

What about its converse? When is a real linear map complex linear? Here we provide two criteria, one theoretical and one computational.

Proposition 1.17. *For a real linear map $L : V \to W$ over complex vector spaces (V, J_V) and (W, J_W), then L is complex linear iff $LJ_V = J_W L$. (Note that as real linear maps, these compositions can all be computed as matrix multiplications.) (Also note that this can be interpreted as $L(iv) = i(Lv)$ for any $v \in V$.)*

Proof. Recall that by definition of the complex structure, $(a+bi)v = (aI + bJ_V)v$ for any $v \in V$, and $(a+bi)w = (aI + bJ_W)w$ for any $w \in W$.

We always have $L((a+bi)v) = L(aI + bJ_V)v = (aL + bLJ_V)v$ and $(a+bi)Lv = (aI + bJ_W)Lv = (aL + bJ_W L)v$. So the two are the same for all $a, b \in \mathbb{R}$ and $v \in V$ iff $bLJ_V = bJ_W L$ for all $b \in \mathbb{R}$, iff $LJ_V = J_W L$. $\square$

Proposition 1.18. *Fix complex vector spaces V, W with any complex basis $v_1, ..., v_n$ for V, and $w_1, ..., w_m$ for W. Any real linear map $L : V \to W$ is complex linear iff its real $2m$ by $2n$ matrix is made up by 2 by 2 blocks like $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ under the real basis $v_1, iv_1, v_2, iv_2, ..., v_n, iv_n$ for V and the real basis $w_1, iw_1, w_2, iw_2, ..., w_m, iw_m$ for W. (Similarly, any real linear malo $L : V \to W$ is complex linear iff its real $2m$ by $2n$ matrix has block form $\begin{pmatrix} B & -C \\ C & B \end{pmatrix}$ for m by n real matrices B, C under the real basis $v_1, ..., v_n., iv_1, ..., iv_n$ and the real basis $w_1, ..., w_m, iw_1, ..., iw_m$.)*

Proof. First let us prove the backward direction. Suppose the $(2j-1)$-th column of L is $\begin{pmatrix} a_1 \\ a_2 \\ ... \\ a_{2m} \end{pmatrix}$. Then it means that $Lv_j = (a_1 + ia_2)w_1 + (a_3 + ia_4)w_2 + ... + (a_{2m-1} + ia_{2m})w_m$. Then $L(iv_j) = i(a_1 + ia_2)w_1 + i(a_3 + ia_4)w_2 + ... + i(a_{2m-1} + ia_{2m})w_m = (-a_2 + ia_1)w_1 + (-a_4 + ia_3)w_2 + ... + (-a_{2m} + ia_{2m-1})w_m$. So the $(2j)$-th column of L is $\begin{pmatrix} -a_2 \\ a_1 \\ -a_4 \\ a_3 \\ ... \\ -a_{2m} \\ a_{2m-1} \end{pmatrix}$. In particular, the two column combined looks like First let us prove the backward direction. Suppose the $(2j-1)$-th column of L is $\begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \\ ... & \\ a_{2m-1} & -a_{2m} \\ a_{2m} & a_{2m-1} \end{pmatrix}$, which is the block structure we desired.

Now we turn to the easy forward direction. For each block $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, we replace it with the complex number $a + ib$. Then we have a m by n complex matrix describing the same map, so the map is complex linear. $\square$

Example 1.19. *Recall that $\mathbb{R}^2$ with $J = \begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix}$ is a complex vector space, and obviously $2I + J$ itself is a complex linear map since $(2I+J)J = 2J + J^2 = J(2I+J)$. Let us now verify this computationally.*

Take $v = (1,0)^T$ and $Jv = (1,-1)^T$ as a basis. Under this new basis, the linear map $2I + J$ now looks like $\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$, which is the correct block form as predicted. This is also exactly the standard matrix for multiplying by $2+i$, the action achieved by $2I + J$!

Proposition 1.20. *For any complex vector space (V, J), pick any complex basis $v_1, ..., v_n$, then under the real basis $v_1, iv_1, ..., v_n, iv_n$, the linear map J looks like $J_{2n} = diag(J_2, J_2, ..., J_2)$ where $J_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. (Similarly, under the real basis $v_1, ..., v_n, iv_1, ..., iv_n$, then J looks like $\begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$.)*

Proof. As you can imagine, applying J to them amounts to multiplying i to each, so v_1 goes to iv_1, and iv_1 goes to $-v_i$ and so on. So in fact J now looks like $J_{2n} = diag(J_2, J_2, ..., J_2)$ where $J_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. $\square$

Corollary 1.21. *$A \in M_{2n}(\mathbb{R})$ is made up by 2 by 2 blocks like $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ iff $AJA^{-1} = J$, where $J = J_{2n}$ as above.*

So in the future, whenever we want to talk about complex vector spaces, we shall always try to find a nice basis where the complex strucutre looks like the matrix J_{2n}. From now on, we shall simply assume that J_{2n} is our complex structure.

Remark 1.22. *In practice, some people prefer the basis order $v_1, ..., v_n, iv_1, ..., iv_n$. Then J would look like $\begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}$. This has certain advantages sometimes. However, we shall not think of it as such in our class.*

1.4 Realization

So it seems that complex matrices are really real matrices with double the dimension and some special structures. We can define the following concepts

Definition 1.23. For a complex matrix $A \in M_{m\times n}(\mathbb{C})$, its ***Realization*** is $\hat{A} \in M_{2m\times 2n}(\mathbb{R})$ where each entry $a + bi$ of A is replaced by the 2 by 2 block $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$.

As a sidenote, we also mention the following process, which looks trivial at first sight but is secretly very powerful.

Definition 1.24. For a real matrix $A \in M_{m\times n}(\mathbb{R})$, its ***Complexification*** is the same matrix except that we now consider A to be inside $M_{m\times n}(\mathbb{C})$. (Note that A is transformed from a linear map $\mathbb{R}^n \to \mathbb{R}^n$ into a linear map $\mathbb{C}^n \to \mathbb{C}^n$, so it becomes a different linear map!)

Example 1.25. *The realization of the complexification of $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ is $\begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 4 & 0 \\ 0 & 3 & 0 & 4 \end{pmatrix}$.*

Example 1.26. *You can see that the realization of iI is $J_{2n} = diag(J_2, J_2, ..., J_2)$. Note that this represent multiplying all vectors by i, so indeed this should be the underlying complex structure.*

Example 1.27. *What is a complex vector? Recall that we are operating under a basis $v_1, iv_1, ..., v_n, iv_n$, so a complex vector $\begin{pmatrix} a_1 + b_1 i \\ \vdots \\ a_n + b_n i \end{pmatrix}$ has a realization of $\begin{pmatrix} a_1 & -b_1 \\ b_1 & a_1 \\ a_2 & -b_2 \\ b_2 & a_2 \\ \vdots & \vdots \\ a_n & -b_n \\ b_n & a_n \end{pmatrix}$. Say we call the first column u, which basically simply listed the real and imaginary parts. Then the two columns are u and Ju. So if we talk about something happening to a complex vector, it is the same as talking about two real vectors u and Ju simultaneously.*

In short, complex stuff happening to a vector u, is equivalent to the corresponding real stuff happening simultaneously to u and Ju.

Remark 1.28. *The above realization is done by assuming $J = J_{2n}$. However, if you prefer $J = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}$, then the corresponding realization should be $A = \begin{pmatrix} \Re(A) & -\Im(A) \\ \Im(A) & \Re(A) \end{pmatrix}$. Here $\Re(A)$ means we only take the real parts and drop all imaginary parts, and $\Im(A)$ means we take the coefficients of the imaginary part (which are real) and drop all real parts. So $\Re(a+bi) = a$ and $\Im(a+bi) = b$. Then all arguments below are still valid except for some formal adjustment into block forms.*

Proposition 1.29. $\widehat{AB} = \hat{A}\hat{B}$.

Proof. A and $\hat{A}$ are merely different representation of the same complex linear map. Now matrix multiplication means function composition, and both sides of the equations are the same composition of the same two complex linear functions, so of course they give the same result. □

Remark 1.30. *Think of each entry $a+bi$ of a complex matrix NOT as a representation of a complex number, but RATHER as a representation of a 2 by 2 block. Then perform block multilications, and you get the same result. So after a fashion, $a+bi$ and $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ are really interchangable.*

Proposition 1.31. *If A has eigenvalues $\lambda_1, ..., \lambda_n$ counting algebraic multiplicity, then $\hat{A}$ has eigenvalues $\lambda_1, ..., \lambda_n$ and $\overline{\lambda_1}, ..., \overline{\lambda_n}$ counting algebraic multiplicity.*

Corollary 1.32. $\det(\hat{A}) = |\det(A)|^2$ *and* $\operatorname{tr}(\hat{A}) = 2\Re(\operatorname{tr}(A))$.

The most important part about realization is the following weird phenomena: for real matrices, there are many theorems involving transpose. But for complex matrixes, miraculously, all these theorems now require transpose CONJUGATE! (Recall that the conjugate of a complex number $a+bi$ is $a-bi$).

Example 1.33.

1. *For two real vectors, their inner product is usually defined as $\begin{pmatrix} a \\ b \end{pmatrix}^T \begin{pmatrix} c \\ d \end{pmatrix} = ac+bd$.*

   But for two complex vectors, their inner product is usually defined as $\overline{\begin{pmatrix} a \\ b \end{pmatrix}}^T \begin{pmatrix} c \\ d \end{pmatrix} = \overline{a}c + \overline{b}d$.

2. *For a real matrix A, if A is symmetric, i.e., $A = A^T$, then A is really nice and satisfy a nice spectal theorem, with real eigenvalues and orthogonal eigenvectors etc.. But for a complex matrix A, we need A to be Hermitian, i.e., $A = \overline{A^T}$, and THEN A would be nice and satisfy a nice spectral theorem, with real eigenvalues and orthogonal eigenvectors and so on.*

3. *For a real matrix A, if $AA^T = I$, then A is an orthogonal matrix and represent our intuiton for rotations. For a complex matrix, if $A\overline{A^T} = I$, then A is called a unitary matrix and represent our intuition for rotations (over the complex number).*

4. *(Recall that the determinants is defined by the Leibniz formula. You can simply still use the same formula for complex matrices, and have complex determinants. A complex determinant might be a complex number and no longer real.)*

5. *For a real invertible matrix A, we can take its cofactor matrix $Cof(A)$, and we have $\det(A)A^{-1} = Cof(A)^T$. For a complex invertible matrix A, its cofactor matrix is computed as computing those minors and then add conjugate. So technically its cofactor matrix is $Cof_{\mathbb{C}}(A) = \overline{Cof_{\mathbb{R}}(A)}$, where $Cof_{\mathbb{R}}$ denote the formula for taking the cofactor matrix for real matrices. Then $\det(A)A^{-1}$ is the transpose conjugate of this cofactor matrix, which is $\overline{Cof_{\mathbb{C}}(A)}^T = \overline{\overline{Cof_{\mathbb{R}}(A)}}^T = Cof_{\mathbb{R}}(A)^T$.*

Definition 1.34. For a complex matrix A, we define the ***adjoint*** of A to be $\overline{A}^T = \overline{A^T}$.

Example 1.35. *We have $\begin{pmatrix} i & 1+i \\ 2+i & 1-i \end{pmatrix}^* = \begin{pmatrix} -i & 2-i \\ 1-i & 1+i \end{pmatrix}$.*

Proposition 1.36. *For a complex matrix A and its realization $\hat{A}$, the adjoint A^* is realized by $\hat{A}^T$. To taking transpose conjugate for A means taking transpose for its realization.*

Corollary 1.37. *A is Hermitian iff the realization of A is symmetric. A is unitary iff the realization of A is orthogonal, etc..*

Example 1.38. *Recall that $a+bi$ can be realized as $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Taking transpose of this realization, we have $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, which is the realization of the complex number $a-bi$, the conjugate of $a+bi$.*

Here is a really interesting fact with application in physics.

Definition 1.39. Let $J = J_{2n}$.

1. An invertible matrix $A \in M_{2n}(\mathbb{R})$ is called ***realizable*** if $AJA^{-1} = J$. (Which is equivalent to saying that A is the realization of some complex matrix, as per Corollary 1.21.)
2. A matrix $A \in M_{2n}(\mathbb{R})$ is called ***symplectic*** if $AJA^T = J$.
3. A matrix $A \in M_{2n}(\mathbb{R})$ is called ***orthogonal*** if $AA^T = I$.
4. A matrix $A \in M_n(\mathbb{C})$ is called ***unitary*** if $AA^* = I$.

Theorem 1.40 (2 out of 3)**.** *(Realizable + Symplectic) = (Realizable + Orthogonal) = (Symplectic + Orthogonal) = (Realizations of Unitary Matrices). In short, any two of the three implies unitary.*

Proof. If A is unitary, then clearly $\hat{A}$ is realizable. And $A^* = A^{-1}$ means $\hat{A}^T = \hat{A}^{-1}$, so $\hat{A}$ is orthogonal. Conversely, if $\hat{A}$ is an orthogonal matrix and realized by the complex matrix A, then $\hat{A}^* = \hat{A}^T = \hat{A}^{-1} = \widehat{A^{-1}}$, so $A^{-1} = A^*$ and A is unitary. (This portion of the proof is almost just reciting definition.)

Now, any two of the equations $AJA^T = J, AJA^{-1} = J, A^{-1} = A^T$ clearly implies the third one. So we are done. $\square$

In quantum mechanics, the dynamics is usually symplectic (as in the previous example involving Hamilton mechanics.) The probability structure is usually orthogonal. As a result, combining the two shows that quantum mechanics are unitary by nature.

Note that, the results here does not really depend on our choice of realization. One can just abstractly use the complex structure J to establish these results. But in practice, it is always a good idea to find a good basis, and under a good basis, complex matrices will always be "reailzed" as in this section.

1.5 Extra stuff insertion

First here is another application of complex structures.

Example 1.41 (Complex romantic relation)**.** *Suppose $f' = kf$, then I'm sure you know that the solution is $f(x) = e^{kx}f(0)$. That is the prerequisite knowledge of this application.*

Suppose two person A, B are in a romantic relation.Their love for each other is a function of time, say $A(t)$ and $B(t)$. Now A is a normal person. For normal people, the more you are loved, the more you love back. In particular, $A'(t) = B(t)$. However, B is an unappreciative person. If you love B, then B take you for granted, and treat you as garbage. If, however, you treat B badly, then B would all of a sudden thinks of you as super charming and attractive. In short, B enjoys things that are hard to get, and think little of the things that are easy to get. In Chinese, we say B is a Jian Ren. Anyway, we see that $B'(t) = -A(t)$.

Now, consider the real vector $v(t) = \begin{pmatrix} A(t) \\ B(t) \end{pmatrix} \in \mathbb{R}^2$. Then for the matrix $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, we see that $v' = -Jv$. Also note that J here defines a complex structure on $\mathbb{R}^2$. Therefore, using this complex structure,

$\mathbb{R}^2$ is now simply $\mathbb{C}$, and v would be like some complex number. And we have $v' = -iv$. So the solution is $v(t) = e^{-it}v(0) = (\cos(t) - i\sin(t))v(0)$.

Then the solution should be $(\cos(t)I - \sin(t)J)\begin{pmatrix} A(0) \\ B(0) \end{pmatrix} = \begin{pmatrix} A(0)\cos(t) + B(0)\sin(t) \\ B(0)\cos(t) - A(0)\sin(t) \end{pmatrix}$. This is indeed the collection of all possible solutions of our system. We have solved the differential equation.

Note that the romantic relation of A and B are necessarily periodic. If you are ever trapped in a relationship which is periodic, (i.e., happy for a week, then fight for a week, and repeat), then maybe you should think about this model a bit more.

Just a side note. Computations are important, but computations are not math. Math is about the thinking and the planning, not the execution.

In the study of linear algebra, there are two kinds of concepts. In the world of linear algebra, the first class citizens are the concepts that are INDEPENDENT from your choice of basis. And the concepts that may change if you change basis are inferior concepts, because they are mainly used for computations.

For example, linear maps are first class citizens, but the specific entries of their matrix representations are NOT important, because they may change as soon as you change basis. Vectors themselves are first class citizens, while their specific coordinates are not. The linear algebra way of thinking should be the following: There are NO coordinates, NO basis, and NO systems of linear equations. There are only linear maps and vectors and their properties that are independent of any change of basis.

Example 1.42. *So, given a set V, is it a real vector space or a complex vector space? This does not depends on V at all. It only depends on the structure you defined on V. And what structure should you put? That should depend completely on the PROBLEMS at your hand. Say you want to study the function $f : \mathbb{C}^2 \to \mathbb{C}^2$ defined as $f(a + bi, c + di) = a + b + c + d$. Then you should NOT treat $\mathbb{C}^2$ as a complex vector space. Why? Because the function of our interests has NOTHING to do with the complex structure at all. It is NOT complex linear.*

Say you want to study $K : \mathbb{C}^2 \to \mathbb{C}^2$ defined as $K(a+bi, c+di) = ((a-c)+(b+d)i, (a+c)+(d-b)i)$. Then on one hand, this map is NOT complex linear. $K(i, 0) = (i, -i)$ while $iK(1, 0) = (i, i)$. On the other hand, it IS complex linear. Why? For the moment, just consider $\mathbb{C}^2$ as a 4-dim real vector space and NOT a complex vector space. Consider the real liear map J that sends a pair of complex numbers (z, w) to $(-\overline{w}, \overline{z})$. This is also not complex linear. However, you may check that $J^2 = -I$, huh. So in fact J defines a DIFFERENT complex structure on $\mathbb{C}^2$, where $i(z, w)$ is now defined to be $(-\overline{w}, \overline{z})$. And under this NEW complex structure, you see that K is the multiplication of $1 + i$ on vectors.

As you see, the same space, same function, they could be real or complex, it all depends on your perspective, on the structure that YOU choose to put on them. Given a function $f : V \to W$, maybe yuo could write f as a real matrix, maybe you can write f as a complex matrix. But the underlying function is the same, and the sets V, W are the same.

Example 1.43. *Out of old habits, I'm sure most people, when they deal with complex numbers, would as a habit divide complex numbers into real parts and complex parts. Let me tell you a secret: these are NOT important. Why? Because they depend completely on your choice of basis!*

Take any complex vector space V with basis v_1, v_2, v_3. Under this basis, the vector $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, which represent the vector $w = v_1 + v_2 + v_3$, seems to be "real". In particular, the real part is $\Re(w) = w$, while the imaginary part is $\Im(w) = 0$.

Now consider iv_1, v_2, v_3. These are also linearly independent over $\mathbb{C}$, and thus they also form a basis. Under this new basis, the w is now represented as $\begin{pmatrix} -i \\ 2 \\ 3 \end{pmatrix}$, since $w = (-i)(iv_1) + v_2 + v_3$. As you can see, now $\Re(w) = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}$ while $\Im(w) = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$. Huh.

Think about it. Whenever you pick a complex basis, they automatically become "real", because they are now $(1,0,...,0),(0,1,0,...,0)$ *and so on.*

The point is this: you should NOT make a habit of taking real parts and complex parts right away. Pick a basis first, try to understand your basis, and then proceed with your calculations.

Example 1.44. *What about complex conjugates? If you no longer have real parts or imaginary parts, how should you define* $\overline{z}$*? Well, it turned out that there are NO fixed way to take complex conjugates.*

Consider a complex vector space V *with complex structure* J*. Then consider the map* $-J : V \to V$*. For* $-J$*, its square is also* $-I$*, so it seems that* $(V, -J)$ *also defines a complex vector space! How peculiar!*

If you like, you can define a complex conjugate to be any complex linear map from (V, J) *to* $(V, -J)$ *which is its own inverse. Say* $C : V \to V$ *is such a map. Then* $CJ = -JC$*, and in particular,* $C(iv) = -iC(v)$*, which is exactly what you want, i.e.,* $\overline{iv} = -i\overline{v}$*. And we want* $C = C^{-1}$*, which implies that* $\overline{\overline{v}} = v$*. Any such linear map can be treated as a complex conjugate.*

Now, what are such maps C*? For each complex basis* $v_1, ..., v_n$*, then you can think of a corresponding complex conjugate map* C *as the real linear map that fixes* $v_1, ..., v_n$ *while negating* $Jv_1, ..., Jv_n$*. It is NOT complex linear if you treat it as a map from* (V, J) *to* (V, J)*, but it IS complex linear if you treat it as a map from* (V, J) *to* $(V, -J)$*.*

Conversely, suppose you have a complex conjugate map $C : V \to V$*. Let* $W_+ = \mathrm{Ran}(I + C)$ *and* $W_- = \mathrm{Ran}(I - C)$*. On one hand,* $2v = (I + C)v + (I - C)v$*, so* $W_+ + W_- = V$*. On the other hand, it is obvious that* $C(v + Cv) = Cv + v$ *while* $C(v - Cv) = Cv - v$*, so* W_+ *are all eigenvectors for 1 while* W_- *are all eigenvectors for* -1*. So* W_+ *and* W_- *are disjoint. In short, we have a decomposition* $V = W_+ \oplus W_-$*, where* W_+ *are vectors fixed by* C*, i.e., the so called "real" vectors, while* W_- *are vectors negated by* C*, i.e., the so called "purely imaginary" vectors. Pick any real basis for* W_+*, say* $v_1, ..., v_n$*. Then* $Jv_1, ..., Jv_n$ *form a real basis for* W_-*. So* $v_1, ..., v_n$ *are complex basis for* V*, underwhich* C *is exactly the corresponding complex conjugate map.*

1.6 Complex Inner Products

This chapter serves to illustrate the intuition behind perpendicularity for complex vectors. First recall the traditional definition for real inner product spaces or Euclidean spaces:

Definition 1.45. For a real vector space V, a real inner product is $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that for any $u, v, w \in V$:

1. (bilinear) $\langle u, av + bw \rangle = a\langle u, v \rangle + b\langle u, w \rangle$ for any real $a, b \in \mathbb{C}$, same fpr the first argument.
2. (symmetry) $\langle v, w \rangle = \langle w, v \rangle$.
3. (positive-definiteness) $\langle v, v \rangle \geq 0$ with equality iff $v = 0$.

For contrast, observe the following. (Note that transpose for real stuff is transpose conjugate for complex stuff, so $v^T w$ becomes $\overline{v^T} w$.)

Definition 1.46. For a complex vector space V, a complex inner product is $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ such that for any $u, v, w \in V$:

1. (sesquilinear) $\langle u, av + bw \rangle = a\langle u, v \rangle + b\langle u, w \rangle$ and $\langle au + bv, w \rangle = \overline{a}\langle u, w \rangle + \overline{b}\langle v, w \rangle$ for any complex $a, b \in \mathbb{C}$.
2. (conjugate-symmetry) $\langle v, w \rangle = \overline{\langle w, v \rangle}$.
3. (positive-definiteness) $\langle v, v \rangle \geq 0$ with equality iff $v = 0$.

A complex vector space with a fixed inner product is called a complex inner product space (or historically also called a unitary space, but no one use this name any more). (Note that we always take complex conjugates on the field $\mathbb{C}$, which is well defined, but we never take complex conjugates on complex vectors, because it is NOT well defined.)

Example 1.47. *The standard complex inner product is* $\langle v, w\rangle = v^*w$ *defined on* $\mathbb{C}^n$*. For example,* $\langle (1,i)^T, (-1,i)^T\rangle = 1(-1) + (-i)i = 0$*. So you can see that* $(1,i)^T$ *and* $(-1,i)^T$ *are perpendicular.*

But what does it mean to be complex perpendicular? Consider the realization for a moment. For the vector $(1,i)^T$*, it is realized as* $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix}$*, whereas the vector* $(-1,i)^T$ *is realized as* $B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix}$*. As you can see, the two columns of* A *and the two columns of* B *are ALL perpendicular to each other.*

In short, to say that $v \perp w$ *over* $\mathbb{C}$ *is the same as saying that* v, Jv *are both perpendicular to* w, Jw*.*

Example 1.48. *Say I want to investigate the complex inner product of* $(1,i)^T$ *and* $(1+i, 1-i)^T$*. Note that their realization is* $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix}$ *and* $Y = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 1 \\ -1 & 1 \end{pmatrix}$*. In the real world, a typical inner product could be something like* X^TY*, which corresponds to* $\overline{(1,i)}\begin{pmatrix} 1+i \\ 1-i \end{pmatrix}$*. Think of this as a motivation for the definition of complex inner products.*

Example 1.49. *Note that suppose* $J = diag(J_2, J_2, ..., J_2)$ *as usual, then the realization of complex vector* v *with purely real coordinates is* (v, Jv)*. Take* $v = (1+0i, 2+0i)^T_{\mathbb{C}} = (1,0,2,0)^T_{\mathbb{R}}$ *as an example.*

Suppose v, w *are two complex vectors and we want to study their inner product. Now we pick a basis that involves* v, w*, so that under this basis, the coordinates of* v, w *are both real. Then under this fixed basis, the realization of* v, w *are* (v, Jv) *and* (w, Jw)*. Their complex inner product* v^*w *would corresponds to* $(v, Jv)^T(w, Jw) = \begin{pmatrix} v^Tw & v^T(Jw) \\ (Jv)^Tw & (Jv)^T(Jw) \end{pmatrix}$*. Since the realization of* J *satisfies* $J^T = -J$ *and* $J^2 = -I$*, we see that the upper left and bottum right entry are the same* v^Tw*, which is some REAL inner product structure! Furthermore,* v^Tw *over the real structure should be the real parts of* v^*w*.*

As you can see, v^*w *over the complex structure and* v^Tw *over the real structure are related. Unfortunately, the relation given here depends on a special kinds of basis. What should we do? We can take trace, which is independent of basis and twice of* v^Tw*. So we want to study the relation of the complex inner product* $\langle v, w\rangle$ *and the real inner product* $2\Re\langle v, w\rangle$*.*

Proposition 1.50. *For any complex inner product* $\langle , \rangle$ *on a complex vector space* V*, then* $[v, w] := 2\Re(\langle v, w\rangle)$ *defines a real inner product on* V*. Conversely, if we have a real inner product* $[,]$ *under which the complex structure* J *is orthogonal, (i.e.,* $[v, w] = [Jv, Jw]$ *for all* $v, w \in V$*,) then* $\langle v, w\rangle = \frac{1}{2}[v, w] + \frac{1}{2}i[Jv, w]$ *is a complex inner product.*

Example 1.51. *Recall that for a real vector space with fixedl basis, i.e.,* $\mathbb{R}^n$*, any inner product is induced by some symmetric matrix* $B = B^T$ *as* $[v, w] := v^TBw$*, and all eigenvalues of* B *are positive. Similarly, it is very easy to verify that for a complex vector space with fixed complex basis, i.e.,* $\mathbb{C}^n$*, any complex inner product is induced by some Hermitian matrix* $H = H^*$ *as* $\langle v, w\rangle := v^*Hw$*, and all eigenvalues of* H *are positive real.*

Now consider a complex inner product induced by H *on* $\mathbb{C}^n$*. We can then realize* H *as a* $2n$ *by* $2n$ *symmetric matrix* $\hat{H}$ *whose eigenvalues are all positive, so* $\hat{H}$ *is in fact a real inner product structrue on* $\mathbb{C}^n$*. This correspondence between the complex inner product by* H *and the real inner product by* $\hat{H}$ *is exactly the relation between* $[,]$ *and* $\langle , \rangle$ *in the above Proposition.*

Also consider the $\langle v, w\rangle = \frac{1}{2}[v, w] + \frac{1}{2}i[Jv, w]$ *in the formula. If we replace* $[,]$ *by* $\langle , \rangle$*, then the right hand side is* $\frac{1}{2}\langle v, w\rangle + \frac{1}{2}i\langle Jv, w\rangle = \frac{1}{2}\langle v, w\rangle + \frac{1}{2}i\langle iv, w\rangle = \frac{1}{2}\langle v, w\rangle + \frac{1}{2}\langle v, w\rangle = \langle v, w\rangle$ *as expected.*

Remark 1.52. *So in essense, a complex inner product structure is simple a real inner product structure under which the complex structure* J *is orthogonal. I.e., multiplying by* i *should preserve angles between vectors and lengths of vectors.*

Interestingly, now we have the following.

Theorem 1.53. *The Gram-Schmidt orthogonalization is still applicable to complex inner product spaces. It works exactly the same way. You can also have QR decomposition where Q is now unitary, or has mutually complex orthogonal columns.*

Remark 1.54.

1. *So in effect, for any complex inner product space, you should always first pick a complex orthonormal basis according to Gram-Schmidt, then your complex inner product now looks like $\langle v, w\rangle = v^*w$ on $\mathbb{C}^n$. Then you can just use your regular intution via realization if you like.*
2. *If you have any abstract comlpex inner product $\langle v, w\rangle$, you could just define the adjoint of A to be the unique matrix B such that $\langle v, Aw\rangle = \langle Bv, w\rangle$. Then you can define Hermitian and Unitary matrices abstractly, and everything will still be fine.*
3. *There is a spectral theorem for all complex matrices A that satisfies $AA^* = A^*A$. These are called normal matrices. They will always be diagonalizable by some unitary matrices (you can find complex orthonormal eigenbasis). Check out the lecture note for Linear Algebra One if you like.*
4. *Just like for symmetric matrices we have a corresponding quadratic form, for Hermitian matrices we have Hermitian form. The theories are basically identical, so you can read it at your own leasure.*

Example 1.55 (Quantum Mechanics)**.** *In Quantum Mechanics, the typical model of the world is the following. Say we want to study some system. First of all, all the possible states of the system form a infinite-dimensional complex inner product space H, also called a Hilbert space. The complex strucutre is used to define the concepts of spins, which I don't quite understand myself, so I will not elaborate. Given two states v, w, the inner product $\langle v, w\rangle$ describes the correlation between the two states. (It is a complex number because we want to take their spins into account into this correlation.)*

An observable is a self-adjoint (i.e., Hermitian) linear map $A : H \to H$ (say postition or velocity or momentum etc.). Don't think of this as a self-adjoint linear map. Instead, think of this as a collection of mutually orthogonal eigenvectors v_λ and corresponding eigenvalues λ. The eigenvalues are the result of your observation at the corresponding eigenstates. Say if the state is at some eigenvector v_λ, then your observation would return the value λ with 100% certainty.

For a generic state v, it is a linear combination of the eigenstates $v = \sum\langle v_\lambda, v\rangle v_\lambda$. So when you perform the observation, you get a probability distribution of results. You have $\frac{|\langle v_\lambda,v\rangle|^2}{|\langle v,v\rangle|^2}$ probability to get the result λ.

So why Hermitian? Because we want the states with "certain results" to be independent to each other, so they should be orthogonal. We also want the eigenvalues, the readings of our observation, to be real. This forces the matrix to be self-adjoint.

Now think about the chang of an observable via time. The time evolution itself must keep independent (orthogonal) states independent, and it must respect the dynamical structure. From HW, you see that this implies it must be both orthogonal and symplectic, and hence unitary. So the evolution through time is some unitary linear map $U_t : H \to H$ and it sends A to $U_t^*AU_t$, i.e., it behaves like a change of basis, which should be expected.

1.7 Fast Fourier Transform

We start by looking at the fourier matrix F_n whose (i, j) entry is $\omega^{(i-1)(j-1)}$ where $\omega = \cos(2\pi/n) + i\sin(2\pi/n)$. As you can check, $1, \omega, ..., \omega^{n-1}$ are all distinct complex numbers, and $\omega^n = 1$, so ω is an n-th root of 1 among the complex numbers. For a typical example, we have $F_4 = \begin{pmatrix}1 & 1 & 1 & 1\\ 1 & \omega & \omega^2 & \omega^3\\ 1 & \omega^2 & \omega^4 & \omega^6\\ 1 & \omega^3 & \omega^6 & \omega^9\end{pmatrix} = \begin{pmatrix}1 & 1 & 1 & 1\\ 1 & i & -1 & -i\\ 1 & -1 & 1 & -1\\ 1 & -i & -1 & i\end{pmatrix}$.

As you can see, it appears that $F_n^T = F_n$. However, it is NOT Hermitian (for example, its diagonal is not real). However, you can verify that it is a multiple of a unitary matrix (also will be proven later in another context), so it satisfy the spectral theorem for normal operators. In particular, you can also check that $\frac{1}{n}\overline{F_n} = F_n^{-1}$.

The fourier matrix is closely related to the Fourier series and Fourier Transforms. In Calculus we learned that Fourier series is very important. For a periodic function $f(x)$ with period 2π, you can try to decompose it into different frequencies via fourier series, and write it as a linear combination of sines and cosines. Say we have maybe $f(x) = \sum c_k e^{kix}$. Here note that $e^{ix} = \cos x + i \sin x$, so e^{ix} is just a lazy way to write sine and cosine simultaneously.

Suppose we have a decomposition $f(x) = c_0 + c_1 e^{ix} + c_2 e^{2ix} + c_3 e^{3ix}$. Given c_0, c_1, c_2, c_3, what do we know about the function $f(x)$? Well, if you apply F_4 to the vector $\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}$, then you can verify that you have $\begin{pmatrix} f(0) \\ f(\pi/2) \\ f(\pi) \\ f(3\pi/2) \end{pmatrix}$. As you can see, you get four points on the graph of $f(x)$. This is the forward direction.

But consider the backward direction as well. Suppose we have $f(x) = c_0 + c_1 e^{ix} + c_2 e^{2ix} + c_3 e^{3ix}$ where the c_i are unknown. How to find the fourier coefficient of $f(x)$? We could evaluate $f(0), f(\pi/2), f(\pi), f(3\pi/2)$, and then compute $F_4^{-1} \begin{pmatrix} f(0) \\ f(\pi/2) \\ f(\pi) \\ f(3\pi/2) \end{pmatrix} = \frac{1}{n}\overline{F_4} \begin{pmatrix} f(0) \\ f(\pi/2) \\ f(\pi) \\ f(3\pi/2) \end{pmatrix}$. As you can see, by evaluating at merely a few points and apply $\frac{1}{n}\overline{F_n}$, we can conveniently obtain the Fourier coefficients.

As a review, let us check some basic properties of the fourier matrix.

Example 1.56.

1. *Since $\frac{1}{\sqrt{n}} F_n$ is unitary, F_n has mutually orthogonal columns.*

2. *The (i,j) entry of F_n^2 is the coordinate-wise product between $(1, \omega^{i-1}, \omega^{2(i-1)}, ..., \omega^{(n-1)(i-1)})^T$ and $(1, \omega^{j-1}, \omega^{2(j-1)}, ..., \omega^{(n-1)(j-1)})^T$, so it is the inner product between $\overline{(1, \omega^{i-1}, \omega^{2(i-1)}, ..., \omega^{(n-1)(i-1)})^T}$ and $(1, \omega^{j-1}, \omega^{2(j-1)}, ..., \omega^{(n-1)(j-1)})^T$. Recall that $\overline{\omega^k} = \omega^{-k}$, so we are looking at the inner product between $(1, \omega^{1-1}, \omega^{2(1-i)}, ..., \omega^{(n-1)(1-i)})^T$ and $(1, \omega^{j-1}, \omega^{2(j-1)}, ..., \omega^{(n-1)(j-1)})^T$, i.e., the inner product between the j-th column and the $(n+2-i)$-th or $(2-i)$-th column. So this is 0 unless $i = j = 1$ or $i + j = n + 2$, in which case this entry is n. For another proof, you can also directly see that the (i,j) entry of F^2 is $\sum_{k=0}^{n-1} \omega^{k(i+j-2)}$. So this is 0 unless $i + j - 2$ is a multiple of n, according to our knowledge from number theory.*

3. *Let $U = \frac{1}{\sqrt{n}} F_n$. Then according to above calculation, $U^2 = \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix}$ where A has ones on the anti-diagonal and 0 everywhere else. Then $U^4 = I$. As a result, all eigenvalues of U are 1 or -1 or i or $-i$.*

4. *Consider $U^2 = \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix}$. Clearly we need $\lfloor \frac{n-1}{2} \rfloor$ column swaps to get U^2 into I. So $\det(U^2) = 1$ if $n = 1, 2 \pmod 4$ and $\det(U^2) = -1$ if $n = 3, 4 \pmod 4$. In particular, you see that these swaps are mutually disjoint, so in fact U^2 has eigenvalue -1 with multiplicity $\lfloor \frac{n-1}{2} \rfloor$, and eigenvalue 1 with multiplicity $n - \lfloor \frac{n-1}{2} \rfloor$. So for U, the sum of multiplicities for 1 and -1 is $n - \lfloor \frac{n-1}{2} \rfloor$, while the sum of multiplicities for i and $-i$ is $\lfloor \frac{n-1}{2} \rfloor$.*

5. *As you can see from $\det(U^2)$, the precise multiplicities of eigenvalues of U is periodic mod 4. You can check the answer in wikipedia under the entry Discrete Fourier transform and find out the precise*

formula for $\det(U)$*, and we have* $\det(F_n) = n^{n/2}\det(U)$*. However, we do know that* $\det(U)$ *is a product of some 1 and -1 and* i *and* $-i$*, so* $\det(U) = \pm 1$ *or* $\pm i$*, which might be enough sometimes.*

Suppose you want to compute the first 1000 fourier coefficients (say you know the rest are probably noises or measurement errors). In effect, you want to quickly apply F_{1000}. How should you do it? By brute fource, this is a 1000 by 1000 matrix, and calculating with it needs millions of calculations. That would take forever. So a better approach is the Fast Fourier Transfour. We start by looking at F_{1024}, reduce it to F_{512}, then reduce it to F_{256}, and so forth, until we reach $F_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. So in 10 steps, we reduce the problem to a much smaller one. In the end, one million calculations will be reduced to merely 5000 calculations. Imagine the gain in speed in signal processing and etc. This is ranked as the top 10 algorithms of the 20-th centry by the IEEE journal Computing in Science and Engineering.

Example 1.57. *Consider* $F_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$*. Observe the relation between its first and third column, and between its second and forth column. You can see that the first and third coordinates of corresponding columns are the same, and the second and forth coordinates are negated.*

Let us now swap the columns to bring the original first and third column togehter, and the original second and forth column together. Then we have $F_4 P_{23} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & i & -i \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -i & i \end{pmatrix}$*. Hey, note that the upper left corner and lower left corner is exactly* $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = F_2$*! In fact, let* $D_2 = diag(1, i)$*, we have* $F_4 P_{23} = \begin{pmatrix} F_2 & D_2F_2 \\ F_2 & D_2F_2 \end{pmatrix} = \begin{pmatrix} I_2 & D_2 \\ I_2 & -D_2 \end{pmatrix} \begin{pmatrix} F_2 & 0 \\ 0 & F_2 \end{pmatrix}$*. So step by step, we have extracted* F_2 *out of* F_4*!*

Theorem 1.58 (Fast Fourier Transform)**.** *We have the following decomposition, where* $D_n = (1, \omega, ..., \omega^{n-1})$ *where* $\omega = \cos(\pi/n) + i\sin(\pi/n)$*, and* P *is a matrix permuting all odd columns to the left and all even columns to the right.*

$$F_{2n} = \begin{pmatrix} I_n & D_n \\ I_n & -D_n \end{pmatrix} \begin{pmatrix} F_n & 0 \\ 0 & F_n \end{pmatrix} P.$$

Proof. Do it yourself. Same idea as Example 1.57. □

Example 1.59. *Here's what happen after a recursion. You will have*

$$F_{4n} = \begin{pmatrix} I_{2n} & D_{2n} \\ I_{2n} & -D_{2n} \end{pmatrix} \begin{pmatrix} I_n & D_n & 0 & 0 \\ I_n & -D_n & 0 & 0 \\ 0 & 0 & I_n & D_n \\ 0 & 0 & I_n & -D_n \end{pmatrix} \begin{pmatrix} F_n & 0 & 0 & 0 \\ 0 & F_n & 0 & 0 \\ 0 & 0 & F_n & 0 \\ 0 & 0 & 0 & F_n \end{pmatrix} P.$$

Here P *is a permutation matrix that put all* $(1 \bmod 4)$ *columns to the left, followed by the* $(3 \bmod 4)$ *columns, followed by the* $(2 \bmod 4)$ *columns, and followed by the* $(4 \bmod 4)$ *columns.*

Proof. Do it yourself. □

How fast is the fast fourier transform? Say we use it on F_{10}, then a million calculations is simplified to 5000 calculations.

Problem 1. *What would happen to* F_{3n}*? Can you do something similar?*

1.8 Fundamental Theorem of Algebra

For our study of matrices, we use complex numbers mainly because of the following theorem.

Theorem 1.60 (Fundamental Theorem of Algebra)**.** *Any complex polynomial of degree n has exactly n roots counting multiplicity.*

Let us first consider a topological proof.

Proof. Since you probably have NOT learned topology yet, this proof is very sketchy, but you can get a vague idea nonetheless.

First, if you can show that any complex polynomial has at least one root, then you can prove the theorem by induction easily. So we prove that instead.

Consider $p(z) = z^k$ for a fixed k. If $|z| = r$, then $|p(z)| = |z^k| = |z|^k = r^k$. So on the complex plane, it maps every circle centered at the origin of radius r to a circle with radius r^k. This map is surjective. If you go around the input circle once, you would go around the output circle exactly k times. You can check these by hand, by considering the inputs $re^{i\theta}$ and consider the period of θ on the output $r^k e^{ik\theta}$.

Consider a generic polynomial $p(z)$, say $p(z) = z^3 + 4z^2 + 5z + 2$ or something. Suppose we only consider input z with extremely large norm, say $|z| = 10^10$. Then since the input is so large, z^3 is also extremely large, and comparatively, the smaller degree terms $4z^2 + 5z + 2$ would appear trivial. So for extremely large input, we have $p(z) \approx z^3$. Then given a circle centered at the origin with extremely large radius r, the output would also approximately be a circle with absurdly large radius r^3. By going around the input circle once, you must approximately go around this super large circle three times, with comparatively tiny perturbations everywhere.

But what about $p(0)$? If this is 0, then our polynomial has at least one root, and we are done. If $p(0) = w$ for some complex number $w \neq 0$, then consider the following: From our extremely large circle with radius $r = 10^10$, we gradually shrink r until it reach 0. Then the output would go from the absurdly large circle around origin to become smaller and smaller circles, until it stabilize towards $p(0) = w$. Then somewhere the output circle must sweep through the origin. So $p(z) = 0$ has a solution.

Essentially, this proof is using a higher-dimensional intermediate value theorem. $\square$

This proof is probably the best proof out there, but unfortunately, since we do not really know topology yet, this proof is not rigorous. So instead, this section aim to provide a rigorous proof, using only linear algebra. First, we stop thinking about polynomials, and think only about matrices.

Lemma 1.61. *For any complex polynomial $p(z)$, there is a complex matrix A such that p is a multiple of the characteristic polynomial of A.*

Proof. Say $p(z) = \sum a_k z^k$ with $a_n = 1$. Then let $A = \frac{1}{a_n}\begin{pmatrix} 0 & 0 & ... & 0 & -a_0 \\ a_n & 0 & ... & 0 & -a_1 \\ 0 & a_n & ... & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & a_n & -a_{n-1} \end{pmatrix}$. As a test of your linear algebra skill, see if you can calculate the determinant of $zI - A$ for an unknown z.

(Hint: you can use column expansion and use induction. Alternatively, you can do a forward elimination, and you will see how the polynomial is build up from this matrix.) $\square$

Now you need to be very careful. Say, if you want to invoke some diagonalization of a matrix, you probably should rethink your argument. Our goal is to prove the fundamental theorem of algebra, so you cannot USE the fundamental theorem of algebra or any of its consequence.

Lemma 1.62. *Let p be the characteristic polynomial of A. Then p has a root iff A has an eigenvector.*

Proof. You should already know this. Here we merely present the proof to ensure you that this part requires no fundamental theorem of algebra, and thus there is no circular logic.

Note that $p(\lambda) = 0$ iff $\det(\lambda I - A) = 0$, iff there is a non-zero vector v in the kernel of $\lambda I - A$, and iff $Av = \lambda v$, an eigenvector! ☐

So, our goal now is the following. We aim to prove that every complex square matrix has at least one eigenvector, which would imply the fundamental theorem of algebra. How to do that? We start simple, by looking at real matrices.

Proposition 1.63. *Every real polynomial of odd degree has at least one root.*

Proof. Say suppose the leading coefficient is positive, then $p(\infty) = \infty$ while $p(-\infty) = -\infty$, so by intermediate value theorem we are done. ☐

Corollary 1.64. *Every real square matrix of odd dimension has at least one real eigenvalue.*

Lemma 1.65. *For any real vector space V of odd dimension and any pair of commuting real matrices A, B on it, then A, B share a common real eigenvector (but maybe for different eigenvalues.)*

Proof. We prove this by induction on dimension of V. (The trouble of using induction is mainly caused by the fact that we cannot use fundamental theorem of algebra....)

If $\dim V = 1$, then the statement is trivial. Now suppose there is a counter example V and A and B. We take a counter example with smallest possible $\dim V$.

Take a real eigenvalue λ of A, which must exist. If $A(v + iw) = \lambda(v + iw)$ for real vectors v, w, then $A(v - iw) = A\overline{(v + iw)} = \overline{A(v + iw)} = \overline{\lambda(v + iw)} = \lambda(v - iw)$. So $Av = \lambda v$ and $Aw = \lambda w$. Since $v + iw \neq 0$, we see that at least one of v, w is non-zero, so A has real eigenvectors.

Consider $\mathrm{Ker}_{\mathbb{R}}(\lambda I - A)$, the collection of all real vectors v such that $Av = \lambda v$, and $\mathrm{Ran}_{\mathbb{R}}(\lambda I - A)$, then range of $\lambda I - A$ over real inputs. Then if $v \in \mathrm{Ker}_{\mathbb{R}}(\lambda I - A)$, $ABv = BAv = \lambda Bv$, so B maps $\mathrm{Ker}_{\mathbb{R}}(\lambda I - A)$ to itself. Similarly, if $v \in \mathrm{Ran}_{\mathbb{R}}(\lambda I - A)$, then $v = (\lambda I - A)w$ for some real vector w. Then $Bv = B(\lambda I - A)w = (\lambda I - A)Bw \in \mathrm{Ran}_{\mathbb{R}}(\lambda I - A)$. So B maps $\mathrm{Ran}_{\mathbb{R}}(\lambda I - A)$ to itself as well.

Now since $\dim \mathrm{Ker}_{\mathbb{R}}(\lambda I - A) + \dim \mathrm{Ran}_{\mathbb{R}}(\lambda I - A)$ is odd, one of the two space must have odd dimension.

If both spaces have smaller dimension than V, then we find a subspace W of odd dimension smaller than $\dim V$, and A, B commute over W. So A, B share does not share a common eigenvectors in W. This is a smaller counter example than V, contradiction.

If $\mathrm{Ker}_{\mathbb{R}}(\lambda I - A) = V$, then $A = \lambda I$, and all vectors are eigenvectors of A. So take any eigenvector of B (recall that V has odd dimension so B does indeed have eigenvectors), and we have a common eigenvector, contradiction.

If $\mathrm{Ran}_{\mathbb{R}}(\lambda I - A) = V$, then $\mathrm{Ker}_{\mathbb{R}}(\lambda I - A) = 0$, so λ is not really an eigenvalue of A, contradiction. ☐

We now proceed to copy these ideas into complex matrices.

Lemma 1.66. *Every complex square matrix of odd dimension has at least one eigenvalue.*

Proof. We will show that $A \in M_n(\mathbb{C})$ has an eigenvector if n is odd.

Consider the set of Hermitian matrices, $\mathcal{H}_n$. These matrices must have real entries on the diagonal, and we only need to specify the upper triangular portion to specify the whole matrix. So the real dimension is $\dim_{\mathbb{R}} \mathcal{H}_n = n + 2(1 + 2 + ... + (n-1)) = n^2$. So this is a real vector space of odd dimension.

Define $L_1 : \mathcal{H}_n \to \mathcal{H}_n$ by $B \mapsto \frac{AB+BA^*}{2}$, and define $L_2 : \mathcal{H}_n \to \mathcal{H}_n$ by $B \mapsto \frac{AB-BA^*}{2i}$. First, you can easily check that if B is Hermitian, then indeed $L_1(B)$ and $L_2(B)$ are Hermitian. Second, you can check that they are linear maps. Third, you can check that $L_1 L_2 = L_2 L_1$ always! So they are commuting matrices over a real vector space of odd dimension! So they share a common eigenvector!

Say $L_1(B) = \lambda_1 B$ and $L_2 B = \lambda_2 B$ for some nonzero Hermitian matrix B. Then $AB = L_1(B) + iL_2(B) = (\lambda_1 + i\lambda_2)B$. So ANY column of B is an eigenvector of A! We are done! ☐

Lemma 1.67. *For any complex vector space V of odd dimension and any pair of commuting matrices A, B on it, then A, B share a common eigenvector (but maybe for different eigenvalues.)*

Proof. Easier than the real version! It is basically the same induction. Take any eigenvalue of A, consider $\mathrm{Ker}(\lambda I - A)$ and $\mathrm{Ran}(\lambda I - A)$, and show that B sends each to itself. $\square$

In fact, let us generalize this a little bit further.

Lemma 1.68. *Fix any interger m. Suppose we see that $A \in M_n(\mathbb{C})$ has an eigenvector for all n not divisible by m, then any commuting $A, B \in M_n(\mathbb{C})$ shares an eigenvector whenever n is not divisible by m.*

Now we attack even dimensions. The key is the following:

Lemma 1.69. *Any degree two complex polynomial has two roots counting multiplicity.*

Proof. They are $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$, as you have learned in high school. $\square$

Proof of FTA. Say $A \in M_n(\mathbb{C})$ for some $n = n'2^k$ with odd n'. We perform induction on k.

If $k = 0$, then n is odd and we are done. So the base case is already done. Also note that commuting matrices in this case share a common eigenvector.

For generic k, suppose we are done with all smaller k. We consider S_n, the set of complex matrices with $A = A^T$. Note that we do NOT take conjugate! Note that S_n is a vector space with dimension $\frac{1}{2}n(n+1) = (n+1)n'2^{k-1}$, so any complex-linear map on S_n must have an eigenvector. Furthermore, using the previous lemma with $m = 2^k$ and our induction hypothesis, any pair of commuting complex-linear map on S_n must share an eigenvalue.

Define $L_1 : S_n \to S_n$ by $B \mapsto AB + BA^T$ and define $L_2 : S_n \to S_n$ by $B \mapsto ABA^T$. Then you can check that they are complex linear and commute. So they share some common eigenvector, say $L_1(B) = \lambda_1 B$ and $L_2(B) = \lambda_2 B$ for some nonzero B.

Then $AB + BA^T = \lambda_1 B$. Multiplying A from the left, we have $\lambda_1 AB = A^2B + ABA^T = A^2B + \lambda_2 B$. So $(A^2 - \lambda_1 A + \lambda_2)B = 0$.

Let $p(z) = z^2 - \lambda_1 z + \lambda_2$. Then $p(z) = (z-a)(z-b)$ for some $a, b \in \mathbb{C}$. So $A^2 - \lambda_1 A + \lambda_2 = (A-aI)(A-bI)$, and now we have $(A - aI)(A - bI)B = 0$.

If $(A - bI)B = 0$, then any column of B is an eigenvector of A. If $(A - bI)B \neq 0$, then any column of $(A - bI)B$ is an eigenvector of A. So we are done. $\square$

2 Jordan Normal Form

Matrices are linear maps. And linear maps are by nature very geometric. Ideally, we should be able to SEE a linear map upon seeing its matrix. And whenever you have proven something about matrices, you should always strive to SEE the corresponding geometric information.

2.1 Cyclic permutation matrices

We first investigate cyclic permutation matrices, i.e., matrices like $P_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$.What do they do? Do they do things? Let's find out.

Example 2.1 (Geometric View)**.** *Obviously one can check that $P_n \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} a_2 \\ \vdots \\ a_n \\ a_1 \end{pmatrix}$. So algebraically, these matrices would simply permute these coordinates.*

The 3-dim picture is very clear then. This linear map send x-axis to y-axis, and y-axis to z-axis, and z-axis to x-axis. As you can imagine, this is a rotation by $\frac{2\pi}{3}$ around the line $x = y = z$. The 2-dim picture *is also clear. P_2 sends x-axis to y-axis and y-axis back to x-axis. So P_2 would "rotate" the whole xy-plane around the line $x = y$ by 180, and thereby flip the plane about the line $x = y$. (We put quotation marks around "rotation", because this is not really a rotation. You can see it as a rotation ONLY from a 3-dim point of view, but inside that 2-dim plane, this is a reflection, not a rotation.) Generically, P_n is a "rotation" around the line $x_1 = ... = x_n$ and send each x_i-axis to x_{i+1}-axis.*

Example 2.2 (Algebraic View)**.** *You can easily see that if $P_nv = \lambda v$, then the real solution is $\lambda = 1$ and $v = \begin{pmatrix} a \\ \vdots \\ a \end{pmatrix}$ for some constant a, and for even n we also have $\lambda = -1$ and $v = \begin{pmatrix} a \\ -a \\ a \\ -a \\ \vdots \\ a \\ -a \end{pmatrix}$. In the odd case, this is precisely the "axis of rotation" as described in the geometric view above! And in the even case, this line is precisely perpendicular to some hyperplane of reflection. The rother eigenvalues must be complex numbers. (And in fact pairs of complex conjugates.)*

Next up we observe that we always have $P_n^n = I$. If P_n has eigenvalues $\lambda_1, ..., \lambda_n$ counting algebraic multiplicity, then P_n^n is suppose to have eigenvalues $\lambda_1^n, ..., \lambda_n^n$ counting algebraic multiplicity. But $P_n^n = I$ and all eigenvalues of I are 1, so we see that all eigenvalues of P_n is an n-th root of 1. Conversely, if ω is a complex number with $\omega^n = 1$, then $P_n \begin{pmatrix} 1 \\ \omega \\ \omega^2 \\ \vdots \\ \omega^{n-1} \end{pmatrix} = \omega \begin{pmatrix} 1 \\ \omega \\ \omega^2 \\ \vdots \\ \omega^{n-1} \end{pmatrix}$. Now, since there are n roots for the polynomial $x^n - 1$, it seems that all distinct n roots are the eigenalues of P_n. As a result, the eigenvalues of P_n are $1, \omega, ..., \omega^{n-1}$ where $\omega = \cos(\frac{2\pi}{n}) + i\sin(\frac{2\pi}{n})$.

We immediately have some funny observations.

1. *Since P_n has n distinct roots, it is diagonalizable. In fact, since P_n is an orthogonal matrix to begin with, the spectral theorem for orthogonal matrices already tells us that P_n is diagonalizable by unitary matrices and all eigenvalues have norm 1.*

2. *In fact, consider the following Fourier matrix $F_n = \begin{pmatrix} 1 & 1 & 1 & ... & 1 \\ 1 & \omega & \omega^2 & ... & \omega^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & ... & \omega^{(n-1)^2} \end{pmatrix}$. (This is also a special type of Vandermonde matrix.) its columns are exactly the eigenvectors for P_n. So in fact $P_n = F_nDF_n^{-1}$ where $D = diag(1, \omega, ..., \omega^{n-1})$. So immediately, we see that F_n has mutually complex orthogonal columns. In fact $\frac{1}{\sqrt{n}}F_n$ would be an unitary matrix.*

3. *Obviously $\mathrm{tr}(P_n) = 0$, and we can see $\det(P_n) = (-1)^{n-1}$ since it is $(n-1)$ column swaps away from becoming I. So we have proven that $\sum_{i=0}^{n-1} \omega^i = 0$ and $\prod_{i=0}^{n-1} \omega^i = (-1)^{n-1}$. These two identities are very fundamental in number theory, complex analysis and representation theories. You can of course prove these some other elementary ways, but our linear algebraic proof here does not even require any calculation.*

Example 2.3 (Deeper geometric picture)**.** *Consider P_4 with determinant -1. Since it has determinant -1, it is NOT really a rotation, but in fact some reflection. What is its precise geometric action on $\mathbb{R}^4$?*

First, it clearly fixes the line parametrized as $(t, t, t, t)^T$. It is also easy to see that it reflect the line parametrized as $(t, -t, t, -t)^T$. The orthogonal complement to these two lines is the 2-dim vector space W

spanned by $v = \Re(1, i, -1, -i)^T = (1, 0, -1, 0)^T$ *and* $w = \Im(1, i, -1, -i)^T = (0, 1, 0, -1)^T$. *You can check that* $P_4 w = v$ *and* $P_4 v = -w$, *so the action of* P_4 *on the two-dim plane* W *under the basis* v, w *is the matrix* $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, *i.e., rotation clock-wise by 90 degree. So what does* P_4 *do? It first perform a quarter-rotation around the axis-plane parametrized as* $(s, t, s, t)^T$, *and then reflect about the hyperplane with equation* $w - x + y - z = 0$ *for vectors* $(w, x, y, z)^T$.

Try to see the deeper geometric picture for P_5 *if you like. You will orthogonally decompose* $\mathbb{R}^5$ *into a line and two planes. The line will be fixed by* P_5 *while the two planes are rotated by* $\frac{2\pi}{5}$.

2.2 Nilpotent Jordan Block

A typical matrix, as shown in the last subsection, can be analyzed by using eigenvalues and eigenvectors as above. We see its eigenstuff and we can have a very clear and very detailed picture of its geometric action. But what about a matrix with almost no eigenvectors?

Example 2.4. *Consider this:* $N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. *This is upper triangular, so its eigenvalues are its diagonal entries, which are all 0. So it has eigenvalue 0 with algebraic multiplicity 3. What about its geometric multiplicity? Well,* $\operatorname{Ker} N$ *is spanned by* $(1, 0, 0)^T$. *So it only has geometric multiplicity 1. Clearly* N *is NOT diagonalizable. What should we do now?*

We should check its orbit. You can check that $N \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} b \\ c \\ 0 \end{pmatrix}$. *So* N *is in effect "shifting upward" your coordinates, and fill up the empty spot with 0. Typically* N *would send* $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ *to* $\begin{pmatrix} b \\ c \\ 0 \end{pmatrix}$ *to* $\begin{pmatrix} c \\ 0 \\ 0 \end{pmatrix}$ *and finally to the 0 vector.*

Funnily, observe that P_3 *is almost identical to* N *except for the lower left entry, and* P_3 *indeed almost behave the same way, except that* $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ *to* $\begin{pmatrix} b \\ c \\ a \end{pmatrix}$. *So* N *almost behaves as performing* P_3 *first, and then delete the third coordinate. In terms of matrices, it says* $N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} P_3$, *and the matrix here essentially means projection to the xy-plane. So now the geomtric action of* N *is very clear. We first do a rotation according to* P_3, *and the project. Repeated application of* N *means we rotate as* P_3 *then project then rotate again then project again and so on. In my mind* N *is like a "spiral-in" process. The energy of the whole space is rotated into the xy-plane, and the xy-plane is rotated into the x-axis, and finally the x-axis is smashed into the origin.*

The action of $N_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ *and so on are very similar.*

Recall the following definition and property (or check out the lecture notes from Linear Algebra One.)

Definition 2.5. A square matrix A is nilpotent if $A^k = 0$ for some integer k.

Proposition 2.6. *An* $n \times n$ *matrix* A *is nilpotent iff* $A^n = 0$ *iff all its eigenvalues are 0.*

The proof of above propositions are in my lecture notes for linear algebra I.

For any nilpotent A, consider the smallest k such that $A^k = 0$. Clearly $k \leq n$. This is sometimes called the index of the nilpotent matrix A.

Example 2.7. *Suppose $A \in M_n$ is nilpotent with largest possible index, i.e., $A^n = 0$ but $A^{n-1} \neq 0$. Then pick a vector $v \notin \operatorname{Ker} A^{n-1}$, and consider $v, Av, A^2v, ..., A^{n-1}v$. None of these vectors are 0. On the other hand, suppose they are linearly dependent. Then we have some linear combination of them $a_0v + a_1Av + ... + a_{n-1}A^{n-1}v = 0$ where some coefficient is non-zero. Let i be the smallest index such that $a_i \neq 0$, then applying A^{n-1-i} to both side of the equation, and use the fact that $A^n = 0$, we see that $a_iA^{n-1}v = 0$, so $A^{n-1}v = 0$, a contradiction! Therefore, $A^{n-1}v, A^{n-2}v, ..., Av, v$ are linearly independent. Since we now have n linearly independent vectors, they form a basis. Note that, under this new basis, the NEW matrix expression of A would now be N_n as in the example above!*

Pick the matrix $B = (A^{n-1}v, A^{n-2}v, ..., Av, v)$. Then we have $A = BN_nB^{-1}$. You can see that any nilpotent matrix of maximal index is essentially N_n under some weird basis. The geometric action of A is exactly the same as the geometric action of N_n, up to some change of basis.

Definition 2.8. A nilpotent Jordan block is a matrix N_k for some k.

Theorem 2.9 (Jordan decomposition for nilpotent matrices)**.** *If A is nilpotent, then A is similar to the block diagonal matrix $diag(N_{k_1}, N_{k_2}, ..., N_{k_t})$ for various nilpotent Jordan blocks of various sizes. We say $diag(N_{k_1}, N_{k_2}, ..., N_{k_t})$ is the Jordan normal form of A. The Jordan normal form is unique up to the permutation of the Jordan blocks. Furthermore, if A is real, then the Jordan normalization can be done via real matrices, i.e., $A = BNB^{-1}$ for some real matrix B and its Jordan normal form N.*

We will note give any explicit proof, but merely provide the following example, from which you should be able to deduce the proof yourself.

Example 2.10. *Suppose $A \in M_7$ is nilpotent with index 4, i.e., $A^4 = 0$ and $A^3 \neq 0$. Say $\dim \operatorname{Ker}(A) = 3, \dim \operatorname{Ker}(A^2) = 5, \dim \operatorname{Ker}(A^3) = 6, \dim \operatorname{Ker}(A^4) = 7$. From $\operatorname{Ker}(A^3)$ to $\operatorname{Ker}(A^4) = \mathbb{R}^7$, it seems to miss only one dimension. So pick any vector $v_1 \in \operatorname{Ker}(A^4) - \operatorname{Ker}(A^3)$. Now $\mathbb{R}^7 = \operatorname{Ker}(A^4)$ is the span of v_1 and $\operatorname{Ker}(A^3)$, and it remains to figure out $\operatorname{Ker}(A^3)$.*

Now we look inside $\operatorname{Ker}(A^3)$. We know Av_1 is already in it, but not in $\operatorname{Ker}(A^2)$. Since the dimension of $\operatorname{Ker}(A^3)$ is only one more than the dimension of $\operatorname{Ker}(A^2)$, we see that $\operatorname{Ker}(A^3)$ is exactly the span of Av_1 and $\operatorname{Ker}(A^2)$. There is nothing to pick at this stage. It remains to figure out $\operatorname{Ker}(A^2)$.

Now inside $\operatorname{Ker}(A^2)$, we know A^2v_1 is already in it, but not in $\operatorname{Ker}(A)$. So inside $\operatorname{Ker}(A^2)$ of dimension 5, we have $\operatorname{Ker}(A)$ and A^2v_1, whose span has dimension 4, which is one dimension short. So we pick any vector v_2 in $\operatorname{Ker}(A^2)$ but NOT in the span of $\operatorname{Ker}(A)$ and A^2v_1. Now $\operatorname{Ker}(A^2)$ is the span of v_2, A^2v_1 and $\operatorname{Ker}(A)$. It remains to figure out $\operatorname{Ker}(A)$.

Finally, inside $\operatorname{Ker}(A)$, we know A^3v_1 and Av_2 are already in it. Since we pick v_2 to be linearly independent from A^2v_1, clearly A^3v_1 and Av_2 are linearly independent as well. Now $\dim \operatorname{Ker}(A) = 3$ but we only have two linearly independent vectors here, so we are one short. Pick any vector v_3 in $\operatorname{Ker}(A)$ but not in the span of A^3v_1 and Av_2. Then $\operatorname{Ker}(A)$ is now the span of A^3v_1, Av_2 and v_3. Everything is now accounted for.

In short, we have figure out the following spanning set for the following spaces:

$$\left(\begin{array}{c|ccccccc}
\operatorname{Ker}(A) & A^3v_1 & Av_2 & v_3 & & & & \\
\operatorname{Ker}(A^2) & A^2v_1 & v_2 & A^3v_1 & Av_2 & v_3 & & \\
\operatorname{Ker}(A^3) & Av_1 & A^2v_1 & v_2 & A^3v_1 & Av_2 & v_3 & \\
\operatorname{Ker}(A^4) & v_1 & Av_1 & A^2v_1 & v_2 & A^3v_1 & Av_2 & v_3
\end{array}\right)$$

And by counting the number of vectors in each spanning sets, we see that they are in fact basis for these spaces. So a basis for the whole $\mathbb{R}^7$ is simply $A^3v_1, A^2v_1, Av_1, v_1, Av_2, v_2, v_3$. Under this new basis, what is the new matrix expression of A? It becomes
$$\left(\begin{array}{cccc|cc|c}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right) = diag(N_4, N_2, N_1).$$

Let $B = (A^3v_1, A^2v_1, Av_1, v_1, Av_2, v_2, v_3)$*, then* $A = BNB^{-1}$ *where* $N = diag(N_4, N_2, N_1)$*.*

Recall this whole process, essentially we are tracking down the following chart from bottum to top, picking new vectors whenever the dimension don't match.

$$\left(\begin{array}{c|ccc} \mathrm{Ker}(A) & A^3v_1 & Av_2 & v_3 \\ \mathrm{Ker}(A^2) - \mathrm{Ker}(A) & A^2v_1 & v_2 & \\ \mathrm{Ker}(A^3) - \mathrm{Ker}(A^2) & Av_1 & & \\ \mathrm{Ker}(A^4) - \mathrm{Ker}(A^3) & v_1 & & \end{array}\right).$$

As a summary, we first figure out v_1*, which immediately filled up its above with* Av_1, A^2v_1, A^3v_1*. Then the second to last row is already filled up. The third to last row has one missing dimension, so we filled up* v_2*, which immediately filled up its above by* Av_2*. Finally the first row is missing one dimension, so we grab* v_3*. This we we find a very nice basis under which* A *is in Jordan normal form.*

I am not rigorously proving anything here, but you can clearly adapt this procedure to any nilpotent matrix to see the Jordan normal form for any nilpotent matrix.

This process is entirely determined by the dimensions of these various $\mathrm{Ker}(A^k)$*, so the Jordan normal form is unique (up to the permutation of these Jordan blocks). You also see that if* A *is real, then all the vector spaces we are talking about can be taken as real vector spaces, then* B *in the decomposition* $A = BNB^{-1}$ *is also real. This example here should give you more than enough to prove the theorem of Jordan normal form on nilpotent matrices.*

Here is a technical lemma for this process.

Lemma 2.11. *For any nilpotent matrix* A *and any positive integer* k*,* $\dim \mathrm{Ker}(A^{k+1}) - \dim \mathrm{Ker}(A^k) \leq \dim \mathrm{Ker}(A^k) - \dim \mathrm{Ker}(A^{k-1})$*.*

Proof. Say $d = \dim \mathrm{Ker}(A^{k+1}) - \dim \mathrm{Ker}(A^k)$. Pick vectors $v_1, ..., v_d$ such that $v_1, ..., v_d, \mathrm{Ker}(A^k)$ together span $\mathrm{Ker}(A^{k+1})$. In particular, $v_1, ..., v_d, \mathrm{Ker}(A^k)$ are linearly independent. Obviously $Av_1, ..., Av_d \in \mathrm{Ker}(A^k) - \mathrm{Ker}(A^{k-1})$, so I only need to show that $Av_1, ..., Av_d, \mathrm{Ker}(A^{k-1})$ are linearly independent as well.

Suppose $\sum a_iAv_i \in \mathrm{Ker}(A^{k-1})$. This is true iff $A^{k-1}(\sum a_iAv_i) = 0$, iff $A^k(\sum a_iv_i) = 0$, iff $\sum a_iv_i \in \mathrm{Ker}(A^k)$, iff all a_i are zero. So we are done. □

Example 2.12. *But wait, in the above example, what do these mean? Consider the subspace* W_1 *spanneed by* $v_1, Av_1, A^2v_1, A^3v_1$*, the subspace* W_2 *spanned by* v_2, Av_2*, and the subspace* W_3 *spanned by* v_3*. Then* A *maps each* W_i *to* W_i *itself, and you can see that* $V = W_1 \oplus W_2 \oplus W_3$*. So we have a decomposition of* V *into three linearly independent subspaces. Furthermore, on the subspaces* W_1, W_2, W_3 *in the given basis, the matrices for* A *are exactly* N_4, N_2, N_1*, the three Jordan blocks making up the Jordan normal form of* A*.*

As an intuition, we see that for any nilpotent matrix, we can chop up the space into several linearly independent subspaces, and on each our matrix would perform some N_k*, or some "spiral-in" process. A nilpotent matrix always eventually kill everything, and now you see that it kill everything via divide and conquer. Now you are able to visualize any nilpotent matrices.*

There are two important takeaways from this. The first takeaway, orbits are important. From both cyclic permutation matrices, to nilpotent matrices, vector sequences $v, Av, A^2v, ...$ seems to play a very major role. Rotation, translation, shear, nilpotency, these can all be told from such orbits. Say we have a 2 by 2 matrix where all orbits are on various circles around the origin, then you immediately and intuitively see that it is a rotation.

The econd takeaway, spatial decomposition is important. For cyclic permutations, spatial decomposition gives you the most detailed geometric picture. And for nilpotent matrices, spatial decomposition gives you the Jordan normal form and the precise pieces that get spiraled in.

2.3 Spatial Decomposition

(Check out Section 24.1 in the lecture note for Linear Algebra One as a preview for this section. In particular, try to be a little familiar with the notion of linearly independent subspaces.)

Our study of both cyclic permutation matrices and nilpotent matrices revealed the following fact: To visualize a matrix, it is BEST to decompose the space V into smaller subspaces, and hopefully study the action of our linear map on this subspace.

Proposition 2.13. *Suppose $V = V_1 \oplus V_2$. Then there are surjective linear maps $p_1 : V \to V_1$ and $p_2 : V \to V_2$ such that* $\mathrm{Ker}(p_1) = V_2$ *and* $\mathrm{Ker}(p_2) = V_1$. *These are called projections onto components. Furthermore, for any $v \in V$, $v = \begin{pmatrix} p_1(v) \\ p_2(v) \end{pmatrix}$ if the basis is combined from first a basis of V_1, and then a basis of V_2.*

Proof. Pick basis $v_1, ..., v_a$ and $w_1, ..., w_b$ for V_1 and V_2. Together they form a basis for V. Under this basis, p_1 means deleting the last b coordinates, while p_1 means deleting the first a coordinates. You can easily check the rest. $\square$

Example 2.14. $\mathbb{R}^2$ *is the direct sum of subspaces spanned by* $(1,1)^T$ *and spanned by* $(1,0)^T$. *Try to imagine the oblique projections onto these components.*

Example 2.15. *Suppose we have a linear map $A : V \to W$, and a decomposition of vector spaces $V = V_1 \oplus V_2$ and $W = W_1 \oplus W_2$. Then we have the following four interesting linear maps: inclusion $inc_1 : V_1 \to V$ and $inc_2 : V_2 \to V$, and the projections $p_1 : W \to W_1$ and $p_2 : W \to W_2$.*

Then we can consider the following four linear maps $A_{11}, A_{12}, A_{21}, A_{22}$, $A_{ij} = p_j \circ A \circ inc_i$. Basically A_{ij} means we are looking at the map A but we only consider it as an action from V_i to W_j, and try to ignore the rest of V and the rest of W. I claim that we shall have a block form $A = \begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix}$, if the basis for V is combined from first a basis of V_1, and then a basis of V_2, and the basis for W is combined from first a basis of W_1, and then a basis of W_2.

Indeed, consider a vector $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in V$ where $v_1 \in V_1$ and $v_2 \in V_2$. Then $A\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = A\begin{pmatrix} v_1 \\ 0 \end{pmatrix} + A\begin{pmatrix} 0 \\ v_2 \end{pmatrix} = A \circ inc_1 v_1 + A \circ inc_2 v_2 = \begin{pmatrix} p_1 \circ A \circ inc_1 v_1 \\ p_2 \circ A \circ inc_1 v_1 \end{pmatrix} + \begin{pmatrix} p_1 \circ A \circ inc_2 v_2 \\ p_2 \circ A \circ inc_2 v_2 \end{pmatrix} = \begin{pmatrix} A_{11} v_1 + A_{21} v_2 \\ A_{12} v_1 + A_{22} v_2 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$.

I highly encourage you to do some concrete example yourself, say $V = W = \mathbb{R}^4$ and pick some decompositions to see how this work.

So whenever you see a matrix in block form, you are ALWAYS trying to see this matrix according to some spatial decomposition. Let us see another example.

Example 2.16. *Consider $A : U \to V$ and $B : V \to W$ and suppose we want to study the composition BA. Say we have a decomposition $V = V_1 \oplus V_2$. Then we have a corresponding block form $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ and $B = (B_1, B_2)$. Then $BA = (B_1, B_2)\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = B_1 A_1 + B_2 A_2$. This is block matrix multiplication.*

But what does this mean? $B_1 A_1$ is itself a composition of linear maps. It means we are going from U to V_1 and then from V_1 to W. On the other hand, $B_2 A_2$ means we are going from U to V_2 and then from V_2 to W. These are the only two possible routes! So we simply add them together. Check out the following diagram.

$$\begin{tikzcd}
U \arrow[r, bend left, "A_1"] \arrow[rrr, bend left, "A_2"] & V_1 \arrow[rrr, bend right, "B_1"'] & \oplus & V_2 \arrow[r, bend right, "B_2"] & W
\end{tikzcd}$$

Compare this with the following diagram which illustrate the multivariable chain rule for a function $f(x, y)$ as $\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$. Here the four vector spaces are all one-dimensional, with the first $\mathbb{R}$ record the variable t, the second $\mathbb{R}$ record the variable x, the third $\mathbb{R}$ record the variable y, and the last $\mathbb{R}$ record the value for $f(x, y)$.

$$\begin{tikzcd}
\mathbb{R}_t \arrow[r, bend left, "\frac{dx}{dt}"] \arrow[rrr, bend left, "\frac{dy}{dt}"] & \mathbb{R}_x \arrow[rrr, bend right, "\frac{\partial f}{\partial x}"'] & \oplus & \mathbb{R}_y \arrow[r, bend right, "\frac{\partial f}{\partial y}"] & \mathbb{R}_f
\end{tikzcd}$$

Note that derivatives are LINEARIZATION of functions, so the similarity here is NOT a coincidence. In fact, you can think of the rule of block matrices multiplication as a proof for the multivariable chain rule. In particular, you can generalize this and prove that for any function $f : \mathbb{R}^d \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^m$, then the ***total derivative matrix*** *$D(g \circ f) = (Dg)(Df)$, and the multiplication on the right is exactly a matrix multiplication.*

Definition 2.17. For any linear map $A : V \to V$, a subspace W of V is an ***invariant subspace*** for A if $A(W) \subseteq W$. We can write $A_W : W \to W$ for the action of A restricted to W alone.

Example 2.18. *The origin is always a 0-dim invariant subspace. Similarly, the whole space is always invariant as well.*

For any linear map A, if v is an eigenvector, then the span of v is an invariant subspace. In fact, any span of eigenvectors is invariant.

Suppose $A = 0$ as a matrix. Then any subspace is an invariant subspace.

Suppose $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The only three invariant subspaces are $\{0\}$, the line spanned by $(1, 0)^T$, and the whole plane.

Proposition 2.19. *Fix a linear map $A : V \to V$. Suppose we have a spatial decomposition into invariant subspaces, say $V = W_1 \oplus W_2 \oplus W_3$, then the block form for A according to this decomposition is block diagonal as $diag(A_{W_1}, A_{W_2}, A_{W_3})$.*

Proof. Pick basis and then it is obvious. □

Example 2.20. *Spatial decomposition into smaller invariant spaces means that, instead of studying the action of A on the whole space, it is enough to study the action of A on each pieces, which is much smaller and much easier to understand. Essentially A would act on each invariant piece independently, so we can study these smaller pieces independently. If originally the space is 10 dimensional, then originally A might have 100 entries. But if we do a decomposition into smaller invariant subspaces, say $10 = 2+2+3+3$, then according to this decomposition, A would be block diagonal. And now we only need to remember $2^2 + 2^2 + 3^2 + 3^2 = 26$ entries. The problem is now easier.*

Ideally, maybe we can decompose V into as small pieces as possible. So maybe we can decompose V into 1-dimensional invariant subspaces! If we can achieve this, then we in fact performed a diagonalization of A. You can check that if $V = W_1 \oplus ... \oplus W_n$ into one-dimensional subspaces, and each W_i is spanned by w_i, then A is diagonal under the basis $w_1, ..., w_n$.

So diagonallization = decomposition into 1-dimensional invariant subspaces. You also see that why diagonalization, if possible, is very powerful. It is because spatial decomposition allows us to hugely simplify whatever problems we come across.

Example 2.21. *What is the Jordan normal form for nilpotent matrices? It gives us a block diagonal form! So it gives us a spatial decomposition into invariant subspaces. In fact, it is the finest decomposition possible, i.e., the pieces are smallest possible.*

Consider N_k. All its eigenvectors are on the same line, the one spanned by $(1, 0, ..., 0)^T$.

Suppose for contradiction that N_k has a block diagonal form $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$. Recall that any matrix must have eigenvectors. Then A would contribute an eigenvector $\begin{pmatrix} v_A \\ 0 \end{pmatrix}$, and B would contribute another eigenvector $\begin{pmatrix} 0 \\ v_B \end{pmatrix}$. The two are clearly linearly independent, so N_k would have 2 linearly independent eigenvectors, a contradiction! So there are no finer block diagonalizations for each N_k.

2.4 Spatial filtration

The whole Jordan decomposition is originally designed to deal with non-diagonalizable matrices, and to get them as close to diagonalized as possible. A half-way point towards being diagonal is being upper-triangular. Therefore, we take a slight detour and first talk about triangularization.

Recall that diagonalizations are essentially a spatial decomposition into 1-dim invariant subspaces. But what if this is not possible? What if, say, maybe NO two invariant subspaces are linearly independent? Then we can only do $V = V$, and no finer decomposition is possible! And this could indeed happen. Those N_k on $\mathbb{C}^k$ already serves as an example. Or consider $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The only three invariant subspaces are $\{0\}$, the line spanned by $(1,0)^T$, and the whole plane. It is NOT diagonalizable.

However, you may notice this very nice fact about these three invariant subspaces: they contain each other! In fact, we have a chain $\{0\} \subseteq \{(t,0)^T : t \in \mathbb{R}\} \subseteq \mathbb{R}^2$. This is called a filtration. Note that our matrix is upper triangular, and we have a very convenient filtration! This is no coincidence.

Definition 2.22. A filtration of a vector space V is a chain $\{0\} \subseteq V_1 \subseteq V_2 \subseteq ... \subseteq V_n = V$, where each V_i is i-dimensional. So each step the dimension is increased by 1. (Some place also call this a flag. Imagine a point-line-plane filtration of $\mathbb{R}^2$, and it has a "flag" feel to it.)

Example 2.23. *The dimension of a vector space is essentially the length of any filtration chain minus 1. In advanced algebra classes this might even be taken as a way to define dimensions.*

Proposition 2.24. *For a linear map A on a vector space V, if we have a filtration of V by invariant subspaces $\{0\} = V_0 \subseteq V_1 \subseteq V_2 \subseteq ... \subseteq V_n = V$, then pick any $v_i \in V_i - V_{i-1}$, under the basis $v_1, v_2, ..., v_n$, the matrix representation of A is upper-triangular.*

Proof. It is enough to prove that Av_i is spanned by $v_1, ..., v_i$ and we do not need $v_{i+1}, ..., v_n$. But since $v_i \in V_i$ and V_i is invariant by A, we see that $Av_i \in V_i$, which is indeed spanned by $v_1, ..., v_i$. So we are done. ☐

I want to note here that filtration is a VERY powerful mathematical idea. Consider the following two examples.

Example 2.25. *Say I think of a random number from 1 to a billion (10^9) in my mind. Your goal is to guess this number. You can only as me yes-or-no questions. How many questions do you need to figure out my number?*

The straight forward way is to guess one by one. Is it 1? Is it 2? This way, you need at most a billion try in the worst case scenario. That is clearly too much.

A better way is the following. You consider a numberical filtration $1, 10, 100, 1000, ..., 10^8, 10^9$. This chain is not by containment, but by divisibility. Looking at this filtration, you see that between each two terms, the quotient is 10. So your strategy is to ask the following: Is the last digit 1? Is the last digit 2? and so on, so in 10 questions you will locate the last digit. This is essentially the gap in the first filtration step from 1 to 10. Then you ask: Is the second to last digit 1? Is the second to last digit 2? and so on, so in 10 questions you will locate the second to last digit. This is essentially the gap in the second filtration step from 10 to 100. Keep doing this, and since your number has at most 9 digits, you can figure int out in at most 90 questions. In fact 80 is enough. As you can see, this strategy is essentially done according to the filtration.

The fastest way is the bisection method. You ask if it is less than half a billion? Suppose yes, the you ask if it is less than 0.25 billion. If no, then you ask if it is less than 0.375 billion, and so on. Each step would half the remaining candidates, and you'll be done in about $\log_2(10^9) \approx 30$ *questions. However, note that this is ALSO a filtration strategy. Our previous filtration strategy allow us to ask one digit as a time in base 10. This new strategy is asking one digit at a time in base 2. So it is in fact also from a filtration* $1, 10, 100, 1000, ...$ *that is expressed in base 2.*

As you can see, a filtration strategy usually makes things exponentially faster.

Example 2.26. *The rubiks cube has more than* 4×10^{19} *configurations. The potential sequences of operations you can do can only be even more. How can you possibly solve it, even with the help of the fastest computers? How many moves do you need to guarantee a solution of the worst case scenario? (This number is also called the god number.)*

If you blindly adopt some algorithms online, your worst case scenario could need hundreds of moves. Previously, no one know how to lower this number much. The first major breakthrough is by Morwan Thistlewaite around 1981, where he reduced this number to 45. This is a drastic improvement. How is this achieved?

Thistlewaite first gives the configurations an algebraic structutre called a group structure. So he has a very big group G_0 *of size more than* 4×10^{19}*. But then he find a filtration of* G_0 *by smaller groups contained in* G_0*, like* $\{solvedstate\} = G_4 \subseteq G_3 \subseteq G_2 \subseteq G_1 \subseteq G_0$*. Even though many of these groups are still big, but the* ***quotient structure*** *between them are small, with a combined size of about a few million, which is much easier to handle. Then Thistlewaite found that you only need 15 steps to go from any position in* G_0 *into the smaller group* G_1*, and the 13 steps to go from any position of* G_1 *into the smaller group* G_2*, then 10 steps into* G_3*, and 7 steps into* G_4 *which only contains the solution as its only element.*

Eventually we see that the god number is 20 (half-turn allowed) or 26 (only quarter turn allowed). But all later progress contain this central ideal of filtration.

2.5 Quotient

As you can see, the power of filtration lies in the "gap", or "quotient", between two adjacent terms. For this reason, we now turn our attention to quotient structure.

Definition 2.27. On a set S, a relation $\sim$ is a subset of $S \times S$. We write $a \sim b$ if $(a, b) \in \sim$. We say that $\sim$ is an equivalence relation if:

1. (Reflexive) $a \sim a$ for all $a \in S$.
2. (Symmetric) $a \sim b$ implies $b \sim a$ for all $a, b \in S$.
3. (Transitive) $a \sim b$ and $b \sim c$ implies $a \sim c$ for any $a, b, c \in S$.

Example 2.28. $=$ *is an equivalence relation on any set. For* $\mathbb{Z}$*, congruence mod n is an equivalence relation. For vector spaces, having the same dimension is an equivalence relation. For dumplings, having the same stuffing is an equivalence relation.*

Definition 2.29. For a set S with an equivalence relation $\sim$, and for any $a \in S$ we can denote $[a] := \{b \in S : b \sim a\}$. This $[a]$ is called the equivalence class of a. Note that maybe $[a] = [b]$, then EVEN if they LOOK different, they are actually the SAME equivalence class. Say if I write $\{[a], [b]\}$, then there is only one element in this set.

Proposition 2.30. *For a set* S *with an equivalence relation* $\sim$*,* $[a] = [b]$ *(as subsets) iff* $a \sim b$*. As a result, different equivalence classes are all disjoint from each other and in fact form a partition of* S*. We denote* $S/\sim = \{[a] : a \in S\}$*, the collection of all equivalence classes. Then there is a unique surjective function* $f : S \to S/\sim$ *such that* $f(a) = [a]$*.*

Proof. Suppose $[a] = [b]$. Then $b \in [a]$, so $a \sim b$. Suppose $a \sim b$. Then $c \in [a]$ iff $c \sim a$ iff $c \sim b$ iff $c \in [b]$. So $[a] = [b]$.

Suppose $[a] \neq [b]$. Suppose for contradiction that $c \in [a] \cap [b]$, then $c \sim a$ and $c \sim b$, so $a \sim b$, and thus $[a] = [b]$, contradiction. So $[a] \cap [b] = \varnothing$.

To show that these equivalence classes form a partition, note that $a \in [a]$ always, so $\bigcup_{a \in S}[a]$ contains all elements of S. So $S \subseteq \bigcup_{a \in S}[a] \subseteq \bigcup_{a \in S} S = S$, so all are equal.

Finally the surjective function portion is trivial. $\square$

Definition 2.31. The function $q : S \to S/\sim$ is called the quotient map of S by $\sim$.

Example 2.32. *Consider any set S with equivalence relation $=$, then each equivalence class contains a single element. $S/=$ is pretty much the same as S, and the quotient function is basically the identity function.*

Consider S the set of all dumplings and $\sim$ the relation of having the same stuffing. Then all dumplings with the same stuffing are combined into a single equivalence class. So $S/\sim$ is the set of all kings of stuffings (egg, pork, spinich, etc.), and the quotient map send each dumpling to its stuffing. When I see a bunch of dumplings, for each dumpling, to determine if I like it or not, I must FIRST consider its stuffing. I like it iff I like its stuffing. So as you can see, I perform my decision by first send each dumpling through the quotient map $q : S \to S/\sim$, and then I check to see if I like its quotient image in $S/\sim$. As you can see, this quotient process is something we do on a daily basis.

Consider $\mathbb{Z}$ with congruence mod 2. Then there are two equivalence classes, the set of all odd numbers and the set of all even numbers. Then $\mathbb{Z}/\mod 2$ or $\mathbb{Z}_2$ has two elements, even and odd. The quotient map send each even number to the value "even", and all odd number to the value "odd". As you probably already know, the even-odd dichotomy is very important. Furthermore, consider the following: from the multiplications of integers and the addition of integers, we see that even+even=odd+odd=even, even+odd=odd, even$\times$even=even$\times$odd=even, odd$\times$odd=odd. So the set $\mathbb{Z}_2$ has a natural addition and multiplication structure. As you can see, quotient is not just a set. It usually inherit whatever structure your pre-quotient set possesses. In this case, for any integer n, $\mathbb{Z}_n$ has an addition and multiplication structure.

As a fun incident, you can verify that when n is a prime, say p, then $\mathbb{Z}_p$ with its addition and multiplication is actually a field. This is the field with p elements for a prime p. We also write it as $\mathbb{F}_p$ sometimes.

Here comes an important observation.

Proposition 2.33. *For a vector space V and any subspace $W \subseteq V$, then the relation $\sim_W$ on V, defined as $a \sim_W b$ iff $a - b \in W$, is an equivalence relation. We sometimes write $V/\sim_W$ simply as V/W.*

Proof. First, clearly $a - a \in W$. Also if $a - b \in W$, then $b - a = -(a - b) \in W$. Finally, if $a - b \in W$ and $b - c \in W$, then $a - c = (a - b) + (b - c) \in W$. So we are done.

For the quotient map, we want to show that $[av + bw] = a[v] + b[w]$. $\square$

As we have noted, quotient usually possesses any pre-quotient structure. So the quotient of a vector space is probably also a vector space.

Proposition 2.34. *On the set V/W, we can try to define $[v] + [w] := \{x + y : x \in [v], y \in [w]\}$ and $k[v] := \{kx : x \in [v]\}$. Then this gives a vector space structure on V/W (over the same field as V), and we have $[v] + [w] = [v + w]$ and $k[v] = [kv]$, and the quotient map $q : V \to V/W$ is linear and surjective with kernel $\mathrm{Ker}(q) = W$.*

Proof. Pick any $x \in [v]$ and $y \in [w]$. Then $x - v, y - w \in W$, so $(x + y) - (v + w) \in W$. So $x + y \in [v + w]$. So $[v] + [w] \subseteq [v + w]$. On the other hand, if $x \in [v + w]$, then $x - v - w \in W$, so $x - v \in [w]$. So $x = v + (x - v)$ where $v \in [v]$ and $x - v \in [w]$. So $x \in [v] + [w]$. In short, $[v] + [w] = [v + w]$.

Similarly, pick any $x \in [v]$, then $x - v \in W$, so $kx - kv \in W$ and $kx \in [kv]$. So $k[v] \subseteq [kv]$. Conversely, if $x - kv \in W$, then $x = k(\frac{1}{k}x)$ where $\frac{1}{k}x - v = \frac{1}{k}(x - kv) \in W$, so $x \in k[v]$. So $k[v] = [kv]$.

Now you can easily verify all rules of a vector space or a linear map. q is surjective because it is a quotient map of some equivalence relation. $\square$

Example 2.35. *Consider $V = \mathbb{R}^2$ with W spanned by the vector $(1,1)^T$. Then elements of V/W are actually lines in $\mathbb{R}^2$ parallel to W. Take a line $(2,3)^T + W$ and $(4,1)^T + W$, then you can verify that if you take any point from one line and any other point from the other, they add up to a point on the line $(5,4)^T + W$.*

Consider $V = \mathbb{R}^3$ and $W = xy$-plane. Then elements of V/W are actually planes parallel to W. You can similarly compute some examples to verify the properties of a quotient space and quotient map. Interestingly, you could see that $\dim(V/W) = \dim V - \dim W$.

As this point, you can happily see that a quotient of V by W is exactly the collection of things in V parallel to W and have the same dimension as W. (Note that they are not necessarily subspaces though. They are affine spaces, if you remember this concept from linear algebra one lecture notes.)

Example 2.36. *Now we do an example to illustrate the next theorem. Let $V = \mathbb{R}^3$ and W be the x-axis. Then V/W is the collection of all lines parallel to the x-axis. We know it is a vector space. What is the origin in the vector space V/W? It is simply $[0] = W$, which corresponds to the x-axis. Now, what is a one-dimensional subspaces in the vector space V/W?*

You pick some vector $[v] \in V/W$. (If you like, take $v = (1,1,1)^T$ to be more concrete.) Then $[v]$ is a line parallel to W through v. What is $k[v]$? Since $k[v] = [kv]$, this is a line parallel to W through kv. In particular, this line is contained in the span of W and v. This span is a plane α (and in fact has equation $y = z$ under our choice of $v = (1,1,1)^T$ and $W = x$-axis). So all multiples of $[v]$ are lines contained in the plane α. Conversely, for any line ℓ contained in α and parallel to W, say $\ell = [u]$, then $u \in V$ must be in the span of W and v. So we must have $u = w + kv$ for some $w \in W$ and $k \in \mathbb{R}$. Then $u \sim kv$ and $[u] = [kv] = k[v]$.

In short, the one-dimensional subspace spanned by $[v]$ in V/W is the collection of all lines parallel to W and inside the plane spanned by W and v. So lines in V/W containing the origin is just planes in V containing W. This is a one-to-one correspondence.

Think about this: points in V/W are lines in V. Lines in V/W are planes in V. It is easy to see the big picture: subspaces of dimension k of V/W are in perfect one-to-one correspondence with subspaces of dimension $k + \dim W$ in V containing W. (And affine subspaces of dimension k of V/W are in perfect one-to-one correspondence with subspaces of dimension $k + \dim W$ in V parallel to W.)

You can also immediately see that $\dim V/W = \dim V - \dim W$.

Proposition 2.37 (Isomorphism Theorems)**.** *Fix any subspace W of a vector space V, then we have the following:*

1. $\dim(V/W) = \dim V - \dim W$.
2. *There is a one-to-one correspondence between subspaces U with $W \subseteq U \subseteq V$ and subspaces U/W of V/W, and $q(U) = U/W$.*

Proof. Consider the rank nullity theorem for the quotient map $q : V \to V/W$. We know q is surjective, so $rank(q) = \dim V/W$, and $\dim \operatorname{Ker}(q) = \dim W$. And we have the desried relation on dimensions.

Say $\dim W = k$. Now V/W is the collection of all k-dim subspaces in V parallel to W, and U/W is the collection of all k-dim subspaces in U parallel to W, so obviously $q(U) = U/W$ is a subspace of V/W. Conversely, for any subspace $T \subseteq V/W$, then $q^{-1}(T)$ is a subspace of V containing $q^{-1}(0) = W$, and you can easily see that $q^{-1}(T)/W = q(q^{-1}(T)) = T$ by surjectivity of q. (Verify this lemma yourself if you don't know it before: for any surjective function $f : S \to T$ and any subset $P \subseteq T$, then $f \circ f^{-1}(P) = P$.) $\square$

Example 2.38. *Next we try for an example for the universal property of quotients. Again $V = \mathbb{R}^3$ and $W = x$-axis.*

Consider the linear map $A : V \to \mathbb{R}$ with matrix $A = (0, 1, -1)$. So $A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = y - z$. As you can easily verify, $W \subseteq \operatorname{Ker}(A)$.

Now take any $\ell \in V/W$. Then A maps it to a subset of $\mathbb{R}$. But for any $v, w \in \ell$, then by definition we have $v - w \in W \subseteq \operatorname{Ker}(A)$, so $A(v) - A(w) = A(v - w) = 0$. So everything in ℓ are going to the SAME image under the map A. (Take ℓ to be spanned by W and $(1,1,1)^T$ for numerical example.)

So even though elements of V/W are subsets of V, nevertheless A maps each subset to a single image. We might as well define a map $A' : V/W \to U$ for this induced map. Then you can easily check that $A'[v] = Av$.

Proposition 2.39 (Universal Property)**.** *For any $A : V \to U$, if $A(W) = \{0\}$, then we have a decomposition $A = A' \circ q$ where $q : V \to V/W$ is the quotient map, and $A' : V/W \to U$ is unique. In particular, we have $A'[v] = Av$. Furthermore,* $\mathrm{Ker}(A') = \mathrm{Ker}(A)/W$.

Proof. For any equivalence class $c \in V/W$, then c is a subset of V, and A maps this subset of V into a subset $A(c)$ of U. I claim that $A(c)$ has only one element. Then we could simply define this element to be $A'(c)$.

To see this, suppose $v, w \in c$. Then $A(v) - A(w) = A(v - w) = 0$ since $v - w \in W \subseteq \mathrm{Ker}(A)$. So indeed $A(v) = A(w)$ for any $v, w \in c$. We also see that $A'[v] = A'c = Av$.

Finally, $A'[v] = 0$ iff $Av = 0$, so we have $\mathrm{Ker}(A') = \mathrm{Ker}(A)/W$. □

Example 2.40. *Suppose $A \in M_{m\times n}$ has rank r. Then $A = BC$ where $B \in M_{m\times r}$ is injective and $C \in M_{r\times n}$ is surjective. How to find this decomposition?*

Consider any linear map $A : V \to W$. Then we immediately have a decomposition of A into $q : V \to V/\mathrm{Ker}(A)$ and $A' : V/\mathrm{Ker}(A) \to W$. Then q is surjective and A' is has kernel $\mathrm{Ker}(A)/\mathrm{Ker}(A) = \{0\}$, so it is injective. Then $A = A' \circ q$ is the desired decomposition.

Corollary 2.41. *For any linear map $A : V \to V$, suppose we have an invariant subspace $W \subseteq V$ for A. Then there is an induced linear map $A/W : V/W \to V/W$ such that $A/W[v] = [Av]$.*

Proof. Consider the map $A : V \to V$ followed by the quotient map $q : V \to V/W$. Then we have a linear map $q \circ A : V \to V/W$. Further more, $q \circ A(W) = q(W) = \{0\}$, so W is in the kernal of this composition. So we have an induced linear map $A/W = (q \circ A)' : V/W \to V/W$ such that $A/W[v] = q \circ A(v) = [Av]$. □

Definition 2.42. The above map A/W is the induced linear map of A (for the invariant subspace W).

Remark 2.43. *Intuitively, A/W is about the overarching structure of A ABOVE W. You perforem A while ignoring everything related to W. Think about it like this. If V is your mother and your girlfriend, both fell into the water, then ideally you save both of them. Say A means saving both of them. Now, if W is the one that knows how to swim (so V/W is the other one that doesn't know how to swim), then you can ignore W, and only perform A/W on V/W, i.e., saving the one that doesn't know how to swim.*

A very common practical strategy is this. Say you are building a mathematical model for something, and you choose to use vector spaces as a model. However, you don't know what information are relevant and what is useless. Then one way is to first put ALL data together and call it V. Then later, if you find out that all useless data form a subspace W, then simply quotient out W everywhere to simplify all expressions of your result.

Intuitively, this is how one might study a linear map A: we find an invariant subspace W. Then $A(W) \subseteq W$, so in fact there is a linear map $A|_W : W \to W$ by mearly restricting A to W. We can also study the induced map A/W. Then $A|_W$ is the action of A INSIDE of W, while A/W is the action of A OUTSIDE of W. Are these enough to recover A? Unfortunately, the answer is no.

Proposition 2.44. *For any vector space V of dim n with a subspace W of dim m, we can pick a basis $v_1, ..., v_n$ for V such that $v_1, ..., v_m$ span W. Then $[v_{m+1}], ..., [v_n]$ would span V/W. And for any linear map A with invariant subspace W, then under the basis $v_1, ..., v_n$, it is $\begin{pmatrix} A|_W & B \\ 0 & A/W \end{pmatrix}$. Here $A|_W$ is the matrix of the restriction of A to W, i.e., $A|_W : W \to W$ under the basis $v_1, ..., v_m$; and A/W is expressed under the basis $[v_{m+1}], ..., [v_n]$; and B is some matrix that we don't know much about.*

Proof. Express everything under the basis $v_1, ..., v_n$. Consider Av where $v = (a_1, ..., a_n)^T$. Then we perform a decomposition $v = v_1 + v_2$ where $v_1 = (a_1, ..., a_m, 0, ..., 0)$. Clearly $Av = Av_1 + Av_2$ and $Av_1 = A|_W v_1$.

Now $[Av_2] = A/W[v_2]$, so they are the same affine subspace parallel to W. So the last $(n-m)$-coordinates of Av_2 is the same as the coordinates of $A/W[v_2]$. So we are done. □

Remark 2.45. *As you can see, quotient spaces and triangular structures are intimately related. If you have a spatial decomposition into invariant subspaces $V = W_1 \oplus W_2$, then A becomes block diagonal. The linear maps $A|_{W_1}$ and $A|_{W_2}$ is enough to obtain the complete information about A.*

But if this is not possible, if you can only find W_1 but CANNOT find W_2, then the quotient structure is the best you can do. Instead of W_1 and W_2, you now have W_1 and V/W_1, and now you have linear maps $A|_{W_1}$ and A/W_1. Furthermore, A is now block upper-triangular, and the two linear maps is NOT engouh to get all the information about A, since $A = \begin{pmatrix} A|_{W_1} & B \\ 0 & A/W_1 \end{pmatrix}$ and we have no clue what B is.

Think about case of $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ as a numerical illustration of above points. $V = \mathbb{R}^2$ and W is x-axis. A fixes everything on W, and A also maps each line parallel to W to itself (since A is a shearing). So both $A|_W$ and A/W are the 1 by 1 matrix (1). However, the two ones are not enough to recover A.

So we see that $A|_W$ and A/W are NOT enough to recover A. Nevertheless, they are still "enough" for the following purpose.

Corollary 2.46. *Let the eigenvalues of $A|_W$ be $\lambda_1, ..., \lambda_m$ and the eigenvalues of A/W be $\lambda_{m+1}, ..., \lambda_n$, both counting algebraic multiplicities. Then the eigenvalues of A are exactly $\lambda_1, ..., \lambda_n$.*

Proof. Easy to see from the upper-triangular block structure of A. □

Now we move on to the triangularization of a matrix.

Lemma 2.47. *Any square matrix A has at least one eigenvector.*

Proof. By the fundamental theorem of algebra, the characteristic polynomial $\det(xI - A)$ always have at least one root, so A has at least one eigenvalue. But an eigenvalue is defined to have at least one corresponding eigenvectors. So A has at least one eigenvector. □

Proposition 2.48. *For any linear map $A : V \to V$, there is a basis of V underwhich A is triangular.*

Proof. Pick an eigenvector v_1 of A, and let W_1 be the span of v_1. Clearly $A(W_1) \subseteq W_1$.

Now consider $A/W_1 : V/W_1 \to V/W_1$. This linear map must also have an eigenvector. Pick any $v_2 \in V$ such that $[v_2]$ is an eigenvector for A/W_1. Since $[v_2] \neq [0]$, v_1 and v_2 are linearly independent. Let W_2 be the span of v_1, v_2. I claim that $A(W_2) \subseteq W_2$. Trust me for now about this claim.

Then consider $A/W_2 : V/W_2 \to V/W_2$. This linear map must also have an eigenvector. Pick any $v_3 \in V$ such that $[v_3]$ is an eigenvector for A/W_2. Since $[v_3] \neq [0]$, v_1, v_2 and v_3 are linearly independent. Let W_3 be the span of v_1, v_2, v_3. I claim that $A(W_3) \subseteq W_3$. Trust me for now about this claim.

You can probably see this pattern now. We repeat this process untill we obtain $W_n = V$. In the end, we would construct linearly independent $v_1, ..., v_n \in V$ and subspaces $\{0\} \subsetneq W_1 \subsetneq W_2 \subsetneq ... \subsetneq W_n = V$. This is a filtration of V by invariant subspaces. As a result, the matrix for A is upper triangular under the basis $v_1, ..., v_n$.

It remains to check my claims that $A(W_i) \subseteq W_i$ for each i. Our construction is inheritly inductive, so obviously this should by done by induction. $A(W_1) \subseteq W_1$ is obvious. Suppose we already have $A(W_{i-1}) \subseteq W_{i-1}$. Recall that v_i is chosen so that $[v_i]$ is an eigenvector of A/W_{i-1}, say for some eigenvalue λ. Then $[Av_i] = A/W_{i-1}[v_i] = \lambda[v_i] = [\lambda v_i]$. So $Av_i - \lambda v_i \in W_{i-1}$. So Av_i is in the span of v_i and W_{i-1}, which is W_i. Now for any vector in W_i, it must be $av_i + w$ for some $w \in W_{i-1}$, so $A(av_i + w) = aAv_i + Aw$. Since $aAv_i \in W_i$ and $Aw \in W_{i-1} \subseteq W_i$, we see that $A(av_i + w) \in W_i$. So $A(W_i) \subseteq W_i$. □

Corollary 2.49. *Any matrix is similar to an upper triangular matrix.*

Proof. A matrix is a linear map. □

Corollary 2.50. *Let $\lambda_1, ..., \lambda_n$ be eigenvalues of A counting algebraic multiplicity, ordered in an arbitrary way. Then there is an upper triangular matrix U similar to A whose (i, i)-entry is λ_i.*

Proof. When you construct the filtration, you could first pick v_1 to be an eigenvector for λ_1, then when you pick v_2, pick it so that $[v_2]$ is an eigenvector for λ_2, and so on. □

So not only we can ALWAYS triangularize ANY square matrices, we can even re-order the diagonal terms however we want.

Example 2.51. *As an illustrative example, consider* $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{pmatrix}$. *Can we change a basis so that it is similar to an upper triangular matrix with diagonal entries* $1, 2, 1$*? Our results above says that we can.*

First, $v_1 = (1, 0, 0)^T$ *is an eigenvector for the eigenvalue 1, so we still use it to construct* W_1*. Now* V/W_1 *means we are ignoring the first coordinates, and each element of* V/W_1 *can be thought of as some line* $(*, b, c)^T$ *for fixed* $b, c \in \mathbb{C}$*. Now* $A(*, b, c)^T = (*, b + 4c, 2c)^T$*, so* A/W_1 *is the matrix* $\begin{pmatrix} 1 & 4 \\ 0 & 2 \end{pmatrix}$ *(exactly as the lower right 2 by 2 block of the original A). We want to find an eigenvector of it for the eigenvalue 2, so we must have* $b + 4c = 2b$*, thus the eigenvector is* $(*, 4, 1)^T$*. Pick* $v_2 = (0, 4, 1)^T$*, and construct* W_2 *as the span of* v_1, v_2*. Finally,* A/W_2 *is one dimensional with its only eigenvalue 1, so any vector in* V/W_2 *is an eigenvector for the eigenvalue 1. Pick any* v_3 *outside of* W_2*, say* $v_3 = (0, 1, 0)^T$*.*

Then under the basis $(v_1, v_2, v_2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$*, our matrix should again be upper triangular with diagonal entries 1,2,1. Let's check this, we have:*

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 11 & 2 \\ 0 & 8 & 1 \\ 0 & 2 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -4 \end{pmatrix} \begin{pmatrix} 1 & 11 & 2 \\ 0 & 8 & 1 \\ 0 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 11 & 2 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

2.6 Jordan Normal Form (Jordan Canonical Form)

Spatial decomposition is the best, but not always possible. Spatial filtration is always achievable, but sometimes it is not enough. Can we make a compromise between the two? The answer is the following: we should first perforem spatial decomposition as much as possible, and then for those that cannot be decomposed, we do a filtration. In the end, we should have the following theorem.

Definition 2.52. A Jordan block of size k for an eigenvalue λ is the $k \times k$ matrix $\begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}$.

Here all empty entries are 0.

Theorem 2.53 (Jordan normal form)**.** *Any square matrix A is similar to a block diagonal matrix* $diag(J_1, J_2, ..., J_a)$ *where all these* J_i *are Jordan blocks. This is called the Jordan normal form of this matrix, and it is unique up to the permutation of these Jordan blocks.*

We now present a first proof of this theorem. The idea is this: we decompose V into invariant subspaces as much as possible. Then we realize that, if an invariant subspace cannote be decomposed further, then $A = aI + N$ for some constant $a \in \mathbb{C}$ and some nilpotent matrix N with maximal nilpotent degree. Then the N is similiar to a nilpotent Jordan block, and thus $aI + N$ is a Jordan block as well.

The first step is to do a spatial decomposition according to eigenvalues.

Lemma 2.54 (Weak Cayley Theorem)**.** *For any square matrix $A \in M_n$, if $p_A(x)$ is its characteristic polynomial, then $p_A(A)^n = 0$. (Note that later we shall actually prove that $p_A(A) = 0$ always.)*

Proof. Assume that we have picked a nice basis so that A is upper triangular with diagonal entries $\lambda_1, ..., \lambda_n$. Then these are eigenvalues of A. So $p_A(x) = \prod(x - \lambda_i)$ and $p_A(A) = \prod(A - \lambda_i I)$. Note that all factor matrices here are upper triangular.

Now consider $A - \lambda_i I$. Its (i,i) entry must be 0. Recall that for upper trianguler matrices U_1, U_2, then the diagonal entries of $U_1 U_2$ is exactly the product of the corresponding diagonal entries of U_1 and of U_2. So in particular, since $A - \lambda_i I$ is a factor of $p_A(A)$, the (i,i) entry of $p_A(A)$ is zero. This is true for all i, so $p_A(A)$ has only zeroes on its diagonal. Since $p_A(A)$ is also upper triangular, all its eigenvalues are zero, so it is nilpotent. So $p_A(A)^n = 0$. ☐

Lemma 2.55. *For any square matrices $A \in M_m$ and $B \in M_n$ with no common eigenvalues, and any matrix $C \in M_{m\times n}$, there is a unique matrix $X \in M_{m\times n}$ such that $XB - AX = C$.*

Proof. Consider the function $L : M_{m\times n} \to M_{m\times n}$ such that $L(X) = XB - AX$. Then we see that $L(X+Y) = L(X) + L(Y)$ and $L(kX) = kL(X)$. So L is in fact a LINEAR map! By rank-nullity, it is enough to show that $\mathrm{Ker}(L) = 0$.

Suppose $XB - AX = 0$. Our goal is to prove that $X = 0$, but how? Well, think about this. Since $AX = XB$, you can check that $A^2X = AXB = XB^2$, and similarly $A^3X = XB^3$ and so on. In fact, it is very easy to see that $p(A)X = Xp(B)$ for any polynomial p.

Then take $p(x) = p_A(x)^n$. Then $0 = p(A)X = Xp(B)$. On the other hand, $p(B) = \prod(B - \lambda_i I)^n$ where $\lambda_1, ..., \lambda_m$ are eigenvalues of A. Since none of them are eigenvalues of B, it seems that $(B - \lambda_i I)$ are all invertible, so $p(B)$ is invertible. So $Xp(B) = 0$ implies that $X = 0$. So we are done. ☐

Proposition 2.56. *For a square matrix $A \in M_n$, if some eigenvalues of A are different from the others, then there is a non-trivial spatial decomposition of $\mathbb{R}^n$ into invariant subspaces of A. (Note that we say non-trivial, because there is always a trivial decomposition $V = V \oplus \{0\}$ into invariant subspaces. And that does not count.)*

Proof. Let a be an eigenvalue of A with algebraic multiplicity g. Let the eigenvalues of A be $\lambda_1, ..., \lambda_g, \lambda_{g+1}, ..., \lambda_n$ counting algebraic multiplicity, such that $\lambda_1 = ... = \lambda_g = a$ and no other eigenvalues is equal to a. Then under the right basis, A is upper triangular with diagonal entries $\lambda_1, ..., \lambda_n$. In partcular, we have $\begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}$ where all eigenvalues of A_1 are a, and NO eigenvalues of A_2 is a. So there exist a UNIQUE X such that $XA_2 - A_1X = -B$.

Consider $\begin{pmatrix} I & X \\ 0 & I \end{pmatrix}\begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}\begin{pmatrix} I & -X \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_1 & B + XA_2 - A_1X \\ 0 & A_2 \end{pmatrix} = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$. Therefore A is similar to $\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$. This is a block diagonal form, so it corresponds to a decomposition of $\mathbb{R}^n$ into invariant subspaces of A. ☐

Corollary 2.57. *Suppose for a linear map $A : V \to V$, there is NO non-trivial way to decompose V into invariant subspaces, then all eigenvalues of A are the same. For any invariant subspace W, then all eigenvalues of $A|_W$ and A/W are the same as the eigenvalue of A.*

Lemma 2.58. *Suppose for a linear map $A : V \to V$, there is NO non-trivial way to decompose V into invariant subspaces. Suppose all eigenvalues of A are k, then $A = kI + N$ for some nilpotent matrix N.*

Proof. Pick any filtration by invariant subspaces $\{0\} = W_0 \subsetneq W_1 \subsetneq W_2 \subsetneq ... \subsetneq W_n = V$. I claim that $A - kI$ maps W_i to W_{i-1}, so $(A - kI)^n$ would maps V to $W_0 = \{0\}$. So $(A - kI)^n = 0$, and $A - kI$ is nilpotent.

To prove my claim, note that all eigenvalues of $A|_{W_i}$ are k, and therefore all eigenvalues of $A|_{W_i}/W_{i-1} : W_i/W_{i-1} \to W_i/W_{i-1}$ are k. However, note that W_i/W_{i-1} is one-dimensional, so $A|_{W_i}/W_{i-1}$ is a one by one matrix, and it must be (k). So it simply multply all vectors of W_i/W_{i-1} by k.

So pick any $v \in W_i$. Then $[Av] = A/W_{i-1}[v] = A|_{W_i}/W_{i-1}[v] = k[v] = [kv]$. So $Av - kv \in W_{i-1}$. Since this is true for all $v \in W_i$, we see that $A - kI$ maps W_i to W_{i-1}. ☐

Corollary 2.59. *Suppose for a linear map $A: V \to V$, there is NO non-trivial way to decompose V into invariant subspaces. The A is similar to a Jordan block.*

Proof. Let $A = kI + N$ for some nilpotent matrix N. Let $N = BJB^{-1}$ for its nilpotent Jordan normal form J. Then $A = B(kI + J)B^{-1}$. So up to similarity, we can pretend that we picked a nice basis and pretend that $A = kI + J$.

Now if J is composed of multiple Jordan blocks, then J is block diagonal with several Jordan blocks on its diagonal. Then so is $kI + J = A$. Then there is a corresponding decomposition of V into invariant subspaces. So J is a single Jordan block, and thus A is a Jordan block. ☐

So now the theorem of Jordan normal form is proven. We simply decompose V into invaraint subspaces as much as possible, then on each "minimal" invariant subspaces that cannot be decomposed further, then A must behaves as a Jordan block. So A is block diagonal with various Jordan blocks on its diagonal.

Example 2.60. *Consider $A = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 2 \\ 3 & 0 & 1 \end{pmatrix}$. Then $\det(xI - A) = \det \begin{pmatrix} x-2 & 0 & 0 \\ 1 & x-1 & -2 \\ -3 & 0 & x-1 \end{pmatrix} = (x - 2)\det \begin{pmatrix} x-1 & -2 \\ 0 & x-1 \end{pmatrix} = (x-2)(x-1)^2$. So it has eigenvalue 1 with algebraic multiplicity 2 and eigenvalue 2 with algebraic multiplicity 1.*

Let us first try to put it in upper triangular form. When we do this, we want to keep in mind to group the eigenvalues of the same value together. So say we require the resulting upper triangular matrix to have diagonal 1,1,2. Then first we need a vector v_1 for eigenvalue 1, say $v_1 = (0,1,0)^T$ that spans a subspace W_1. Next we look at V/W_1, whose vector has the form $(a,*,b)^T$ and $A(a,*,b)^T = (2a,*,3a+b)^T$. So it seems that $(0,*,1)^T$ or $[(0,0,1)]$ is an eigenvector for A/W_1. So let $v_2 = (0,0,1)^T$ and W_2 be the span of v_1, v_2. Finally, V/W_2 is one dimensional so $A/W_2 = (2)$, so pick any v_3 outside of W_3, say $v_3 = (1,0,0)^T$, and we are done.

Under the basis v_1, v_2, v_3, we have A similar to $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 2 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \left(\begin{array}{cc|c} 1 & 2 & -1 \\ 0 & 1 & 3 \\ \hline 0 & 0 & 2 \end{array}\right) = \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}$ with $A_1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $A_2 = (2)$ and $B = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$.

Now, since A_1 and A_2 has NO eigenvalue in common, we know that there is a unique $X \in M_{2\times 1}$ such that $A_2X - XA_1 = -B$. Then A is similar to $\begin{pmatrix} I & X \\ 0 & I \end{pmatrix}\begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}\begin{pmatrix} I & -X \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$.

To be more explicit, if $X = \begin{pmatrix} x \\ y \end{pmatrix}$, then $XA_2 - A_1X = -B$ would translate into $\begin{pmatrix} 2x \\ 2y \end{pmatrix} - \begin{pmatrix} x+2y \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$, which means that $x = -5$ and $y = -3$. Then A is similar to $\left(\begin{array}{cc|c} 1 & 0 & -5 \\ 0 & 1 & -3 \\ \hline 0 & 0 & 1 \end{array}\right) \left(\begin{array}{cc|c} 1 & 2 & -1 \\ 0 & 1 & 3 \\ \hline 0 & 0 & 2 \end{array}\right) \left(\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & 3 \\ \hline 0 & 0 & 1 \end{array}\right) = \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 2 \end{array}\right)$. This corresponds to a spatial decomposition of V into invariant subspaces.

Finally, $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is a Jordan block, and $A_2 = (2)$ is already a Jordan block. So A is similar to $\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 2 & 0 \\ \hline 0 & 0 & 1 \end{array}\right) \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 2 \end{array}\right) \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ \hline 0 & 0 & 1 \end{array}\right) = \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 2 \end{array}\right)$ is block diagonal with Jordan blocks on the diagonal.

Remark 2.61. *The proof here is nice and simple, but INEFFECTIVE. It does not give us any specific way to CONSTRUCT the Jordan decomposition $A = BJB^{-1}$. (We merely proved the existence of X to solve $A_2X - XA_1 = -B$, but we did not show how to find this X.)*

This is very annoying. An ineffective theorem is a lot easier to prove than the effective version. Consider the following theorem: There is a positive even integer $n \in \mathbb{Z}$ such that Goldbach's conjecture (i.e., any even integer at least 4 is the sum of two primes) is true iff it is true for all integers less than n.

Intuitively, it seems that this is a VERY powerful theorem. Once you established this theorem, you merely need to check Goldbach's conjecture for the finitely many cases $4, 6, 8, ..., n$ and then you will be able to prove or disprove Goldbach's conjecture!

And this theorem is indeed true, BUT it is in fact useless, because it is INEFFECTIVE. Let's see a proof. The proof goes like this. Suppose Goldbach's conjecture is true, then set $n = 4$ and we are done. Suppose Goldbach's conjecture is false, let k be the smallest counter example, and set $n = k + 2$ and we are done.

You can see that the proof is extremely trivial, and you can imagine that such a trivial theorem must actually be completely useless.

The effective version of this theorem should tell us EXPLICITLY what is n. THEN this theorem would be very powerful, as our intuition first suggested.

2.7 Generalized Eigenstuff

In the last session, we proved the Jordan normal form by first performing a decomposition by grouping eigenvalues of the same value together. The key lies in solving an equation $A_2X - XA_1 = -B$, and this is where our proof becomes ineffective. We merely proved the existance of a unique solution, but failed to actually provide the solution.

In this subsection, we will essentially use a new method to achieve Corollary 2.57. The rest would proceed the same way as before. This way we circumvented the ineffective portion, and the resulting proof would be effective. As you will see, this needs us to TRULY understand Jordan normal form and the associated decomposition. It starts with the following question: What should our space decompose into?

Definition 2.62. We say that v is a generalized eigenvector of A for eigenvalue λ if $(A-\lambda I)^k v = 0$ for some integer k. The generalized eigenspace for λ is the subspace V_λ formed by all generalized eigenvectors for λ and the zero vector.

Proposition 2.63 (Geometric meaning of algebraic multiplicity)**.** *For any linear map $A : V \to V$, then $\dim V_\lambda$ is the algebraic multiplicity of the eigenvalue λ.*

Proof. Suppose λ has algebraic multiplicity g.

Say A has eigenvalues $\lambda_1, ..., \lambda_n$ such that $\lambda_1 = ... = \lambda_g = \lambda$. Then we can find a basis $v_1, ..., v_n$ such that A under this basis is upper triangular with diagonal entries $\lambda_1, ..., \lambda_n$. Then $A = \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}$ where all eigenvalues of A_1 are λ, and NO eigenvalue of A_2 is λ, and both A_1, A_2 are upper triangular. In particular, $(A-\lambda I)^k = \begin{pmatrix} 0 & B' \\ 0 & (A_2-\lambda I)^k \end{pmatrix}$ for all $k \geq g$. Since NO eigenvalue of A_2 is λ, therefore $(A_2-\lambda I)^k$ is invertible, so $(A-\lambda I)^k$ has rank $n-g$ for all $k \geq g$. So $\mathrm{Ker}(A-\lambda I)^k$ is g-dimensional for all $k \geq g$. Since $\mathrm{Ker}(A-\lambda I)^k$ is a non-decreasing chain of subspaces, this means it equal to $\mathrm{Ker}(A-\lambda I)^g$ for all $k \geq g$.

So $V_\lambda = \bigcup \mathrm{Ker}(A-\lambda I)^k = \mathrm{Ker}(A-\lambda I)^g$ and it is g dimensional. ☐

Lemma 2.64. *Let $V_1, V_2, ..., V_k$ be the generalized eigenspaces for different eigenvalues $\lambda_1, ..., \lambda_k$ of A. Then these subspaces are all linearly independent.*

Proof. Let their algebraic multiplicities be $g_1, ..., g_k$.

As usual, pick any non-zero vectors $v_1, ..., v_k$ such that $v_i \in V_i$. Suppose some linear combination $\sum a_i v_i = 0$.

For each i, construct a polynomial $p_i(x) = \frac{\prod_j (x-\lambda_j)^{g_j}}{(x-\lambda_i)^{g_i}}$. Then You can check that $p_i(A)$ kills all V_j with $j \neq i$ but restricted to a bijection on V_i. So apply $p_i(A)$ to both sides of $\sum a_j v_j = 0$, we see that $a_i p_i(A) v_i = 0$ and thus $a_i = 0$. Since this is true for each i, we see that $v_1, ..., v_k$ are linearly independent. ☐

Corollary 2.65. *Say A has eigenvalues $\lambda_1, ..., \lambda_k$ NOT counting algebraic multiplicity with corresponding generalized eigenspaces $V_1, ..., V_k$, then $V = \bigoplus V_i$ is a decomposition of V into invariant subspaces. So A is block diagonal with blocks $A|_{V_i}$, where each block $A|_{V_i}$ only has eigenvalue λ_i.*

Proof. The algebraic multiplicities always add up to n. □

Now we hae prove Corollary 2.57 again, and the rest is identical to the last proof.

Example 2.66. *Again consider $A = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 2 \\ 3 & 0 & 1 \end{pmatrix}$. We know its eigenvalues are 1,1,2. So it must has a generalized eigenspace V_1 for the eigenvalue 1 of dimension 2 and a generalized eigenspace V_2 for the eigenvalue 2 of dimension 1.*

What is V_1? It is $\operatorname{Ker}(A-I)^2 = \operatorname{Ker}\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 2 \\ 3 & 0 & 0 \end{pmatrix}^2 = \operatorname{Ker}\begin{pmatrix} 1 & 0 & 0 \\ 5 & 0 & 0 \\ 3 & 0 & 0 \end{pmatrix}$*, which is spanned by* $(0,1,0)^T, (0,0,1)^T$.

And restricted to this subspace V_1, under this basis, we have $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, and $A - I = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$ in indeed nilpotent. Now our theorem on nilpotent Jordan normal form tells us that we could pick basis $v_1 = (A-I)v_2$ and $v_2 = (0,0,1)^T$ as the right basis for V_1.

What is V_2? It is $\operatorname{Ker}(A - 2I) = \operatorname{Ker}\begin{pmatrix} 0 & 0 & 0 \\ -1 & -1 & 2 \\ 3 & 0 & -1 \end{pmatrix}$ *which is spanned by $v_3 = (1,5,3)^T$. Obviously A restricted to V_2 is just (2) and there is nothing to do here.*

So the best basis for V should be $(v_1, v_2, v_3) = \begin{pmatrix} 0 & 0 & 1 \\ 2 & 0 & 5 \\ 0 & 1 & 3 \end{pmatrix}$. And under this basis, the new matrix for A should be $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Let us check this. Indeed, we have:

$$\begin{pmatrix} 0 & 0 & 1 \\ 2 & 0 & 5 \\ 0 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \\ 2 & 0 & 5 \\ 0 & 1 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 0 & 2 \\ 2 & 2 & 10 \\ 0 & 1 & 6 \end{pmatrix}\begin{pmatrix} -5/2 & 1/2 & 0 \\ -3 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 2 \\ 3 & 0 & 1 \end{pmatrix} = A$$

Example 2.67. *Let us have a more complex example. Say $A = \begin{pmatrix} -10 & 9 & -7 & -1 & 7 \\ -17 & 13 & -9 & -2 & 12 \\ -14 & 9 & -6 & -1 & 10 \\ -13 & 9 & -7 & 0 & 9 \\ -12 & 9 & -7 & -1 & 9 \end{pmatrix}$. You can find its characteristic polynomial and check that its eigenvalues are 1,1,1,1,2.*

The eigenvalue 2 is simple. It has algebraic and geometric multiplicity 1, and you can find its corresponding eigenvector is $v_5 = (5,9,8,7,6)^T$.

For the eigenvalue 1, consider $A-I = \begin{pmatrix} -11 & 9 & -7 & -1 & 7 \\ -17 & 12 & -9 & -2 & 12 \\ -14 & 9 & -7 & -1 & 10 \\ -13 & 9 & -7 & -1 & 9 \\ -12 & 9 & -7 & -1 & 8 \end{pmatrix}$. You can check that $\dim\operatorname{Ker}(A-I) = 2, \dim\operatorname{Ker}(A-I)^2 = 3, \dim\operatorname{Ker}(A-I)^3 = 4$*, and we don't need to continue once we reach dimension 4, because 1 only has algebraic multiplicity 4.*

Pick any $v_3 \in \operatorname{Ker}(A-I)^3 - \operatorname{Ker}(A-I)^2$, and set $v_2 = (A-I)v_3$ and $v_1 = (A-I)v_2$, and find any v_4 such that v_1, v_4 span $\operatorname{Ker}(A-I)$. One possible choice is $v_3 = (1,2,2,1,1)^T$, then $v_2 = (-1,-1,-1,-1,-1)^T$, *and then* $v_1 = (3,4,3,3,3)^T$*. Then you can pick say* $v_4 = (1,3,3,2,1)$*. Note that since* v_3 *goes to* v_2*, which goes to* v_1*, and* v_4 *stands alone, therefore the corresponding nilpotent Jordan block is* $\left(\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right)$.

So under the basis $B = (v_1, v_2, v_3, v_4, v_5)$*, we have* A *in Jordan normal form* $J = \left(\begin{array}{ccc|c|c} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 2 \end{array}\right)$*. In particular,* $A = BJB^{-1}$*.*

2.8 Understanding Jordan Blocks

Recall that if a matrix A is block diagonal, say $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$, then essentially we can just decompose the space into smaller invariant spaces, on which A would act in a easier way. In short, you can think of A as a "sum" of A_1 and A_2, or that the action of A is simply glueing the action of A_1, A_2 together. It is enough to just understand A_1, A_2.

Then Jordan normal form tells us that to understand ANY matrix, we just need to understand Jordan blocks, because up to a change of basis, Jordan blocks make up all matrices. But how do Jordan blocks behave?

Algebraically, we know that nilpotent Jordan blocks act on column vectors by shifting coordinates up. And geometrically they are rotation plus a projection, a combination that folds the space into smaller and smaller dimensions. Let us look at some other Jordan blocks here.

Example 2.68. $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ *is the standard shearing. In general, consider* $E = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$*. It sends rectangles, with sides parallel to the coordinate-lines, into parallelograms of the same height. Draw a few graphic examples and shapes to see this better. And precisely because it is a shearing, you can see that* $\det(EA) = \det(A)$*, i.e., shearing preserves area in dimension two.*

If you repeatedly apply $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ *to a vector, say* $(0,1)^T$*, you get* $(1,1)^T, (2,1)^T, (3,1)^T$*, and so on. Basically the second coordinates are always the same, while the first coordinate keep progressing. The so the orbits of* A *are lines parallel to the* x*-axis.*

Example 2.69. *Now consider* $J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$*. It sends* $(0,0,1)^T$ *to* $(0,1,1)^T$*, then to* $(1,2,1)^T$*, then to* $(3,3,1)^T$*, then to* $(6,4,1)^T$*, and so on. This is EXACTLY the left three entries of the Pascal's triangule (Yang Hui triangle, or binomial coefficients, etc.)! So to see* $J^k(0,0,1)^T$*, you ca imagine that you are doing* $(x+1)^k$*, and read out the last three coefficients. You can also see that* $J^k(0,1,0)^T = (k,1,0)^T$*, which is basically the last three coefficients of* $x(x+1)^k$*. In general,* $J^k(a,b,c)^T$ *is the last three coefficients of* $(ax^2+bx+c)(x+1)^k$*. Funny, no?*

Is this really true? Well, let P_2 *be set of polynomials mod* x^3*. I.e., we consider two polynomials to be the same as long as their last three coordinates are identical. Consider the linear map* $L : \mathbb{R}^3 \to P_2$ *that sends* $(a,b,c)^T$ *to* ax^2+bx+c*. This is a linear isomorphism, so we can think of the two as the same vector space. Then how does* J *behaves on* P_2*? It sends* 1 *to* $x+1$*, and* x *to* x^2+x*, and* x^2 *to* x^2*, which is the same as* x^3+x^2 *mod* x^3*. So* J *behaves exactly by multiplying polynomials by* $(x+1)$*. So* $J^k(ax^2+bx+c) = (ax^2+bx+c)(x+1)^k \pmod{x}^3$*.*

This algebraic picture can be generalized to Jordan blocks with eigenvalue 1 of arbitrary size.

Example 2.70. *What is the geometric behavior of* $J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$*? Say what are its orbits (smooth curves C such that J always maps each point in C back to some point in C)?*

Well, in general, $(a, b, c)^T$ would goes to $(a + b, b + c, c)^T$, and then to $(a + b + b + c, b + c + c, c)^T$, and then to $(a + b + b + c + b + c + c, b + c + c + c, c)^T$ and so on. So after k steps, J^k would maps it to $(a + kb + (0 + 1 + ... + (k - 1))c, b + kc, c)^T = (a + kb + \frac{1}{2}(k^2 - k)c, b + kc, c)^T$.

So generically, the orbits are $p(t) = (\frac{c}{2}t^2 + (b - \frac{1}{2}c)t + a, ct + b, c)^T$. As you can see, the third coordinate never change, so the orbit curves stays on a plane (parallel to the xy-plane). On this plane, the first coordintae is in fact a degree two polynomial of the second coordinate. So on this plane, we would actually see a graph of a parabola. So orbits of J are various parabolas parallel to the xy-plane.

Note that for each parabola on a plane $z = c \neq 0$, when $t = -\frac{b}{c}$, then the parabola would go through the xz-plane. So if you want to find all parabolas on the plane $z = c$, then they are $p(t) = (\frac{c}{2}t^2 - \frac{c}{2}t + a, ct, c)^T$, or the parabola $p(t) = (\frac{c}{2}t^2 - \frac{c}{2}t, ct, c)^T$ shifted along the x-axis. Furthermore, since we only care about the curve, not how it is parametrized, we can further more substitute t by t/c. Then we have $p(t) = (\frac{1}{2c}t^2 - \frac{1}{2}t, t, c)^T$ shifted along the x-axis.

I highly recommand you to draw these parabolas on $z = 1, z = 2, z = -1$ to see what would happen. Also fell free to draw the picture on the plane $z = 0$, and see why this is the limiting case for $z > 0$ and $z < 0$.

If you want to see the geometric behavior, you can try to generlize this further. Say you want a size 4 Jordan block with eigenvalue 1. Then for any orbit curve, again the last coordinate is constant for some $d \in \mathbb{C}$. If the third coordinate is t, then the second coordinate would again be $\frac{1}{2d}t^2 - \frac{1}{2}t$ shifted around by some constant. And finally, the first coordinate would be a degree 3 polynomial in t. It would look like some form of spiral. Consider $(1, t, t^2, t^3)$ for a idea of this kind of spirals.

Example 2.71. *Consider a Jordan block with eigenvalue, say* $J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$*. Then it sends $(0, 0, 1)^T$ to $(0, 1, 2)^T$, then to $(1, 4, 4)^T$ and so on. It looks like you are doing $(x + 2)^k$.*

Indeed, algebraically $J^k(a, b, c)^T$ is the last three coordinates of $(ax^2 + bx + c)(x + 2)^k$ for the same reason as before. Now you generalize this to get the algebraic behavior of all Jordan blocks of all size for all eigenvalues.

What about its geometric behavior? Suppose we start at some vector $(a_0, b_0, c_0)^T$, and we construct $J(a_{n-1}, b_{n-1}, c_{n-1})^T = (a_n, b_n, c_n)^T$. Then we see that $c_n = 2^n c_0$.

We can see that $b_n = 2b_{n-1} + c_{n-1}$. Divide this by 2^n on both sides (because we know all three sequences must be related to 2^n somehow, as 2 is the eigenvalue), we see that $\frac{b_n}{2^n} = \frac{b_{n-1}}{2^{n-1}} + \frac{c_0}{2}$. So the sequence $\frac{b_n}{2^n}$ is arithmetic and $\frac{b_n}{2^n} = \frac{b_0}{2^0} + \frac{c_0}{2}n$. So $b_n = 2^n b_0 + n2^{n-1}c_0$.

Finally, $a_n = 2a_{n-1} + b_{n-1}$. By a similar argument, $\frac{a_n}{2^n} = \frac{a_{n-1}}{2^{n-1}} + \frac{b_0}{2} + (n - 1)\frac{c_0}{4}$. So $\frac{a_n}{2^n}$ is a degree two polynomial in n, and specifically you can see that $\frac{a_n}{2^n} = \frac{b_0}{2}n + \frac{c_0}{4}(0 + 1 + 2 + ... + (n - 1)) = \frac{c_0}{8}n^2 + (\frac{b_0}{2} - \frac{c_0}{8})n$. So $a_n = n^2 2^{n-3}c_0 + n2^{n-3}(4b_0 - c_0)$.

So a typical curve looks like $p(t) = (t^2 2^{t-3}c_0 + t2^{t-3}(4b_0 - c_0), 2^t b_0 + t2^{t-1}c_0, 2^t c_0)^T$. By a change in parametrization, we can choose $2^t c_0$ as the new parameter t, then the curve is $p(t) = t(a(t), b(t), c(t))^T$ where a, b, c here are polynomials in $\ln t$ of degree 2,1,0.

Also note that asymptotically for super large n, $\lim \frac{b_n^2}{2a_n c_n} = 1$. Therefore these curves has asymptotic surface $xz = y^2$. What is this surface? It is a cone around the line $\{y = 0\} \cap \{x = z\}$. So all these orbital curves will eventually get closer and closer to this cone.

Example 2.72. *As shown in the example above, the geometric picture of a Jordan block is not always easy to compute. However, let us try to do another case,* $J = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$ *for some extremely large λ. Then since λ is so large, comparatively the 1 is ignorable. So $J \approx \lambda I$. This geometric picture is very easy now, it is approximately just stretch everything by λ. So the orbits are approximately just rays shooting from the origin, with minor perturbations.*

3 Applications of Jordan normal form

3.1 Cayley-Hamilton Theorem

We want to establish the following Cayley Hamilton Theorem. We shall give two proofs.

Theorem 3.1 (Cayley-Hamilton Theorem)**.** *Let $p_A(x)$ be the characteristic polynomial of A. Then $p_A(A) = 0$.*

The first proof is by using Jordan blocks.

Lemma 3.2. *$p(J) = 0$ for a nilpotent Jordan block J of size n iff x^n divides $p(x)$.*

Proof. Suppose x^n does not divides $p(x)$. Let $a < n$ be the largest integer such that x^a divides $p(x)$. Then $p(x) = x^{a+1}q(x) + kx^a$ for some nonzero k.

Now we have $J^{n-a-1}p(J) = J^n q(J) + kJ^{n-1} = kJ^{n-1} \neq 0$. Contradiction. (Note that this step requires $n - a - 1 \geq 0$, since we cannot take the inverse of J. And luckily, our requirement $a < n$ guarantees this to be possible.) □

Corollary 3.3. *$p(J) = 0$ for a Jordan block J of size n and eigenvalue λ iff $(x - \lambda)^n$ divides $p(x)$.*

Lemma 3.4. *For any square matrix A, let its eigenvalues be $\lambda_1, ..., \lambda_k$ NOT counting multilicity. Let d_i be the size of the largest Jordan block of A for the eigenvalue λ_i. Then $p(A) = 0$ iff $(x - \lambda_i)^{d_i}$ divides $p(A)$ for all λ_i.*

Proof. Say $A = diag(J_1, ..., J_m)$ in Jordan normal form. Then $p(A) = diag(p(J_1), ..., p(J_n))$ for all polynomial p. So some p kills A iff it simultaneously kills each Jordan block. And we are done. □

Definition 3.5. The minimal polynomial of a matrix A is a polynomial p with leading coeffecient 1 and smallest possible degree such that $p(A) = 0$.

Corollary 3.6. *The minimal polynomial is unique (and hence we say "the" minimal polynomial rather than "a" minimal polynomial). For any square matrix A, let its eigenvalues be $\lambda_1, ..., \lambda_k$ NOT counting multilicity. Let d_i be the size of the largest Jordan block of A for the eigenvalue λ_i. Then the minimal polynomial is $\prod(x - \lambda_i)^{d_i}$.*

Corollary 3.7. *$p(A) = 0$ iff the minimal polynoimal divides $p(x)$.*

Now, in the characteristic polynomial, the exponent for each $x - \lambda_i$ is its algebraic multiplicity, which is always larger than the size of largest Jordan block for λ_i! So we have happily proven the Cayley Hamilton Theorem. (And established a super nice criteria of when $p(A) = 0$.)

The first proof is long but gives us more insight into the theories of polynomials of matrices. The second proof here is shorter but with less insight, and more geometric intuition.

Lemma 3.8. *Cayley-Hamilton Theorem is true for diagonalizable polynomials.*

Proof. If A is diagonalizable, then it has a basis made of eigenvectors. It is enough to show that $p_A(A)$ kills all eigenvectors of A.

Suppose $Av = \lambda v$. Then since $(x - \lambda)$ divides $p_A(x)$, we can write $p_A(x) = (x - \lambda)q(x)$. And we see that $p_A(A)v = q(A)(A - \lambda I)v = q(A)0 = 0$. So we are done. □

Lemma 3.9. *The function $f : M_n \to M_n$ that sends each A to $p_A(A)$ is continuous. (Note that M_n is naturally an n^2 dimensional space, so f is just a function sending n^2 variables to n^2 variables. So you can talk about continuity and so on.)*

Proof. First, the Leibniz formula for determinant tells us that determinants are polynomials of the entries of a matrix, so taking determinant is continuous. Then $p_A(x) = \det(xI - A)$ depends on A continuously. Then $p_A(A)$ depends on A continuously, so we are done.

To be more precise, we know then the process of going from A to $\det(xI - A)$ is continuous for each fixed x. Let $b_k = \det(kI - A)$ for $k = 0, 1, ..., n$, and let $M = \begin{pmatrix} 1 & 0 & ... & 0 \\ 1^0 & 1^1 & ... & 1^n \\ 2^0 & 2^1 & ... & 2^n \\ \vdots & \vdots & \ddots & \vdots \\ n^0 & n^1 & ... & n^n \end{pmatrix}$. Then say $p_A(x) = \det(xI - A) = \sum a_k x^k$. Then $M \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix}$. Furthermore, M is invertible since we can compute its determinant as a Vandermonde matrix. So coefficients of p_A depends linearly on $b_0, ..., b_n$, while these b_k are computed as polynomials of entries of A.

So coefficients of p_A are continuous in entries of A. So in the expression $p_A(A)$, everything depends continuously on entries of A. So f is continuous. $\square$

Lemma 3.10. *Any matrix is a limit of diagonaliable matrices with distinct eigenvalues. (In fact, you can slightly modify this proof to further require the matrices to be invertible.)*

Example 3.11. *Before the proof, here is an intuitive example. Say in Jordan normal form we have* $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. *Then we consider* $A_t = \begin{pmatrix} 1+t & 1 & 0 \\ 0 & 1+2t & 0 \\ 0 & 0 & 2+3t \end{pmatrix}$. *Then the eigenvalues of* A_t *would be* $1+t, 1+2t, 2+3t$, *and as long as* $0 < |t| < 1$, *these are distinct. And* $\lim_{t\to 0} A_t = A$.

Proof. Take any matrix A. Say its eigenvalues are $\lambda_1, ..., \lambda_k$ NOT counting multiplicity. Let g be the smallest gap, i.e., $g = \min |\lambda_i - \lambda_j|$. We suppose A is already in it Jordan normal form.

Now let $D = diag(1, 2, ..., n)$, and consider $A_t = A + tD$, then obviously $\lim_{t\to 0} A_t = A$. I claim that, for $0 < |t| < \frac{g}{n}$, then A_t must have distinct eigenvalues (and thus diagonalizable).

Note that A_t is still upper triangular, so the eigenvalues are the diagonal entries. Then any two diagonal entries of A_t is like $\lambda_i + \alpha t$ and $\lambda_j + \beta t$ where $1 \le i, j \le k$ and $1 \le \alpha, \beta \le n$ with $\alpha \neq \beta$.

Consider $|(\lambda_i + \alpha t) - (\lambda_j + \beta t)| \ge ||\lambda_i - \lambda_j| - |\alpha - \beta||t||$. Now, if $\lambda_i = \lambda_j$, then we are looking at $|\alpha - \beta||t|$ where $\alpha \neq \beta$ and $t \neq 0$, so this is positive. If $\lambda_i \neq \lambda_j$, then $|\lambda_i - \lambda_j| \ge g$, while $|\alpha - \beta||t| < n\frac{g}{n} = g$. So $||\lambda_i - \lambda_j| - |\alpha - \beta||t|| > 0$. Either way, we see that any two eigenvalues of A_t must be distinct if $0 < |t| < \frac{g}{n}$. $\square$

Proof of Cayley-Hamilton. For any matrix A, take diagonalizable A_t with $\lim A_t = A$. Then $f(A) = f(\lim A_t) = \lim f(A_t) = \lim 0 = 0$. Here you see that continuity of f is essential. $\square$

This idea can be used in many settings. The idea is this: for continuous behavior, you can first prove it on diagonalizable (and maybe also invertible) matrices. Then use continuity, you automatically proved it for ALL matrices. Here is an example.

Proposition 3.12. *For cofactor matrices, we have* $Cof(AB) = Cof(A)Cof(B)$.

Proof. Consider the function f that sends a pair of matrices A, B to $Cof(AB) - Cof(A)Cof(B)$. This is a function from $2n^2$ variables to n^2 variables, and it can be computed completely as polynomials of entries of A and B, so it is continuous.

Now suppose A, B are invertible. Then $Cof(A) = \det(A)(A^T)^{-1}$, $Cof(B) = \det(B)(B^T)^{-1}$, and $Cof(AB) = \det(AB)((AB)^T)^{-1} = \det(A)\det(B)(B^TA^T)^{-1} = \det(A)\det(B)(A^T)^{-1}(B^T)^{-1} = \det(A)(A^T)^{-1}\det(B)(B^T)^{-1}$ $Cof(A)Cof(B)$. So $f(A, B) = 0$.

Now take invertible A_t with limit A and B_t with limit B. Then $f(A, B) = f(\lim(A_t, B_t)) = \lim f(A_t, B_t) = \lim 0 = 0$. $\square$

Finally, wikipedia also gives a bazillion other proofs. You can check them out if you like.

3.2 Functions of Matrices

Polynomials of matrices are simple. Say $p(x) = x^3 + 2x^2 + 3x + 4$, then $p(A) = A^3 + 2A^2 + 3A + 4I$. Then since we have polynomials, it is very natural to move on to power series.

Definition 3.13. Suppose f is an analytic function (a function that equals to its Taylor expansion, say exponential functions or sine functions). Then $f(x) = \sum_{i=0}^{\infty} a_i x^i$ for some sequence a_i, and thus we can define $f(A) = \sum_{i=0}^{\infty} a_i A^i$.

Example 3.14. *You can see that* $e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \ldots$. *In particular, suppose* $A = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}$. *Then* $A^2 = -\theta^2 I$. *Therefore* $A^{2k} = (-1)^k \theta^{2k} I$, *while* $A^{2k+1} = (-1)^k \begin{pmatrix} 0 & -\theta^{2k+1} \\ \theta^{2k+1} & 0 \end{pmatrix}$. *So we have*

$$
\begin{aligned}
e^A &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix} + \frac{1}{2!}\begin{pmatrix} -\theta^2 & 0 \\ 0 & -\theta^2 \end{pmatrix} + \frac{1}{3!}\begin{pmatrix} 0 & \theta^3 \\ -\theta^3 & 0 \end{pmatrix} + \ldots \\
&= \left(\begin{array}{c|c} 1 - \frac{1}{2!}\theta^2 + \frac{1}{4!}\theta^4 - \ldots & -\theta + \frac{1}{3!}\theta^3 - \frac{1}{5!}\theta^5 + \ldots \\ \hline \theta - \frac{1}{3!}\theta^3 + \frac{1}{5!}\theta^5 - \ldots & 1 - \frac{1}{2!}\theta^2 + \frac{1}{4!}\theta^4 - \ldots \end{array}\right) \\
&= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.
\end{aligned}
$$

Interestingly, note that the matrix A here can be thought of as a representation of the complex number $i\theta$, *and the output* $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ *is representing the complex number* $\cos\theta + i\sin\theta$. *So we have proven that* $e^{i\theta} = \cos\theta + i\sin\theta$. *In particular, we have* $e^{i\pi} + 1 = 0$, *the famous Euler identity.*

Of course, computing power series is a pain. We should be able to find some shortcut.

Proposition 3.15. *Suppose* $A = BDB^{-1}$ *for diagonal* $D = diag(a_1, ..., a_n)$. *Then* $f(A) = Bf(D)B^{-1}$ *and* $f(D) = diag(f(a_1), f(a_2), ..., f(a_n))$.

Proof. Analytic functions are limits of polynomials. Since the statement is true for polynomials, and all calculations involved are continuous, therefore the statement is true for analytic functions. □

So diagonalizable functions are easy to compute. What about non-diagonalizable matrices? Again we invoke the density argument. They are limits of diagonalizable functions.

Example 3.16. *Consider* $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$. *Let* $A_t = \begin{pmatrix} 2 & 1 \\ 0 & 2+t \end{pmatrix}$, *then* A_t *are diagonalizable for* $t \neq 0$, *and* $\lim_{t\to 0} A_t = A$. *To diagonalize* A_t *for* $t \neq 0$, *you can do* $A_t = B_t D_t B_t^{-1}$ *where* $D_t = diag(2, 2+t)$ *and* $B_t = \begin{pmatrix} 1 & \frac{1}{t} \\ 0 & 1 \end{pmatrix}$.

Now $f(A) = f(\lim A_t) = \lim f(A_t) = \lim B_t \begin{pmatrix} f(2) & 0 \\ 0 & f(2+t) \end{pmatrix} B_t^{-1} = \lim \begin{pmatrix} f(2) & \frac{f(2+t)-f(2)}{t} \\ 0 & f(2+t) \end{pmatrix} = \begin{pmatrix} f(2) & f'(2) \\ 0 & f(2) \end{pmatrix}$. *Curiously, the derivatives creep into it!*

Thinking back about this, this should NOT be a surprise. A Jordan block of size two means we have a repeated eigenvalue, and $f'(2)$ *is the secant line of repeated points on the graph of* f. *The multiplicity is innately present in the derivative.*

Proposition 3.17. *For a Jordan block* $J = \begin{pmatrix} a & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & a \end{pmatrix}$ *of size n and eigenvalue a, and any analytic function f, we see that* $f(J) = \begin{pmatrix} f(a) & f'(a) & \frac{1}{2!}f''(a) & \dots & \frac{1}{(n-1)!}f^{(n-1)}(a) \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{1}{2!}f''(a) \\ & & & \ddots & f'(a) \\ & & & & f(a) \end{pmatrix}$.

Proof. It is enough to prove this for all polynomials (since f is a limit of them). And then, it is enough to prove it for power functions $p(x) = x^k$ for some k, since all polynomials are linear combinations of them.

Now if $p(x) = x^k$, then $p(J) = J^k$ which is J shifted up $(k-1)$ times. You can check that the entries are exactly the values in our conjectured formula. □

So in general, how to calculate $f(A)$? First you put A in Jordan normal form, then you apply f on each Jordan block using the formula above, and you are done. In particular, we see the following curious thing:

Corollary 3.18. *Given a matrix A, with eigenvalues $\lambda_1, ..., \lambda_k$ and maximal Jordan block sizes $d_1, ..., d_k$ for each eigenvalue, the entries of $f(A)$ depends only on $f^{(i)}(a_j)$ where $1 \leq j \leq k$ and $0 \leq i \leq d_j$. In particular, if f and g agrees at all eigenvalues of A in values and in derivatives and in enough higher order derivatives, then $f(A) = g(A)$. We say that f and g agrees on the Spectrum of A.*

Definition 3.19. Turns out that we don't need to focus only on analytic functions. Fix a matrix A, with eigenvalues $\lambda_1, ..., \lambda_k$ and maximal Jordan block sizes $d_1, ..., d_k$ for each eigenvalue. For any function f that is d_i-times differentiable at λ_i, then we can define $f(A)$ using the formula for $f(J)$ for each Jordan block J of A. This allows us to define $f(A)$ for a super wide range of functions.

We have some other goodies as well.

Corollary 3.20. *If the eigenvalues of A are $\lambda_1, ..., \lambda_n$ counting algebraic multiplicity, then the eigenvalues of $f(A)$ are $f(\lambda_1), ..., f(\lambda_n)$ counting algebraic multiplicity.*

Corollary 3.21. $e^{\mathrm{tr}(A)} = \det(e^A)$.

Proposition 3.22. $f(A^T) = f(A)^T$*, and* $f(\overline{A}) = \overline{f(A)}$*, and* $f(A^*) = (f(A))^*$.

Proof. Recall that $f(A)$ is defined via polyomial series, i.e., $f(A) = \sum a_i A^i$. Then $f(A^T) = \sum a_i (A^T)^i = \sum a_i (A^i)^T = (\sum a_i A^i)^T = f(A)^T$. The proof for complex conjugates and adjoints are the same. □

3.3 Matrix Exponentials

There are many applications of functions of matrices. For example, check out Chapter 2 of the book Functions and Matrices: Theories and Computation by Nicholas J. Higham. For our purpose here, we only present a simple one, solving differential equations using e^A.

To start, observe that $e^{a+b} = e^a e^b$. Is this still true for matrices? Not always.

Proposition 3.23. $e^{A+B} = e^A e^B$ *if* $AB = BA$.

Proof. If $AB = BA$, then by the power series definition of the exponential function, you can easily verify that $e^{A+B} = e^A e^B$. For details,

$$\begin{aligned}
e^A e^B =&(\sum_n \frac{A^n}{n!})(\sum_m \frac{B^m}{m!}) \\
=&\sum_{m,n} \frac{A^n B^m}{n!m!} \\
&\text{(Now we re-index the sum by setting } s = m+n. \\
&\text{This is similar to a change of variable for integrations.} \\
&\text{Note that this requires absolute converging sequences)} \\
=&\sum_s \sum_{0\leq n\leq s} \frac{A^n B^{s-n}}{n!(s-n)!} \\
=&\sum_s \frac{1}{s!} \sum_{0\leq n\leq s} \frac{s!}{n!(s-n)!} A^n B^{s-n} \\
=&\sum_s \frac{1}{s!}(A+B)^s \\
=&e^{A+B}.
\end{aligned}$$

Note that the binomial thing involved here REQUIRES $AB = BA$. You also see that if $AB \neq BA$, then the deduction above can very easily fall off at the binomial part. □

Corollary 3.24. *If $A^T = -A$ for a real matrix A, then e^A is an orthogonal matrix with determinant one (i.e., a rotation). Similarly, if $A^* = -A$ for a complex matrix A, then e^A is a unitary matrix. Conversely, all orthogonal matrices with determinant one are equal to the exponential of some real skew-symmetric matrices, and all unitary matrices are equal to the exponential of some skew-Hermitian matrices.*

Proof. Suppose $A^* = -A$. Then A and A^* commute. So $I = e^{A^*+A} = e^{A^*}e^A = (e^A)^*e^A$. So e^A is unitary in general, and it is orthogonal if A is real.

For a real matrix A, if $A^T = -A$, then all diagonal entries of A are zero, so $\mathrm{tr}(A) = 0$. Then $\det(e^A) = e^{\mathrm{tr}(A)} = e^0 = 1$.

Now for the converse. Suppose A is unitary, then by the spectral theorem, $A = BDB^{-1}$ for some unitary B and diagonal $D = diag(d_1, ..., d_n)$. Since A is unitary, by spectral theorem all d_k have norm 1. So for each d_k we have $d_k = e^{i\theta_k}$. Simply set $C = diag(i\theta_1, ..., i\theta_n)$. (Here we are like taking the logarithm of D.) Then we would have $e^C = D$, and $e^{BCB^{-1}} = Be^CB^{-1} = A$. Now we only need to show that BCB^{-1} is skew-Hermitian. Indeed, since B is unitary and C is purely imaginary, we have $(BCB^{-1})^* = BC^*B^{-1} = B(-C)B^{-1} = -BCB^{-1}$.

In the case where A is real orthogonal with determinant one, by a corollary of spectral theorem we have $A = BDB^{-1}$ where B is real orthogonal, and $D = diag(I_m, R(\theta_1), ..., R(\theta_t))$ is block diagonal where I_m is an m by m identity matrix, and $R(\theta)$ represents a 2 by 2 rotation matrix with angle θ counterclockwise. Now denote $\begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}$ as $S(\theta)$, and set $C = diag(0_m, S(\theta_1), ..., S(\theta_t))$. (Here we are like taking the logarithm of D.) Then we would have $e^C = D$, and as the complex case, we now have $e^{BCB^{-1}} = A$. Finally, since B is orthogonal and C is skew-symetric, $(BCB^{-1})^T = BC^TB^{-1} = -BCB^{-1}$. □

Example 3.25. *Let us consider a 3D rotation. This must be e^A for some 3 by 3 skew-symmetric matrix A. Say $A = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}$. Then immediately we see that $v = \begin{pmatrix} c \\ -b \\ a \end{pmatrix}$ is an eigenvector for the eigenvalue 0* *of A. As a result, it is an eigenvector for the eigenvalue $e^0 = 1$ of the matrix e^A, i.e., $e^A v = v$. So v is the AXIS OF ROTATION!*

So, how to build a 3D rotation around some axis $(x, y, z)^T$? Simply take $e^{k\begin{pmatrix} 0 & z & -y \\ -z & 0 & x \\ y & -x & 0 \end{pmatrix}}$, and this will always be a rotation around $(x, y, z)^T$. By setting k to be various value, you will obtain a rotation of various degree.

Furthermore, for an axis of rotation $v = (a, b, c)^T$, consider the eigenvalues of $A = \begin{pmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{pmatrix}$. Its characteristic polynomial is $x^3 + (a^2 + b^2 + c^2)x = 0$. So the three eigenvalues are $0, ||v||i, -||v||i$. So if we pick v to be a unit vector, $e^{\theta A}$ would have eigenvalues 0,$i\theta$, $-i\theta$, so it would corresponds to a rotation by θ.

So the quickest way to set up a 3D rotation around some axis by θ is to first pick a unit vector for the axis of rotation $v = (a, b, c)^T$, and then the desired matrix is $e^{\theta\begin{pmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{pmatrix}}$.

The most amazing thing about e^x is that it equal to its own derivative. In particular, $\frac{d}{dt}e^{kt} = ke^{kt}$. Funnily, we have the following:

Proposition 3.26. $\frac{d}{dt}e^{At} = Ae^{At}$.

Proof. First note that $\frac{e^{A(t+dt)} - e^{At}}{dt} = \frac{e^{At}(e^{Adt} - I)}{dt} = \frac{e^{Adt} - I}{dt}e^{At}$. Therefore, we only need to prove that $\lim_{t\to 0} \frac{e^{At} - I}{t} = A$.

Now $e^{At} = I + At + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \ldots$. So the statement becomes obvious. □

Theorem 3.27. *Suppose we have functions $f_1(x), f_2(x), \ldots, f_n(x)$ such that the derivative on each f_i is a linear combination of $f_1, \ldots, f_n$. I.e., if we write $f(x) = (f_1(x), \ldots, f_n(x))^T$, then we have $f'(x) = Af(x)$ for some constant matrix A. Then f is a linear combination of columns of e^{Ax}.*

Proof. Columns of e^{Ax} are obviously solutions. Furthermore, $\det(e^{Ax}) = e^{\mathrm{tr}(Ax)} \neq 0$, so the n columns of e^{Ax} are linearly independent. It is now enough to show that the solution space is n-dimensional.

If A is diagonalizable, say $A = B^{-1}DB$ with $D = diag(d_1, \ldots, d_n)$, then note that $f' = Af$ implies $(Bf') = D(Bf)$. Therefore $(Bf)' = D(Bf)$. Since B is invertible, it is enough to show that $f' = Df$ has a solution space with dimension n. But $f' = Df$ means that $f_i' = d_i f_i$ for all i, so the solution space is $(k_1e^{d_1x}, k_2e^{d_2x}, \ldots, k_ne^{d_nx})^T$ for arbitrary $k_1, \ldots, k_n$. So the solution space is n dimensional.

If A is NOT diagonalizable, we can similarly put A in Jordan normal form, and then we only need to prove this for each Jordan block. First let us do a simple case, say J is a m by m Jordan block with eigenvalue 0, and $f' = Jf$. Then $f_i' = f_{i+1}$ and $f_m' = 0$. So f_m is some constant k_m, and we must have $f_{m-1}(x) = k_mx + k_{m-1}$, and $f_{m-2}(x) = \frac{k_m}{2!}x^2 + k_{m-1}x + k_{m-2}$, and so on until $f_1(x) = \sum \frac{k_i}{i!}x^i$. So they are all polynomials, and there are n unknowns, forming an n-dim space as predicted.

What if we have eigenvalue a? Then $f_i' = af_i + f_{i+1}$ and $f_m' = af_m$. On one hand, it is reasonable to assume that polynomials are involved. On the other hand, we see that $f_m(x) = k_me^{ax}$ is exponential. So it is not hard to guess that f_i are polynomials times e^{ax}.

To rigorously prove this, consider the sequence $\frac{f_i}{f_m}$ instead. We have $(\frac{f_i}{f_m})' = \frac{f_i'f_m - f_if_m'}{f_m^2}$. But remember that $f_m' = af_m$. So we have

$$(\frac{f_i}{f_m})' = \frac{f_i'f_m - af_if_m}{f_m^2} = \frac{f_i'}{f_m} - a\frac{f_i}{f_m} = \frac{f_{i+1}}{f_m}.$$

So since $\frac{f_m}{f_m} = 1$, it follows that $\frac{f_i}{f_m}$ is a polynomial of degree $i - 1$. So $f_i(x) = (k_ix^{i-1} + \ldots + k_m)e^{ax}$. There are n unknowns, forming an n-dim space as predicted. □

Corollary 3.28 (Solving the initial value problem)**.** *Suppose $f' = Af$, then $f(t) = e^{At}f(0)$.*

3.4 Hermite Interpolation

So far, we have been using polynomials to study matrices. However, it goes the other way as well. A very interesting theorem is the Hermite Interpolation. (Yes, the Hermite here is Charles Hermite, and Hermitian matrices and so on are all named after the same guy.)

Example 3.29. *Take number 1,2,3. Can you find a polynomial $p(x)$ such that $p(1) = p(2) = p(3) = 0$? Of course. This is simple, as the answer is $p(x) = (x-1)(x-2)(x-3)$.*

Can you find a polynomial $p(x)$ such that $p(1) = 2, p(2) = 1$ and $p(3) = 4$? This is more tricky. Let $p_i(x) = \prod_{j\neq i} \frac{(x-j)}{(i-j)}$, then $p_i(i) = 1$ while $p_i(j) = 0$ for all $j \neq i$. Now set $p(x) = 2p_1(x) + p_2(x) + 4p_3(x)$, and we are done. This trick is called the Lagrange interpolation. This idea has a highly "linear algebraic" feel. Essentially we find a "basis" for such a phenomena, i.e., p_1, p_2, p_3, and then the answer is a linear combination of them. Indeed, let L be the linear map from the space of polynomials to $\mathbb{C}^3$ such that $L(p) = (p(1), p(2), p(3))^T$, then $L(p_1), L(p_2), L(p_3)$ form a basis in $\mathbb{C}^3$. Then any vector in $\mathbb{C}^3$ can be achieved as a linear combination of $L(p_1), L(p_2), L(p_3)$, and thus it is the image of L from a linear combination of p_1, p_2, p_3. (Note that by replacing $\mathbb{C}$ by $\mathbb{R}$, you will achieve a version over real numbers and so on.)

Can you find a polynomial $p(x)$ such that $p(1) = 2, p(2) = 1, p(3) = 4$ and $p'(1) = 5$? Well well well, this is not so obvious, but the answer is still yes. Consider $p_i(x) = \frac{\prod_{j\neq i}(x-j)}{(i-j)}$ as before. Now, their linear combination can achieve any value at 1,2,3. However, how can we control the derivative at 1? Again, let L be the linear map from the space of polynomials to $\mathbb{C}^4$ such that $L(p) = (p(1), p(2), p(3), p'(1))^T$. You can check that this is indeed linear. So we need a forth polynomial q such that $q(1) = q(2) = q(3) = 0$ and $q'(1) = 1$. Suppose we find this, then $L(p_1) = (1, 0, 0, *)^T, L(p_2) = (0, 1, 0, *)^T, L(p_3) = (0, 0, 1, *)^T, L(q) = (0, 0, 0, 1)^T$ would form a basis, and we can then find p as a linear combination of p_1, p_2, p_3, q. In particular $p(x) = 2p_1(x) + p_2(x) + 4p_3(x) + (5 - 2p_1'(1) - p_2'(1) - 4p_3'(1))q(x)$.

We can find this $q(x)$ as $q(x) = (x-1)p_1(x)^2$.

Can we always find a polynomial that takes prescribed values and derivatives and higher derivatives at prescribed places? The answer is yes. This is Hermite Interpolation. You can probably already get a proof from above examples.

Theorem 3.30. *Take distinct complex numbers $a_1, ..., a_k$, and non-negative integers $n_1, ..., n_k$, and arbitrary complex numbers b_{ij} for $1 \leq i \leq k$ and $0 \leq j \leq n_i$. Then there is a polynomial $p(x)$ such that $p^{(j)}(a_i) = b_{ij}$ for all $1 \leq i \leq k$ and $0 \leq j \leq n_i$.*

We now proceed to prove this WITHOUT constructing any polynomial. How? The key to finding a polynomial lies in the commuting behavior of matrices! We first see a mini version.

Proposition 3.31. *If A has distinct eigenvalues, then $AB = BA$ iff A, B can be simlutaneously diagonalized.*

Proof. Say A has distinct eigenvalues $\lambda_1, ..., \lambda_n$ with corresponding eigenvectors $v_1, ..., v_n$. Note that all eigenspaces for A are necessarily one-dimensional.

Suppose $AB = BA$. Now $ABv_i = BAv_i = B(\lambda_i v_i) = \lambda_i Bv_i$. So Bv_i is also an eigenvector for λ_i. Since all eigenspaces for A are necessarily one-dimensional, it follows that Bv_i is a multiple of v_i. So v_i is an eigenvector for B.

So $v_1, ..., v_n$ are all eigenvectors for B. So under this basis, B is diagonal as well. So under this basis A, B are simultaneously diagonal.

Conversely, if A, B are simultanously diagonal under some basis, then we know diagonal matrices all commute with each other, so we have $AB = BA$. □

Corollary 3.32. *If A has distinct eigenvalues, then $AB = BA$ iff $B = p(A)$ for some polynomial A.*

Proof. The necessity is trivial. For sufficiency, suppose $AB = BA$. Then under the right basis, A, B are both diagonal, say $A = diag(a_1, ..., a_n)$ and $B = diag(b_1, ..., b_n)$. Since a_i are all distinct, pick any polynomial such that $p(a_i) = b_i$, then $p(A) = B$. You can construct this polynomial via Lagrange Interpolation. □

Above is just a mini example. We don't really need it. What we need is a generalized version of it. We will show that, for a matrix A, if it only has ONE Jordan block for each eigenvalue, then $AB = BA$ iff $B = p(A)$ for some polynomial p. (We will prove this without constructing any polynomials explicitly.)

First let us look deeper into our assumptions.

Proposition 3.33. *A has only one Jordan block for each eigenvalue iff the minimal polynomial of A is the same as the characteristic polynomial of A.*

Proof. Sufficiency is trivial.

To show necessity, for each eigenvalue λ, its algebraic multiplicity g is the multiplicity of the factor $x - \lambda$ in the characteristic polynomial of A. Since the minimal polynomial is the same, g is also the multiplicity of the factor $x - \lambda$ in the minimal polynomial. Then λ has a Jordan block of size g. On the other hand, the algebraic multiplicity g is also the sum of the sizes of all Jordan blocks. So it turned out that λ can only have one Jordan block. ☐

Corollary 3.34. *$A \in M_n$ has only one Jordan block for each eigenvalue iff the minimal polynomial of A has degree n.*

Proof. The minimal polynomial must divides the characteristic polynomial. ☐

The new characterization has a very fascinating consequence.

Lemma 3.35. *The minimal polynomial of $A \in M_n$ has degree n iff we can find a vector v such that $v, Av, ..., A^{n-1}v$ are linearly independent (and thus they form a basis). Say the minimal polynomial is $p(x) = x^n + a_{n-1}x^{n-1} + ... + a_0$. Then under this basis, the matrix of A is $(e_2, e_3, ..., e_n, w)$ where $e_i = (0, ..., 0, 1, 0, ..., 0)$ with one in the i-th coordinate, and $v = (-a_0, -a_1, ..., -a_{n-1})$.*

Proof. We only need to prove the first statement, as the second is immediate.

We find v by eliminating all bad vectors. What is a bad vector? If $v, Av, ..., A^{n-1}v$ are linearly dependent, then there is a polynomial $p(x)$ of degree at most $n - 1$, such that $p(A)v = 0$. Say we find such a p with smallest degree. Let $q(x)$ be the minimal polynomial of A, then we also have $q(A)v = 0$. Then any linear combination of $p(A)$ and $q(A)$ would kill v. So we can immitate the Euclidean algorithm to find the greatest common divisor f of p and q, and $f(A)$ would also kill v. So unless $f = p$, otherwise f would have smaller degree than p, a contradiction. So p divides the minimal polynomial.

So any bad v is in the kernel of $p(A)$ for some proper factor $p(x)$ of the minimal polynomial. (Proper factor means that p is a factor of the minimal polynomial, but it is NOT the minimal polynomial itself.) Since the minimal polynomial only has finitely many proper factors (up to a multiplicative constant), therefore there are only finitely many such kernels. So the bad v are contained in finitely many subspaces of dimension at most $n - 1$. Geometrically, these can NOT hope to fill up $\mathbb{R}^n$, so there must be some good vector left. ☐

Proposition 3.36. *If the minimal polynomial of $A \in M_n$ has degree n, then $AB = BA$ iff B is a polynomial of A.*

Proof. Find v as above. Then since $v, Av, ..., A^{n-1}v$ form a basis, we see that Bv is a linear combination of them, say $Bv = \sum a_i A^i v$. Then let $p(x) = \sum a_i x^i$, we see that $Bv = p(A)v$.

But then $BA^k v = A^k Bv = A^k p(A)v = p(A)A^k v$. So B agrees with $p(A)$ on each vector in the basis $v, Av, ..., A^{n-1}v$. So $B = p(A)$ after all. ☐

Corollary 3.37 (Hermite Interpolation)**.** *Fix any distinct $a_1, ..., a_k$ and any non-negative integers $d_1, ..., d_k$. Then for any numbers b_{ij} where $1 \leq i \leq k$ and $0 \leq j \leq d_i$, we can find a polynomial $p(x)$ such that $p^{(j)}(a_i) = b_{ij}$.*

Proof. Let A be a matrix in Jordan normal form $diag(J_1, ..., J_k)$, where each Jordan block J_i has eigenvalue a_i and size d_i. Let B be a matrix $diag(B_1, ..., B_k)$, where each B_i is d_i by d_i and upper triangular. The diagonal entries of B_i are b_{i0}, and the entries above the diagonal are $\frac{b_{i1}}{1!}$, and then the entries above are $\frac{b_{i2}}{2!}$, and so on. So take a 4 by 4 example, B_i should look like this $\begin{pmatrix} b_{i0} & b_{i1} & b_{i2}/2 & b_{i3}/6 \\ 0 & b_{i0} & b_{i1} & b_{i2}/2 \\ 0 & 0 & b_{i0} & b_{i1} \\ 0 & 0 & 0 & b_{i0} \end{pmatrix}$.

It is easy to see that each B_i commute with J_i. In particular, you might even be able to tell that, say in the 4 by 4 example above, then $B_i = \frac{b_{i3}}{3!}(J_i - a_i I)^3 + \frac{b_{i2}}{2!}(J_i - a_i I)^2 + b_{i1}(J_i - a_i I) + b_{i0}$. Basically each B_i is a polynomial of J_i (but maybe for different polynomials). Any way, this tells us that $AB = BA$. But then this implies that B is a polynomial of A. So we are done. □

The intuition here is that, as long as the inputs are distinct, then you can find a polynomial satisfying any restrictions on values and arbitrary-order derivatives at these inputs.

3.5 Commuting matrices

We start with case 1, matrices with distinct eigenvalues, as a motivation. Then we do a "case 0" as a general analytical perspective to make things clearer. Then we do case 2,3,4 as some special applications of case 0.

Case 1: A has Distinct Eigenvalues.

In this subsection, we tried to answer the following question: what does commutativity for matrices mean? Think about it. What would commute with A? Well for starters, if A is diagonal, then all other diagonal matrices commute with A. Indeed, we have the following.

Proposition 3.38. *If A has distinct eigenvalues, then $AB = BA$ iff A, B can be simlutaneously diagonalized.*

Proof. Say A has distinct eigenvalues $\lambda_1, ..., \lambda_n$ with corresponding eigenvectors $v_1, ..., v_n$. Note that all eigenspaces for A are necessarily one-dimensional.

Suppose $AB = BA$. Now $ABv_i = BAv_i = B(\lambda_i v_i) = \lambda_i Bv_i$. So Bv_i is also an eigenvector for λ_i. Since all eigenspaces for A are necessarily one-dimensional, it follows that Bv_i is a multiple of v_i. So v_i is an eigenvector for B.

So $v_1, ..., v_n$ are all eigenvectors for B. So under this basis, B is diagonal as well. So under this basis A, B are simultaneously diagonal.

Conversely, if A, B are simultanously diagonal under some basis, then we know diagonal matrices all commute with each other, so we have $AB = BA$. □

For another idea, the Jordan blocks of A clearly have distinct eigenvalues. So $AB = BA$ iff B is a polynomial of A.

Corollary 3.39. *If A has distinct eigenvalues, then $AB = BA$ iff $B = p(A)$ for some polynomial A.*

Proof. The necessity is trivial. For sufficiency, suppose $AB = BA$. Then under the right basis, A, B are both diagonal, say $A = diag(a_1, ..., a_n)$ and $B = diag(b_1, ..., b_n)$. Since a_i are all distinct, pick any polynomial such that $p(a_i) = b_i$, then $p(A) = B$. This is possible by Hermite interpolation. □

So it seems that commutativity has to do with simultaneous diagonalization and/or polynomial relation.

Case 0: Spatial Decomposition and Reduction to Nilpotency

What would happen if we have repeated eigenvalues? Then neither of above is true. Consider this:

Example 3.40. *Set $A = I$. Then $AB = BA$ for ALL B. Take $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, then obviously A, B cannot be simultaneously diagonalized. Furthermore, since $p(A) = p(1)I$ is always diagonal, so B is NOT a polynomial of A.*

However, I claim the following: If A is diagonalizable (but may have repeated eigenvalues), then $AB = BA$ implies that A, B can be simultaneously Jordanized, i.e., put into Jordan normal form. (But not the other way around!) And furthermore, $AB = BA$ iff there is a matrix C such that $A = p(C)$ and $B = q(C)$ for some polynomials $p(x), q(x)$.

What if we simply know nothing about either A or B? We already know that, if each eigenvalue of A corresponds to only ONE Jordan block, then $AB = BA$ iff B is a polynomial of A. If some eigenvalues of A possess more than one Jordan block, then we don't know anything detail, unless the dimension is tiny (say $n = 3$ or $n = 4$). We do still have some grand pictures about A and B though.

Let us get around to seeing these grand pictures first. The central idea here is to consider subspaces invariant under BOTH A and B. In fact, if $AB = BA$, then they MUST share many invariant subspaces.

Lemma 3.41. *If $AB = BA$, then for any polynomials $p(x), q(x)$, we have $p(A)q(B) = q(B)p(A)$.*

Corollary 3.42. *If $AB = BA$, then for any $\lambda \in \mathbb{C}$ and $k \in \mathbb{Z}$, the subspace $\operatorname{Ker}(B - \lambda I)^k$ is invariant under A.*

Proof. Suppose $v \in \operatorname{Ker}(B - \lambda I)^k$. Then $(B - \lambda I)^k Av = A(B - \lambda I)^k v = A0 = 0$. So $Av \in \operatorname{Ker}(B - \lambda I)^k$. $\square$

Corollary 3.43. *If $AB = BA$, then for any $\lambda \in \mathbb{C}$ and $k \in \mathbb{Z}$, the subspace $\operatorname{Ker}(B - \lambda I)^k$ is invariant under both A and B. In particular, eigenspace or generalized eigenspace of either is invariant under both.*

Think about what this means. Essentially, it means that if we perform a spatial decomposition or filtration using eigen-realated spaces of B, then these subspaces are ALL invariant under A. Say we bring B into block diagonal form according its generalized eigenspaces, then A will automatically be in a corresponding block diagonal form as well! (Even though the blocks of A might be a lot messier than the nice blocks of B.)

Remark 3.44. *In general, for ANY invariant subspace W for B, then $B(AW) = A(BW) = AW$, so AW is another invariant subspaces for B. So A would ALWAYS PERMUTE the invariant subspaces for B. Sometimes A would fix invariant subspaces, sometimes A might swap two invariant subspaces, and so on. Furthermore, above Lemma also identified tons of B-invariant subspaces that must be FIXED by A.*

Corollary 3.45. *Suppose $AB = BA$. Then there is a spatial decomposition $V = \bigoplus V_i$ such that V_i are invariant under both A and B, and B on each V_i is a multiple of identity plus a nilpotent matrix.*

Proof. Just pick each V_i to be a generalized eigenspace of B. $\square$

We have broken down V according to B. But for each subspace, we can further break it down according to A.

Lemma 3.46. *If $A = kI + N$ for some $k \in \mathbb{C}$ and nilpotent N, then for any invariant subspace W for A, $A|_W = kI|_W + N'$ for some nilpotent $N' : W \to W$.*

Proof. Consider $(A - kI)$. Its power would eventually kills everythin in the domain, so its power would also eventually kills everything in any subspace W. So $A|_W - kI|_W = (A - kI)|_W$ is nilpotent. $\square$

Corollary 3.47. *Suppose $AB = BA$. Then there is a spatial decomposition $V = \bigoplus W_i$ such that V_i are invariant under both A and B, and both A, B on each W_i are multiples of identities away from being nilpotent.*

Proof. For each generalized eigensubspace V_i for B, since it is invariant under both, we can restrict our attention to V_i completely. Then we have $A|_{V_i}B|_{V_i} = B|_{V_i}A|_{V_i}$, so we can further decompose each V_i into generalized eigenspaces of $A|_{V_i}$. These spaces are the W_i, invariant under both, and on which A is now identity away from being nilpotent. Since B is identity away from being nilpotent on V_i, this would also be true for any subspace of W_i. But So we have a spatial decomposition $V = \bigoplus W_i$ as desired. $\square$

Now we have obtained a small enough decomposition. A, B on each small subspace are close to nipotent. What does commutativity mean in this setting?

Lemma 3.48. *f $A = kI + N$ for some $k \in \mathbb{C}$ and nilpotent N, then all vectors are generalized eigenvectors of A for the eigenvalue k.*

Proof. Some power of $A - kI$ is 0, so obviously all vectors are in the kernel of some power of $A - kI$. □

Corollary 3.49 (Basis of Common Generalized Eigenvectors)**.** *If $AB = BA$ on a vector space V, then there is a basis for V made by common generalized eigenvectors (but may be for different eigenvalues and different Jordan block size and so on).*

Proof. We do the decomposition $V = \bigoplus W_i$ as before. Then all vectors in each W_i are generalized eigenvectors for both A and B. So pick a basis for each W_i and put them together to get a basis for V. □

This is a very nice result indeed. It means that if $AB = BA$, you can find a basis for V that is nice for BOTH A and B. However, you need to be careful here.

Example 3.50. *Having a basis of common generalized eigenvectors does NOT mean that A and B can be simultaneously Jordanized into their Jordan normal form.*

Consider $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ *and* $B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. *Note that $A = B^2$, so obviously $AB = BA$. And since both are nilpotent, any basis for V is made of common generalized eigenvectors.*

However, note that any Jordan normal form of B must be B itself, while the Jordan normal form of A is $J_A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ *or* $J_A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. *If they can be simultaneously Jordanized, that would implies that $BJ_A = J_AB$, which is impossible for either choice of J_A.*

For another point of view, since any Jordan normal form of B must be B itself, any attempt to put B in Jordan normal form must looks like $B = PBP^{-1}$. But then $A = B^2 = PB^2P^{-1} = PAP^{-1}$. So any attempt to Jordanize B must leave A unchanged, which is NOT the Jordan normal form of A.

Remark 3.51. *In essense, what we have is the following: we can decompose the space V into subspaces W_i invariant under both, and A, B restricted to W_i are multiples of identity away from being nilpotent. So it is enough to figure out the meaning of $AB = BA$ when A, B are both multiples of identity away from being nilpotent.*

But if $A = aI + N$ and $B = bI + N'$ for some nilpotent N, N', then $AB = abI + aN' + bN + NN'$ and $BA = abI + aN' + bN + N'N$, so $AB = BA$ iff $NN' = N'N$. So it is enough to figure out when nilpotent matrices commute.

To sum up, our achievement is the following: To study the meaning of $AB = BA$, it is enough to understand $AB = BA$ when both A and B are nilpotent.

Case 2:* A *is Diagonalizable.

Now we go for our claim about diagonalizable matrices.

Lemma 3.52. *A matrix is diagonalizable iff its minimal polynomial $p(x)$ has distinct roots.*

Proof. Recall that for any eigenvalue λ, $(x - \lambda)^k | p(x)$ iff λ has a Jordan block of size at least k. So $p(x)$ has distinct roots iff all Jordan blocks have size 1 iff A is diagonalizable. □

Corollary 3.53. *If A is diagonalizable, then its invariant subspaces must be sums of subspaces of its eigenspaces. In particular, A is also diagonalizable in any invariant subspace.*

Proof. Let $p(x)$ be the minimal polynomial of A, then $p(x)$ has distinct roots, and $p(A) = 0$. Then for any invariant subspace W, A restricted to W should also satisfy $p(A) = 0$, so its minimal polynomial is a factor fo $p(A)$, and thus has distinct roots as well. So this restriction is also diagonalizable. □

Proposition 3.54. *Suppose A is diagonalizable. Then $AB = BA$ implies that A, B can be simultaneously Jordanized. Furthermore, $AB = BA$ iff you can find a matrix C such that $A = p(C)$ and $B = q(C)$ for some polynomial $p(x), q(x)$.*

Proof. Suppose $AB = BA$. Do the decomposition $V = \bigoplus W_i$ as always. Then on each W_i, A has only one eigenvalue. But since W_i is invariant under A, A is also diagonalizable. So A is a multiple of identity on each W_i.

So pick a basis on each W_i that Jordanize B, and put these basis together to get a basis for V. Then under this basis, B is block diagonal where each block is in Jordan normal form, so B is in Jordan normal form. A is block diagonal where each block is diagonal, so A is diagonal. So A, B are simultaneously Jordanized.

Now let us find C. In the above Jordanization, let $d_k = \dim W_k$. Say $A = diag(a_1 I_{d_1}, a_2 I_{d_2}, ..., a_m I_{d_m})$ and $B = diag(J_{b_1,d_1}, ..., J_{b_m,d_m})$. Then we let $C = diag(J_{1,d_1}, ..., J_{m,d_m})$. Then we can find $p(x)$ such that $p(k) = a_k$ and $p'(k) = ... = p^{(n)}(k) = 0$. Then $p(C) = A$. And we can also find $q(x)$ such that $q(k) = b_k$ and $q'(k) = 1$ and $q''(k) = ... = q^{(n)}(k) = 0$. Then $q(C) = A$. You can find this polynomial by Hermite interpolation.

The necessity of the second statement is trivial. □

Case 3: A has only One Jordan Block for Each Eigenvalue.

We know that in this case, $AB = BA$ iff B is a polynomial of A. But what about simultaneous Jordanization though? Well, we are unlucky there, as shown in Example 3.50.

Remark 3.55. *Note that "having distinct eigenvalues" = "diagonalizable" + "minimal polynomial has degree n". So Case 1 is the intersection of Case 2 and Case 3.*

Case 4: Simultaneous Triangularization.

In general, $AB = BA$ may NOT implies polynomial relation, or simultaneous Jordanization. However, we can still try to do something nice. We can always simultaneously triangularize.

Lemma 3.56. *Suppose $AB = BA$, then A, B share a common eigenvector (but maybe for different eigenvalues).*

Proof. Take any eigenspace of A, say $\mathrm{Ker}(A - \lambda I)$. Then this is invariant under B, so the restriction $B : \mathrm{Ker}(A - \lambda I) \to \mathrm{Ker}(A - \lambda I)$ must also have an eigenvector, which would be an eigenvector for B. □

Corollary 3.57. *If $AB = BA$, then there is a filtration $\{0\} = V_0 \subsetneq V_1 \subsetneq V_2 \subsetneq ... \subsetneq V_n = V$ by subspaces invariant under both A and B.*

Proof. Find common eigenvector v_1 to span W_1, then consider V/W_1. Find common eigenvector $[v_2]$ in V/W_1, let v_1, v_2 span W_2, then consider V/W_2. So on so forth. Basically just the usual strategy as always. □

Corollary 3.58. *If $AB = BA$, then they can be simultaneously triangularized.*

Can we go one step further? Say maybe we can simultaneously triangularize them, and we are less greedy now, and ONLY turn ONE of them into its Jordan normal form? Unfortunately, this is NOT always possible.

Consider the following two examples where we can actually do this, and we can in fact find ALL the basis that do this.

Example 3.59. *Say* $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ *and* $B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$*. Then* $AB = BA = 0$*. Also note that both matrices are nilpotent. They also both have the same Jordan normal form* $J_1 = \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right)$*, or alternatively* $J_2 = \left(\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 0 & 0 & 1 \\ 0 & 0 & 0 \end{array}\right)$*. They also have the same kernel. Their range are both 1-dimensional.*

Now, under whatever basis, they must not be equal. So if they are simultaneously Jordanized, then one must be J_1 *while the other must be* J_2*. But then* $J_1J_2 \neq J_2J_1$*, which contradict the fact that* $AB = BA$*. So they CANNOT be simultaneously Jordanized.*

Pick v_1 *that span* $\mathrm{Ran}(A)$ *and pick* v_3 *such that* $Av_3 = v_1$*. Since* $Av_3 \neq 0$*,and* $\mathrm{Ker}(A) = \mathrm{Ker}(B)$*, it follows that* $v_2 := Bv_3$ *is non-zero. And under the basis* v_1, v_2, v_3*, we would turn* A *into uppoer triangular* $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ *and* B *into its Jordan canonical form* J_2*.*

Example 3.60. *Consider* $C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$*, and* $A = \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix}$ *and* $B = \begin{pmatrix} 0 & C \\ C & 0 \end{pmatrix}$*. Then obviously* $AB = BA = 0$*, and* A *is already in Jordan normal form. You may check that the Jordan normal form of* B *is simply* A*. So since they are not equal under whatever basis, they cannot simultaneously be* A*, so they cannot simultaneously be Jordanized.*

Suppose we want to change basis simultaneously by some matrix T*, such that* A *is still in JNF but* B *is upper triangular. If afterwards* A *is still in Jordan normal form, then* $TAT^{-1} = A$*.*

Let P *be the* $(2,3)$ *swap matrix. Then* $PAP^{-1} = \begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix}$*. From now on, we change basis throughout. Then* $(PTP^{-1})(PAP^{-1})(PTP^{-1})^{-1} = PTAT^{-1}P^{-1} = PAP^{-1}$*. So* PTP^{-1} *commutes with* $PAP^{-1} = \begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix}$*. Therefore* PTP^{-1} *must be of the form* $\begin{pmatrix} X & Y \\ 0 & X \end{pmatrix}$ *and we can compute its inverse as* $\begin{pmatrix} X^{-1} & -X^{-1}YX^{-1} \\ 0 & X^{-1} \end{pmatrix}$*.*

Now $PBP^{-1} = \begin{pmatrix} 0 & R \\ 0 & 0 \end{pmatrix}$ *where* $R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$*. Then* $P(TBT^1)P^{-1} = (PTP^{-1})(PBP^{-1})(PTP^{-1})^{-1} = \begin{pmatrix} X & Y \\ 0 & X \end{pmatrix}\begin{pmatrix} 0 & R \\ 0 & 0 \end{pmatrix}\begin{pmatrix} X^{-1} & -X^{-1}YX^{-1} \\ 0 & X^{-1} \end{pmatrix} = \begin{pmatrix} 0 & XRX^{-1} \\ 0 & 0 \end{pmatrix}$*. So pick any* X *that upper triangularize* R*, say* $X = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$*. Then* $P(TBT^{-1})P^{-1} = \left(\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$*.*

As a result, we have $TBT^{-1} = P^{-1}\left(\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)P = \left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right)$*.*

So as you can see, you can pick any $T = P\begin{pmatrix} X & Y \\ 0 & X \end{pmatrix}P^{-1}$ *with any* X *upper-triangularizing* R *and any* Y*. These are all possible choices of basis* T*.*

Now consider the next example, where we can do this, but there is no control over WHICH matrix get put into Jordan normal form.

Example 3.61. *Consider* $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ *and* $B = A^2$*. We are already good since* A *is in Jordan normal form and* B *is upper triangular.*

However, there is no way to put B in Jordan normal form while keeping A upper triangular. Suppose TAT^{-1} is upper triangular. Then it is still nilpotent, and its rank is still 2. So $TAT^{-1} = \begin{pmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{pmatrix}$ for some unknown $a, b, c \in \mathbb{C}$. Then $TBT^{-1} = (TAT^{-1})^2 = \begin{pmatrix} 0 & 0 & ac \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. So if A is upper triangular, B may never be in its Jordan normal form.

Finally, consider the following example where this is not possible at all. Whenever one is in JNF, the other cannot be upper triangular.

Example 3.62. *Let $C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ and $D = C^2$ be as in the last example. Let $A = diag(C, D)$ and $B = diag(D, C)$. The intuition is that, in the first block, you cannot Jordanize the D portion withouth ruining the upper triangular structure of C, but in the second block, the situation is reversed. So it turned out that neither can be Jordanized without ruining the other.*

Of course, who knows if there is some super weird change of basis that end up achieving the desired result? To rigorously prove the impossibility, first note that $AB = BA = 0$. Suppose, for contradiction, that we find a basis A^2u, Au, u, Av, v, w, such that A becomes its Jordan normal form $\left(\begin{array}{ccc|cc|c} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$ and B is upper triangular. Note that since $BA = 0$, we see that $B(A^2u) = B(Au) = B(Av) = 0$. So the first, second and forth columns of B must be all 0. Since $AB = 0$, by looking at A, the second, third and fifth rows of B are also all 0. Since B is nilpotent and upper triangular, its entries on the diagonal and below the diagonal are all 0. So B has the form $\begin{pmatrix} 0 & 0 & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$. Then you can check that $B^2 = 0$. However, this is not true. In the original basis, $B^2 = diag(0, C^2) = diag(0, D) \neq 0$. Contradiction

3.6 Geometric and Dynamic Meaning of Commutativity.

Example 3.63 (Things commute if they are disjoint)**.** *Consider elementary matrices E_{ij} where all diagonal entries are one, and the (i, j) entry is one, and all other entries are 0. These are "shearings". Consider $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. It sends rectangles, with sides parallel to the coordinate-lines, into parallelograms of the same height. Draw a few graphic examples and shapes to see this better.*

In general E_{ij} is shearing the i-th coordinate in the direction of the j-coordinate. Then interestingly, E_{ij} commutes with $E_{k\ell}$ if $\{i, j\}$ and $\{k, \ell\}$ are disjoint. Think about it. $E_{12} = \left(\begin{array}{cc|cc} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$ will NOT change the last two coordinates of a vector, while $E_{34} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$ will NOT touch the first two coordinates of *a vector. When you multiply them, i.e., composing them as functions, you see that the non-trivial portion of these matrices simply avoid touching each other. This is the key reason why they commute.*

In general, disjoint geometric or dynamical actions often commute. Think about this: I can rotate the left face of the Rubik's cube, and call this operation L. I can also rotate the right face of the Rubik's cube, and call this operation R. Then you can see that $LR = RL$, because the two operations are disjoint. They avoid each other.

For a more generalized view point, block diagonal matrices commute if the corresponding blocks commute. Consider $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ *and* $\begin{pmatrix} C & 0 \\ 0 & D \end{pmatrix}$*. Then A and D are disjoint actions, and B and C are disjoint actions. In general, blocks at different location are disjoint actions, so they always commute. So it is enough to just consider the pair A and C, and the pair B and D, anc check if they commute.*

The idea of identifying disjoint actions and block diagonalizations are very powerful, as you have already seen. Now we consider another end of the spectrum.

Example 3.64 (Things commute if they have a polynomial relation)**.** *What if $E_{i,j}$ and $E_{k\ell}$ are NOT disjoint? Well, the opposite of being disjoint is being identical. Funnily, you also get commutativity here. If $A = B$, then obviously $AB = BA$ no matter what.*

In fact they don't have to be identical. If B is a polynomial in A, then you have $AB = BA$. Even more general, if you can find C such that A, B are both polynomials in C, then you have $AB = BA$.

Interestingly, polynomials and Jordan blocks are closely related. Consider the following.

Definition 3.65. A matrix $M \in M_{m\times n}$ is ***neat-triangular*** if it is $(0, U)$ or $\begin{pmatrix} U \\ 0 \end{pmatrix}$ for some square upper triangular matrix U, and each "diagonal" of U contains entries of the same value. I.e., say $\begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. (I made up this terminology myself.)

Lemma 3.66. *Suppose $J_1 \in M_m$ and $J_2 \in M_n$ are Jordan blocks, and for a matrix $M \in M_{m\times n}$, we have $J_1M = MJ_2$. Then M is neat triangular.*

Proof. Say $J_1 = aI + N_1$ and $J_2 = bI + N_2$ for some $a, b \in \mathbb{C}$ and nilpotent Jordan blocks N_1, N_2. Then $J_1M = MJ_2$ means $aM + N_1M = bM + MN_2$. So far it is not clear what would happen. But let us tweak things a bit.

Since $J_1M = MJ_2$, it follows that $(J_1 - bI)M = M(J_2 - bI) = MN_2$. Furthermore, since N_2 is nilpotent, $0 = MN_2^n = (J_1 - bI)^n M$. But if $a \neq b$, then $J_1 - bI$ has all eigenvalues equal to $a - b$, so it is invertible. So $0 = (J_1 - bI)^n M$ implies that $M = 0$. So you see, unless $a = b$, otherwise the only matrix to have $J_1M = MJ_2$ is the zero matrix.

Now we consider the case when $a = b$. Then $J_1M = MJ_2$ means $aM + N_1M = bM + MN_2$, and $a = b$ implies that $N_1M = MN_2$ for two nilpotent Jordan blocks N_1, N_2. When a nilpotent Jordan blocks are acting from the left, they shift all rows up, and fill the last row with zeroes. When a nilpotent Jordan blocks are acting from the right, they shift all columns to the right, and fill the left-most column with zeroes. So M shifted right and M shifted up are the same matrix. So M is neat triangular. □

Proposition 3.67. *Suppose $AB = BA$. Then they can be simultaneously Jordanized iff they can be simultaneously block diagonalized, where each block position, one has a single Jordan block while the other is either a same-sized Jordan block or a multiple of identity.*

Proof. The necessity is clear. We only need to show sufficiency.

Say A, B are simultaneously Jordanized, and in the Jordan form $A = diag(J_1, ..., J_k)$. Now, for any $1 \leq i, j \leq k$, we let B_{ij} be the subblock of B corresponding to the block structure of A with the same rows as J_i in A and the same columns as J_j in A. Note that it could be rectangular.

First, since $AB = BA$, we see thatn $J_i B_{ij} = B_{ij} J_i$. So B_{ij} is neat triangular. In particular, unless $J_i = J_j = (a)$ for some $a \in \mathbb{C}$, otherwise the lower left entry of B_{ij} must be zero.

Now we know B is also in Jordan normal form, so each diagonal block B_{ii} are either a multiple of the identity or a single Jordan block of the same form. Suppose some B_{ij} is not entirely 0 for some $i \neq j$, then we must have $j = i + 1$ and $J_i = J_j = (a)$ for some $a \in \mathbb{C}$.

So we see that each non-trivial Jordan block of B must either corresponds to a Jordan block of A in the same location, or corresponds to a diagonal block of A that is a multiple of the identity. $\square$

Corollary 3.68. *Suppose $AB = BA$ and they can be simultaneously Jordanized. Then there is a matrix C such that $A = p(C)$ and $B = q(C)$ for some polynomials $p(x), q(x)$.*

Proof. Same strategy as before. See how we handle the case of $AB = BA$ when A is diagonalizable. $\square$

Geometrically, simultaneous Jordanization is like we are doing a decomposition, and on each piece, either A, B are basically the same (a single Jordan block of the same size), or one of them is trivial (a multiple of the identity). This case is essentially a mixture of "disjoint actions commute" and "identical actions commute". However, there are more polynomial relations than simply being identical. Simultaneous Jordanization implies polynomial relation, but NOT the other way around.

Example 3.69 (Skipping commutativity)**.** *We have seen this already. Let $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ and $B = A^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Then $AB = BA$, but any attempt to put B in normal form would result in A NOT being upper triangular. So they cannot simultaneously be put in normal form.*

This kind of commutativity feels like some sort of "skipping" activity. Consider the standard basis v_1, v_2, v_3. Then A would sends v_3 to v_2, and v_2 to v_1. This is the central geometric feeling of the action of A. What about B? It sends v_3 to v_1, skipping v_2! and it sends v_2 to 0, skipping v_1! Graphically, it feels like this:

$$\begin{tikzcd} v_3 \arrow[r, bend left, "A"] \arrow[rr, bend right, "B"'] & v_2 \arrow[r, bend left, "A"] \arrow[rr, bend right, "B"'] & v_1 \arrow[r, bend left, "A"] & 0 \end{tikzcd}$$

Compare this with a case of disjoint actions. Say $A = \left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$ and $B == \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right)$. Then graphically it feels like this:

$$\begin{tikzcd} v_1 \arrow[r, bend left, "A"] & v_2 \arrow[l, bend left, "A"] & v_3 \arrow[r, bend left, "B"] & v_4 \arrow[l, bend left, "B"] \end{tikzcd}$$

Now if you are neither disjoint, nor identical, and no polynomial relation, then you must be partially entangled. In general, this usually results in non-commuting behavior.

Example 3.70 (Entangled things may not commute)**.** *Consider E_{12} and E_{23}. $E_{12}E_{23} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ while $E_{23}E_{12} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. You can do these calculations easily by seeing them as column operations or row* *operations. Then $E_{12}E_{23}$ means adding the third row to the second and then add the second (which has now been added by the third row) to the first. So in the end, the first row got added by both the second and the third row. On the other hand, $E_{23}E_{12}$ means adding the second row to the first, and then the third row to the second. So the third row never got a chance to get into the first row. Under the standard basis v_1, v_2, v_3, it feels like this:*

$$\begin{tikzcd} v_3 \arrow[r, bend left, "E_{23}"] & v_2 \arrow[r, bend left, "E_{12}"] & v_1 \end{tikzcd}$$

And you can see that they failed to commute because they are "passing the baton". This is pretty much the most common reason to fail commutativity. I change something, and then you carry away the change I made, to further change something else.

However, partially entangled may NOT always results in non-commutativity. Consider the following.

Example 3.71 (Parallel things commute)**.** *Consider E_{12} and E_{13}. They commute, and in fact the two shearings are shearing parallel things! Curiously, E_{23} and E_{13} also commute. The two shearings are also parallel, but in a different meaning. Now we are shearing non-parallel things, but into parallel directions! In fact, you can check that any orbit of E_{23} is a line parallel to any orbit of E_{13}, which is also a line.*

In general, parallel actions commute. Furthermore, above pairs are neither disjoint, nor can you find any polynomial relation.

(Say $A = E_{12}$ and $B = E_{13}$, and let v_1, v_2, v_3 be the standard basis. Suppose there is a matrix C such that $A = p(C)$ and $B = q(C)$. How to show that this is impossible? Well, any eigenvector of C must also be an eigenvector of both A and B. However, all eigenvectors of A are spanned by v_1, v_3, while all eigenvectors of B are spanned by v_1, v_2. Therefore, any eigenvector of C must be in the intersection of these two subspaces. So eigenvectors of C MUST be a multiple of v_1. In particular, all eigenvalues of C are the same, and the corresponding geometric multiplicity must be 1. So under maybe some other basis, the Jordan normal form of C looks like $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$. And any polynomial of C must looks like $\begin{pmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{pmatrix}$. Now, $A - I$ is nilpotent of rank 1. If a polynomial of C is nilpotent, it must looks like $\begin{pmatrix} 0 & b & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix}$. If it has rank 1, then the second and third column are parallel, so in particular $b = 0$ as well. So $A - I$ is a multiple of $\begin{pmatrix} 0 & 0 & c \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. But for the same reason, $B - I$ must also be a multiple of $\begin{pmatrix} 0 & 0 & c \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Then $A - I$ and B_I are multiples of each other, a contradiction.)

Graphically it looks like this:

$$\begin{tikzcd} v_3 \arrow[rr, bend left, "E_{13}"] & v_2 \arrow[r, bend left, "E_{12}"] & v_1 \end{tikzcd}$$

$$\begin{tikzcd} v_3 \arrow[r, bend left, "E_{23}"] \arrow[rr, bend left, "E_{13}"] & v_2 & v_1 \end{tikzcd}$$

As you can see, even though they are entangled, the entanglement is actually quite smooth! The arrow and is always the arrow end, and the arrow source is always the source, there is no "arrow-contradiction" here.

Example 3.72 (Balancing Entanglement)**.** *Parallel actions are NOT the only non-polynomial commuting phenomena. Sometimes there are entanglements that are not suppose to commute. However, they end up balancing each other.*

Consider $A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ *and* $B = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. *You can check that they commute. However, they are non-parallel and there is no polynomial relation.*

(Same deal. Suppose there is a C. Then for the same reason, all eignevalues of C must be multiples of v_1, so under some basis C is just a single Jordan block. Then $A - I, B - I$ are rank 2 and nilpotent, so they look like $\begin{pmatrix} 0 & 0 & a & b \\ 0 & 0 & 0 & a \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. *Then they have the same kernel, contradiction.)*

Graphically, they look like this:

$$\begin{tikzcd} & v_2 \arrow[dl, "A"] & \\ v_1 & & v_4 \arrow[ul, "B"'] \arrow[dl, "A"] \\ & v_3 \arrow[ul, "B"] & \end{tikzcd}$$

As you can see, there are two "passing the baton" phenomena, but they balanced out.

Example 3.73. *Let us also try to understand the above phenomena from yet another perspective, by looking at a related phenomena. Consider the nilpotent version* $A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ *and* $B = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, *and then we symmetrize B so that* $B_s = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$.

Then note that $B_s A B_s^{-1} = A$, and B_s essentially permute the two Jordan blocks of A. Yet $B_s A B_s^{-1} = A$ is equivalent to $B_s A = A B_s$. So we see that they commute.

4 Dual space

At this juncture, you are no longer learning elementary mathematics, or so called "Gao Deng Shu Xue" in Chinese. You should keep a close mind that your goal is NOT to learn, but to understand.

The concept of duality in vector spaces arises from row vectors. You have column vectors, and you have row vectors. You have row operations on matrices, and you have column operations on matrices. You have matrix A, and you could have its transpose A^T. What are these?

Consider the system of linear equations $Ax = b$, where we are trying to solve x. What is the meaning of this? The column view is asking the following: What linear combinations of columns of A will give us b? Think about this. If $A = (v_1, ..., v_n)$ and $x = (x_1, ..., x_n)^T$, then $Ax = \sum x_i v_i$.

What is the row view? Recall that $Ax = b$ can be thought of as a short hand for writing a system of linear equations. Then each row is in fact representing an equation! And each linear equation represent a hyperplane in $\mathbb{R}^n$! So the row view is asking the following: What is the intersection of these hyperplanes?

Now here is a tricky thing to think about. Above I have posed $Ax = b$ as two different questions. These two different questions does not seem to have anything in common in appearances, yet their mathematical formulation and solution are exactly the same! Why? What intuition should you extract from this? I'll leave this to you for the moment.

4.1 Dual and adjoint

Column vectors, row vectors, they have one thing in common: they are all matrices! Let us think about how do they behave as linear map. For the moment, let us restrict our attention to real numbers, where our geometric intuitions are better.

Focus on $\mathbb{R}^3$ for the moment. Say you have $(a,b,c)^T$. This is a column vector, and it is also a 3 by 1 matrix. So it is a linear map from $\mathbb{R}$ to $\mathbb{R}^3$. In particular, column vectors are in one-to-one correspondence with linear maps from $\mathbb{R}$ to $\mathbb{R}^3$. What about a row vector (a,b,c)? It is a 1 by 3 matrix, so it maps $\mathbb{R}^3$ to $\mathbb{R}$. In particular, row vectors are in one-to-one correspondence with linear maps from $\mathbb{R}^3$ to $\mathbb{R}$!

Definition 4.1. For a vector space V, we define its dual V^* to be the vector space of linear maps from V to $\mathbb{R}$. We call vectors in V^* as dual vectors, or linear functionals.

Example 4.2. *The dual of the space of column vectors is the space of row vectors. The converse is true as well.*

Example 4.3. *Take P_n, the space of polynomials of degree at most $n-1$. Take any $a \in \mathbb{C}$, then the map $p \mapsto p(a)$ is a linear functional from P_n to $\mathbb{C}$! The map $p \mapsto p'(a)$ is also a linear functional! In general, evaluating a function at a point, or evaluating some higher order derivatives at a point, these are all "row vectors" for the space of functions.*

Example 4.4. *Let F be the space of nice functions. Above example shows that local evaluations, on functions or their derivatives, are all "dual functions", or linear functionals. On the other hand, consider integration, say $\int_a^b$. For any pair a,b, this is also a linear map from F to $\mathbb{R}$. So global evaluations are also dual functions!*

Remark 4.5. *In general, you may think about dual vectors or linear functionals as evaluations, either local or global. For example, if you want to study the set J of all jobs, then you want to know the salaries of these jobs $p : J \to \mathbb{R}$, the stress level of these jobs $s : J \to \mathbb{R}$, and the working hours of these jobs $h : J \to \mathbb{R}$. Only by knowing these, can you gain an understanding of J.*

Similarly, you cannot TRULY understand a vector space V without knowing its dual V^*. In fact, when we learn about vectors in high school, we are already dealing with V^* ALL the time. Think about it. In high school, when you think about a vector, the FIRST thing you want to figure out are the values of the coordinates. But to go from a vector to its value of one of its coordinates, is in fact a linear map from V to $\mathbb{R}$! You would have to admit, that you are already using dual vectors to understand vectors all along.

We first establish some basic properties.

Proposition 4.6. *For any basis $v_1, ..., v_n$ for a vector space V, there is a unique "dual" basis $v_1^*, ..., v_n^*$ such that $v_i^*(v_j) = 0$ if $i \neq j$ and $v_i^*(v_i) = 1$. In particular, $\dim V^* = \dim V$.*

Proof. Take a basis $v_1, ..., v_n$ for V. Then under this basis, there is a one-to-one correspondence between linear maps from V to $\mathbb{R}$ and 1 by n matrices. So $V^* \simeq \mathbb{R}^n$. The dual basis is made by linear maps that corresponds to the matrices $(1,0,0,...,0), (0,1,0,...,0), ..., (0,...,0,1)$, and it is obviously unique. □

Example 4.7. *Pick a basis $v_1, ..., v_n$, so that we can write each vector in coordinates. Then finding the coordinates of a vector v is the same as applying the dual basis $v_1^*, ..., v_n^*$ to v.*

Example 4.8. *You need to be extra careful here. For each v, there is NO specific vector that is dual to v. Only when you treat v as part of a basis, can you then define a vector to be dual to v.*

Consider $\mathbb{R}^2$ with its usual basis $(1,0)^T, (0,1)^T$. Then a dual basis is $(1,0)$ and $(0,1)$. However, suppose we take the basis $(1,0)^T, (1,1)^T$. Then the dual basis is in fact $(1,-1)$ and $(0,1)$. As you can see, the "dual" of $(1,0)^T$ might change if you put it in the context of a different basis. In particular, I urge you to never say "the dual of a vector", and only say the dual to a basis.

Corollary 4.9. *For a vector $v \in V$, if $\alpha(v) = 0$ for all $\alpha \in V^*$, then $v = 0$.*

Proof. Pick basis $v_1, ..., v_n$. Then $v_i^*(v) = 0$ for all i. So all coordinates of v are 0, and thus $v = 0$. □

Now, if you have the concept of dual spaces, then you have the distinction between rows and columns. So naturally you would have something akin to taking the transpose, that should send some m by n matrix to some n by m matrix.

Proposition 4.10. *There is a one-to-one correspondence between linear maps from V to W and linear maps from W to V. For each $A : V \to W$, the corresponding $A^* : W^* \to V^*$ is called the adjoint of A.*

Proof. Pick $A : V \to W$. Then for each linear functional $\alpha \in W^*$, you can recall that $\alpha : W \to \mathbb{C}$ as a linear map. Then $\alpha \circ A : V \to \mathbb{C}$ would be a linear map from V to $\mathbb{C}$, and thus an element in V^*. We call this the adjoint of A, and write it as A^*. In particular, $A^*(\alpha) = \alpha \circ A$.

(In particular, we are doing a "pull-back". Imagin that V is the set of all people, and W is the set of all jobs, and A is some function that assign each person to a job. Then properties of the job can be "pulled back" to the person via the assignment A. For example, "salary" is a property of jobs, and it is in fact a function $\alpha : W \to \mathbb{R}$. Then A would pull back the salary of a job to the salary of the person, $A^*(\alpha) = \alpha \circ A : V \to \mathbb{R}$.)

Obviously A^* is a fucntion from W^* to V^*. Furthermore, consider $A^*(k\alpha) : V \to \mathbb{R}$. We wish to show that this is in fact $kA^*\alpha$. How to show that two functions are the same? We do this by showing that they are the same for whatever input. Then for each $v \in V$, $A^*(k\alpha)(v) = (k\alpha) \circ Av = k\alpha(Av) = kA^*(\alpha)(v)$. Since this is true for all $v \in V$, we see that as functions $A^*(k\alpha) = kA^*(\alpha)$. Similarly you can verify that $A^*(\alpha + \beta) = A^*\alpha + A^*\beta$. So A^* is linear.

It remains to show that this is a one-to-one correspondence. Suppose $\dim V = n$ and $\dim W = m$. First note that linear maps fro V to W form a vector space of dimension mn (due to the matrix representation). Similarly, linear maps from W^* to V^* would also form a vector space of dimension mn. I now claim that the process $A \mapsto A^*$ is linear and injective. Then it follows that this is in fact bijective and linear for dimensional reason.

Consider $(kA)^*$. Then for any $\alpha \in W^*$ and $v \in V$, $(kA)^*(\alpha)(v) = \alpha(kAv) = k\alpha(Av) = k(A^*\alpha)v = k(A^*)(\alpha)(v)$. So $(kA)^* = kA^*$ Similarly $(A + B)^* = A^* + B^*$.

To see injectivity, suppose $A^* = 0$. Then $A^*(\alpha) = 0$ for all $\alpha \in W^*$. Then $\alpha \circ Av = 0$ for all α and all v. Then Av has all coordinates zero for all v. So $A = 0$. □

Corollary 4.11. $A = 0$ *iff* $A^* = 0$.

Proposition 4.12. $(AB)^* = B^*A^*$.

Proof. Say $A : V \to W$ and $B : U \to V$. Pick any $\alpha \in W^*$ and $u \in U$, then $B^*A^*\alpha(u) = B^*(\alpha \circ A)(u) = (\alpha \circ A \circ B)(u) = (AB)^*\alpha(u)$. □

Example 4.13. *For V and W, take basis $v_1, ..., v_n$ and $w_1, ..., w_m$, and take the corresponding dual basis for V^* and W^*. Now everything can be written in coordinates, and any linear maps are in matrix form.*

Take any linear map $A : V \to W$. What are columns of A? Well, they are $Av_1, ..., Av_n$ written in coordinates according to $w_1, ..., w_m$. And what are the coordinates? They are the result of applying $w_1^*, ..., w_m^*$. In particular, the (i, j) entry of A is in fact $w_i^*Av_j$.

(At the start of the section, I ask you why the two interpretation of $Ax = b$ have the same solution. The answer is associativity, that $(w_i^*A)v_j = w_i^*(Av_j)$. And why do we have assoiativity? The answer is that function composition are always associative.)

Now consider A^*. Its columns are $A^*w_1^*, ..., A^*w_m^*$. For each $\alpha \in V^*$, what are its coordinates in terms of $v_1^*, ..., v_n^*$? We can find these coordinates by evaluating on $v_1, ..., v_n$. So the i-th coordinate of $A^*w_j^*$ is $A^*w_j^*(v_i) = (w_j^* \circ A)(v_i) = w_j^*Av_i$.

The conclusion is this: If you take basis for vector spaces and take the corresponding dual basis for their dual basis, then A and A^* in matrix form are transpose to each other.

Remark 4.14. *If you were in my linear algebra one class, you would recall that back then, we could only define transpose with the help of an inner product. However, here I managed to define transposition withouth inner product?*

However, you must be careful here. Our transposition is NOT the transposition you are used to. For example, for the moment, for any $A : V \to V$, you will never have $A = A^*$ because they will have different domain and codomain! $A^* : V^* \to V^*$ and it has nothing to do with V itself. So you still cannot define symmetric matrix, orthogonal matrix and so on.

In the end, symmetric matrix and so on MUST always depends on an inner product structure. Nevertheless, we have the following interesting results.

Proposition 4.15. *Take any $\alpha \in W^*$ and $v \in V$ and $A : V \to W$. Then $\alpha(Av) = (A^*\alpha)v$. (Compare this with the following: you will always have $v^T(Aw) = (A^T v)^T w$. Here in the formula the T in v^T and the second T in $(A^T v)^T$ are not adjoint, but more like a declaration that we are treating the stuff under it as linear functionals.)*

Proof. This is just the definition of A^*. □

Lemma 4.16 (Separation Lemma)**.** *For any $v \in V$ and proper subspace $W \subsetneq V$ that does NOT contain v, then there is a linear functional α on V such that $\alpha(W) = 0$ but $\alpha(v) \neq 0$.*

Proof. Consider the quotient map $q : V \to V/W$. Then $q(v) \neq 0$. Take any nonzero $\beta \in (V/W)^*$ such that $\beta(q(v)) \neq 0$, and then we should have $(q^*\beta)(v) = \beta(qv) \neq 0$. On the other hand, $q^*(\beta)(W) = \beta \circ q(W) = 0$. So $q^*\beta$ is the desired α. □

Lemma 4.17 (Preservation of Exactness)**.** *If* $\operatorname{Ran}(A) = \operatorname{Ker}(B)$, *then* $\operatorname{Ran}(B^*) = \operatorname{Ker}(A^*)$.

Proof. Say $A : V \to W$ and $B : W \to U$

First you can see that $BA = 0$. Take any $\alpha = B^*\beta$. Then $A^*\alpha = A^*B^*\beta = (BA)^*\beta = 0$. So $\operatorname{Ran}(B^*) \subseteq \operatorname{Ker}(A^*)$.

Now suppose $A^*\alpha = 0$. Then $\alpha(Av) = 0$ for any $v \in V$. So $\operatorname{Ker}(B) = \operatorname{Ran}(A) \subseteq \operatorname{Ker}(\alpha)$. Pick any $w \notin \operatorname{Ker}(\alpha)$, then $Bw \neq 0$. So we should have a linear functional β over U such that $\beta(Bw) \neq 0$ while $\beta(B(\operatorname{Ker}(\alpha))) = 0$. Then in particular, $(B^*\beta)(w) \neq 0$ and $(B^*\beta)(\operatorname{Ker}(\alpha)) = 0$.

Now $B^*\frac{\beta}{\alpha(w)}$ and α agrees on w and on $\operatorname{Ker}(\alpha)$. But w and $\operatorname{Ker}(\alpha)$ spans the whole W! So $\alpha = B^*\frac{\beta}{\alpha(w)} \in \operatorname{Ran}(B^*)$. So $\operatorname{Ker}(A^*) \subseteq \operatorname{Ran}(B^*)$. □

Proposition 4.18. *A is surjective iff A^* is injective, and A is injective iff A^* is surjective. Furthermore, A and A^* have the same rank (i.e. their image have the same dimension.)*

Proof. Pick basis and dual basis. Then A and A^T are transpose of each other, and all is obvious.

Let us now prove this without invoking basis. For any linear map $A : V \to W$, let $e : \operatorname{Ker}(A) \to V$ be the inclusion map, and let $q : W \to W/\operatorname{Ran}(A)$ be the obvious quotient map. Then we have chains:

$$0 \longrightarrow \operatorname{Ker}(A) \xrightarrow{e} V \xrightarrow{A} W \xrightarrow{q} W/\operatorname{Ran}(A) \longrightarrow 0$$

$$0 \longleftarrow \operatorname{Ker}(A)^* \xleftarrow[e^*]{} V^* \xleftarrow[A^*]{} W^* \xleftarrow[q^*]{} (W/\operatorname{Ran}(A))^* \longleftarrow 0$$

Here we did not specify the map related to 0, because there can only be ONE possible map from 0 to any vector space (sending 0 to the origin) or from any vector space to 0 (sending everything to 0).

Note that $A \circ e = 0$ and $q \circ A = 0$. So, on the top chain, you go two step and you always die. Not only so, you can easily check that the image of one map in the chain is ALWAYS EXACTLY the kernel of the next map. Such a sequence is called an exact sequence.

It follows that the bottom sequence is also exact. This is due to the "preservation of exactness" lemma we just proved. So q^* is injective, e^* is surjective, and $\operatorname{Ker}(A^*) = \operatorname{Ran}(q^*)$ and $\operatorname{Ran}(A^*) = \operatorname{Ker}(e^*)$. In particular, $\dim \operatorname{Ran}(A^*) = \dim \operatorname{Ker}(e^*) = \dim V^* - \dim \operatorname{Ker}(A)^* = \dim V - \dim \operatorname{Ker}(A) = \dim \operatorname{Ran}(A)$. □

Remark 4.19. *The exact sequence in fact tell us the following:* $\operatorname{Ker}(A^*)$ *is all about the dual of* $W/\operatorname{Ran}(A)$, *or the structures of W ABOVE* $\operatorname{Ran}(A)$.

Exact sequences are super powerful tools. We shall certainly see them later as well.

There is one last basic property to talk about.

Proposition 4.20. *There is a CANONICAL isomorphism between V and $(V^*)^*$.*

Remark 4.21. *First of all, ALL vector spaces of the same finite dimension are isomorphic. Just pick basis and match up, and you are done. So having an isomorphism is NOT hard.*

However, having a canonical isomorphism is a different matter. Canonical means it behaves well with respect to whatever context, so it is ALWAYS the BEST isomorphism. In this particular setting, canonical means the following: there is a way to assign an isomorphism $e_V : V \to (V^*)^*$ for EACH V, such that for any $A : V \to W$, we have the following commutative diagram (i.e., $e_W \circ A = (A^*)^* \circ e_V$):

$$\begin{tikzcd} V \arrow[r, "A"] \arrow[d, "e_V"'] & W \arrow[d, "e_W"] \\ (V^*)^* \arrow[r, "(A^*)^*"] & (W^*)^* \end{tikzcd}$$

Proof. Pick a basis $v_1, ..., v_n$. Let $w_1, ..., w_n$ be the dual basis of the dual basis of $v_1, ..., v_n$. Then the map sending each v_i to w_i is the desired map. Note that for any $\alpha \in V^*$, $w_i(\alpha)$ essentially means evaluating α at v_i.

For a construction independent of basis, for each v, define $e_V(v) \in (V^*)^*$ to be the evaluation of each $\alpha \in V^*$ to $\alpha(v)$. Then for any nonzero v, there is certainly some linear functional α such that $\alpha(v) \neq 0$, so $e_V(v) \neq 0$. So e_V is injective. You can also check that it is linear, so it is bijective for dimensional reason.

To see that this is canonical, take any $v \in V$ and $\alpha \in W^*$. Then $((A^*)^* e_V(v))(\alpha) = e_V(v)(A^*\alpha) = A^*(\alpha)(v) = \alpha(Av) = e_W(Av)(\alpha)$. So $e_W \circ A = (A^*)^* \circ e_V$ as desired. ☐

Remark 4.22. *Essentially, a canonical isomorphism means that V is not just isomorphic to $(V^*)^*$. It is in fact equal to $(V^*)^*$. From now on, we simply write $(V^*)^* = V$, since the conversion between the two is always unambiguous: we should ALWAYS use the canonical isomorphism.*

Now you see the duality feel. V taking dual is V^*, and V^* taking dual is back to V. Let us see some examples here. We can also say that the dual basis of a dual basis is the original basis, since we are identifying $v_1, ..., v_n$ with $(v_1^*)^*, ..., (v_n^*)^*$.

Example 4.23. *Let P_n be the space of polynomials of degree at most $n-1$. Then evaluations at some specific point is a linear functional. Say $e_1, ..., e_n$ are evaluations at distinct input $a_1, ..., a_n$. I claim that these are linearly independent.*

To see this, let $p_i = \prod_{j \neq i} \frac{x - a_j}{a_i - a_j}$. Then it has degree $n-1$ and takes value 0 at all a_j with $j \neq i$, and $p_i(a_i) = 1$. Then $p_1, ..., p_n$ are all in P_n. They form a basis because any $(n-1)$ of them share a common root, but all n of them has NO common root, hense no polynomial is a linear combination of the rest. You can check that $e_1, ..., e_n$ are then exactly the dual basis to $p_1, ..., p_n$, so they must be linearly independent.

In fact, our Hermite interpolation can also be proven from the perspective of dual spaces. Say we are thinking of P_{mn}, take distinct $a_1, ..., a_n \in \mathbb{C}$, and let e_{ij} be the evaluation $e_{ij}(p) = p^{(j)}(a_i)$ for $1 \leq i \leq n$ and $0 \leq j \leq m-1$. Hermite interpolation claims that these e_{ij} are linearly independent (and hence given $b_{ij} \in \mathbb{C}$, the requirement $e_{ij}(p) = b_{ij}$ has a unique solution).

The dual basis is harder to construct. However, the linear independency is not hard to see. Suppose they have linear dependency, then the intersection of all of there kernels is non-trivial (since it means we have less that mn equations to solve in a mn dimensional space). Say $e_{ij}(p) = 0$ for all i, j. This means that at each a_i, $0 = p(a_i) = p'(a_i) = ... = p^{(m-1)}(a_i)$. So a_i is a root of p with multiplicity at least m. Since this is true for each a_i, we now have found mn roots of p. In particular, p must be a multiple of $\prod(x - a_i)^m$. But p has degree at most $mn-1$. So it follows that $p = 0$. So Hermite interpolation is true.

4.2 Complex dual basis

(V, J) and $(V^*, -J^*)$.

4.3 Riesz Representation Theorem

So far, $A^* = A$ makes no sense due to different domain and codomain. Can we change this somehow? As mentioned before, this would require an inner product structure.

Theorem 4.24 (Riesz representation theorem)**.** *Say V is a vector space with an inner product $\langle , \rangle$. Then for each $v \in V$, the map $\langle v, - \rangle : w \mapsto \langle v, w \rangle$ is a linear functional, and all linear functionals arises this way uniquely.*

Proof. That this is a linear functional is obvious. We only need the second statement.

Take any $\alpha \in V^*$. Take any vector v perpendicular to $\mathrm{Ker}(\alpha)$ under the inner product structure, then $\alpha(v) \neq 0$ since $v \notin \mathrm{Ker}(\alpha)$. Consider $\frac{\alpha(v)v}{\langle v,v \rangle}$. For any $w \in \mathrm{Ker}(\alpha)$, then $\alpha(w) = 0 = \langle \frac{\alpha(v)v}{\langle v,v \rangle}, w \rangle$. And furthermore, $\alpha(v) = \alpha(v)\frac{\langle v,v \rangle}{\langle v,v \rangle} = \langle \frac{\alpha(v)v}{\langle v,v \rangle}, v \rangle$. So α agrees with $\langle \frac{\alpha(v)v}{\langle v,v \rangle}, - \rangle$ on the span of v and $\mathrm{Ker}(\alpha)$, which is all of V. $\square$

Corollary 4.25. *$v = w$ iff $\langle u, v \rangle = \langle u, w \rangle$ for all u.*

As you recall, normally, there is no "canonical dual" to a vector $v \in V$. However, with the help of an inner product, v is now canonically dual to $\langle v, - \rangle$. This gives us a canonical identification between V and V^*.

Now, it might be very tempting to say that for inner product spaces, $V = V^*$. HOWEVER, the canonical identification function is NOT always linear!

Example 4.26. *Consider the complex plane $\mathbb{C}$. Then $\langle w, z \rangle = \overline{w}z$. So in particular, $\langle iw, z \rangle = \bar{i}\overline{w}z = -i\overline{w}z = -i\langle w, z \rangle$. So it is NOT complex linear!*

In general, V and V^* is always the same REAL vector space but with DISTINCT complex structures. In particular, take any vector $v + iw \in V$ where v, w are real. Say $\alpha(u) = \langle v, u \rangle$ and $\beta(u) = \langle w, u \rangle$. Then $\langle v + iw, u \rangle = (\alpha - i\beta)(u)$. As you can see, it is as if the column vector $v + iw$ is identified with the row vector for $v - iw$, like taking a complex conjugate!

Example 4.27. *Pick any (not necessarily orthonormal) basis $v_1, ..., v_n$ for an inner product space V. Then there is a dual basis $v_1^*, ..., v_n^*$ for V^*. But we also know that each functional must in fact looks like $\langle v, - \rangle$ for some unique v. So suppose $v_i^* = \langle w_i, - \rangle$. Then we say the basis $w_1, ..., w_n$ is dual or bi-orthogonal to $v_1, ..., v_n$.*

For bi-orthogonal basis, you always have $\langle w_i, v_j \rangle = \delta_{ij}$, which is 1 if $i = j$ and 0 if $i \neq j$. (These δ_{ij} are very famous and very useful. Why? For starters, the identity matrix is $I = (\delta_{ij})$.)

If you want to write a vector v in terms of coordinates under the basis $v_1, ..., v_n$, you can write $v = \sum \langle v, w_i \rangle v_i$.

In particular, an orthogonal basis is a basis that is dual/bi-orthogonal to itself.

Example 4.28. *Say we take $(1,0)^T, (0,i)^T$ as a basis for $V = \mathbb{C}^2$. Then the dual basis for V^* would be row vectors $(1,0)$ and $(0,-i)$. In particular, you need $(0,-i)$ to send the second vector to 1 and the first to 0. Note that a conjugation has happened. Interesting, yes? Now we identify V and V^* via the dot product. Note that applying $(1,0)$ is like sending $(a,b)^T$ to a, which is like taking dot product with $(1,0)$. But applying $(0,-i)$ is like sending $(a,b)^T$ to $-ib$, which is $\langle (0,i)^T, (a,b)^T \rangle$. So under this identification, we see that $(1,0)$ is identified with $(1,0)^T$, while $(0,-i)$ is identified with $(0,i)^T$. So the basis $(1,0)^T, (0,i)^T$ is bi-orthogonal to itself, and thus it is an orthonormal basis.*

Example 4.29. *Recall that if $A : V \to W$, then $A^* : W^* \to V^*$. Say $\alpha \in W^*$ is $\langle w, - \rangle$. Then what is $A^*(\alpha)$? It must be $\langle v, - \rangle$ for some $v \in V$. So we might call this v as $A^{ad}w$. So far we only know that it is a function, but we also know the following: $\langle A^{ad}w, v \rangle = (A^*\alpha)(v) = \alpha(Av) = \langle w, Av \rangle$. This is the most important property of A^{ad}.*

A^{ad} is in fact a linear map. $\langle A^{ad}(w_1 + w_2), v \rangle = \langle w_1 + w_2, Av \rangle = \langle w_1, Av \rangle + \langle w_2, Av \rangle = \langle A^{ad}w_1, v \rangle + \langle A^{ad}w_2, v \rangle = \langle A^{ad}w_1 + A^{ad}w_2, v \rangle$. Furthermore, $\langle A^{ad}(kw), v \rangle = \langle kw, Av \rangle = \overline{k}\langle w, Av \rangle = \overline{k}\langle A^{ad}w, v \rangle = \langle kA^{ad}w, v \rangle$. So we have complex linearity.

Recall that the identification of w and $\langle w, -\rangle$ is canonical (even though non-linear) between V and V^*. Under this identification, in fact A^* would be exactly the same as A^{ad}, and both are genuin complex linear maps. So in practice, most scholars do NOT distinguish between the two at all, and simply write A^* for both.

In particular, we see that $\langle w, Av\rangle = \langle A^*w, v\rangle$, and now if $A : V \to W$, then $A^* : W \to V$. You can now compose A and A^*. And if $A : V \to V$, then $A^* : V \to V$ as well, and you can define symmetric (or Hermitian in the complex case) as $A = A^*$.

Example 4.30. *Here is a tricky thing. If you take a basis $v_1, ..., v_n$ for V, $w_1, .., w_m$ for W and a dual basis $v_1^*, ..., v_n^*$ for V^* and $w_1^*, ..., w_m^*$ for W^*, then $A : V \to W$ and $A^* : W^* \to V^*$ are the transpose of each other, both over $\mathbb{R}$ and over $\mathbb{C}$.*

However, if you exploit the inner product structure and identify V, W with V^*, W^*, this process would be the same as taking a conjugate. So A^{ad} would in fact be the conjugate transpose, not just the transpose!

Again say we take $(1, 0)^T, (0, i)^T$ as a basis for $V = \mathbb{C}^2$. Then the dual basis for V^* would be row vectors $(1, 0)$ and $(0, -i)$, yet the two columns vector are identified with $(1, 0)^T, (0, i)^T$ via the inner product.

Now take $A : V \to V$ such that $A(1, 0)^T = (i, 1)^T$ and $A(0, i)^T = (1 + i, 1 + i)^T$. In particular, under the standard basis we have $A = \begin{pmatrix} i & 1-i \\ 1 & 1-i \end{pmatrix}$, while under our chose basis of V, the matrix of A is $\begin{pmatrix} i & 1+i \\ -i & 1-i \end{pmatrix}$.

Now consider $A^* : V^* \to V^*$. Then A^* send row vectors to row vectors. Say $v = (a, b)$. Then $[A^*(1, 0)](v) = (1, 0)(Av) = (1, 0)(ai + b - bi, a + b - bi)^T = ia + (1 - i)b = (i, 1 - i)v$, so A^* maps $(1, 0)$ to $i(1, 0) + (1 + i)(0, -i)$. So the first column of A^* is $\begin{pmatrix} i \\ 1+i \end{pmatrix}$.

And similarly $[A^*(0, -i)](v) = (0, -i)(ai + b - bi, a + b - bi)^T = -ia + (-1 - i)b = (-i, -1 - i)v$. So A^* maps $(0, -i)$ to $-i(1, 0) + (1 - i)(0, -i)$ So the second column of A^* is $\begin{pmatrix} -i \\ 1-i \end{pmatrix}$.

So the matrix for $A^* : V^* \to V^*$ under the dual basis is $\begin{pmatrix} i & -i \\ 1+i & 1-i \end{pmatrix}$, which is EXACTLY the transposition of the matrix representation of A!

Now consider $A^{ad} : V \to V$. Then the bi-orthogonal basis now is again $(1, 0)^T, (0, i)^T$. To find $A^{ad}(1, 0)^T$, note that $\langle A^{ad}(1, 0)^T, (a, b)^T\rangle = \langle(1, 0)^T, A(a, b)^T\rangle = \langle(1, 0)^T, (ai + b - bi, a + b - bi)^T\rangle = ai + (1 - i)b = \langle(-i, 1 + i)^T, (a, b)^T\rangle$. So $A^{ad}(1, 0)^T = (-i, 1 + i)^T = -i(1, 0)^T + (1 - i)(0, i)^T$. So the first column of A^{ad} is $\begin{pmatrix} -i \\ 1-i \end{pmatrix}$. Similarly the second column is $\begin{pmatrix} i \\ 1+i \end{pmatrix}$. So the matrix representation of A^{ad} is $\begin{pmatrix} -i & i \\ 1-i & 1+i \end{pmatrix}$, which is EXACTLY the transpose conjugate of the matrix representation of A!

As you can see from this example, writing A^{ad} and A^* in the same way is probably a bad idea and may lead to confusion. Well, the well-accepted notation is not always a good notation, but you just have to adapt sometimes.... The key to untangle this mess is to ALWAYS keep in mind what is the domain and codomain.

Example 4.31. *Adjoint of Adjoint Note that $(A^{ad})^{ad} = A$. You can either take matrix form and check that the conjugate transpose of conjugate transpose is the original matrix, or you can use the definition to play Ping Pong, and see that $\langle v, (A^{ad})^{ad}w\rangle = \langle A^{ad}v, w\rangle = \langle v, Aw\rangle$. (Any Ping Pong phenomena must point out some duality.)*

Example 4.32 (Fundamental Theorem of Linear Algebra)**.** *Now consider $A : V \to W$ for some inner product space V, W and we have $A^* : W^* \to V^*$ and $A^{ad} : W \to V$. Take orthonormal basis, so that A^{ad} is the conjugate transpose of A. We wish to establish the fundamental theorem of linear algebra between A and A^{ad}. Specifically, we want to see that $\dim \mathrm{Ran}(A^*) = \dim \mathrm{Ran}(A^{ad})$, and we want to see that $\mathrm{Ran}(A)$ and $\mathrm{Ker}(A^{ad})$ are orthogonal complements of each other, and $\mathrm{Ran}(A^{ad})$ and $\mathrm{Ker}(A)$ are orthogonal complements of each other.*

We already know that $\dim \mathrm{Ran}(A) = \dim \mathrm{Ran}(A^*)$. And furthermore, A^* and A^{ad} are essentially the same map (they can be translated into each other via a conanical identification of the domain and codomain), therefore $\dim \mathrm{Ran}(A^*) = \dim \mathrm{Ran}(A^{ad})$.

Now let us consider $\mathrm{Ker}(A)$ *and* $\mathrm{Ran}(A^{ad})$. *If* $v \in \mathrm{Ker}(A)$ *and* $u \in \mathrm{Ran}(A^{ad})$, *say* $u = A^{ad}w$, *then* $\langle u, v\rangle = \langle A^{ad}w, v\rangle = \langle w, Av\rangle = \langle w, 0\rangle = 0$. *So the orthogonal relation is clear. Furthermore,* $\dim \mathrm{Ker}(A) + \dim \mathrm{Ran}(A^{ad}) = \dim \mathrm{Ker}(A) + \dim \mathrm{Ran}(A) = \dim V$. *So we are done. The other orthogonality statement can be proven similarly.*

Example 4.33. *Many infinite dimensional versions of Riesz Representation Theorem exist, and they are quite useful.*

Suppose we are trying to solve the differential equation $-f''(x) + b(x)f(x) = q(x)$, *given the initial condition* $f'(0) = f'(1) = 0$. *For any nice function* $\phi(x)$, *integration by parts gives* $\int_0^1 f''(x)\phi(x)dx = -\int_0^1 f'(x)\phi'(x)dx$. *On the other hand, we also have* $\int_0^1 f''(x)\phi(x)dx = \int_0^1 b(x)f(x)\phi(x)dx - \int_0^1 q(x)\phi(x)dx$. *Rearranging terms, we have* $\int_0^1 (f'(x)\phi'(x) + b(x)f(x)\phi(x))dx = \int_0^1 q(x)\phi(x)dx$. *Note that this should be true for all nice functions* $\phi(x)$.

Now you need to realize the following things. First, the pairing defined as $\langle f, \phi\rangle = \int_0^1 (f'(x)\phi'(x) + b(x)f(x)\phi(x))dx$ *is actually an inner product! Next, the operator* T *such that* $T(\phi) = \int_0^1 q(x)\phi(x)dx$ *is a dual vector. Then by a fancy Riesz Representation Theorem, any "nice" dual vector must be in fact represented by taking inner product with sometime. So you can find a function* f_0 *such that* $T(\phi) = \langle f_0, \phi\rangle$ *for all* ϕ. *So this* f_0 *is your solution.*

4.4 Tangent vectors and cotangent vectors

Consider a paraboloid S defined as $x^2 + y^2 = z$ in $\mathbb{R}^3$. Take any point, say $p = (0,0,0)^T \in P$. What is the tangent plane of S at p? The intuitive definition is this: imagin that you live ON this paraboloid, and cannot move out of it. Now, you are standing at the point p, and you must remain on the paraboloid, then which direction can you move? The collection of all such directions form a vector space, i.e., the tangent plane. In this case, we see that T_p is the span of $(1,0,0)^T$ and $(0,1,0)^T$. If you take $p = (1,0,1)^T$, then you can check that T_p is the span of $(1,0,2)^T$ and $(0,1,0)^T$.

Remark 4.34. *How to rigorously define this? Well, we want to collect all possible movements at* p. *So the key is to look at all CURVES through* p. *Let* $\mathcal{C}$ *be the set of all differentiable curves* $r(t)$ *with* $r(0) = p$. *Next we define an equivalence relation. We say that two curves* $r(t)$ *and* $s(t)$ *are equivalent,* $r \sim s$, *iff* $r'(0) = s'(0)$. *Then the quotient* $\mathcal{C}/\sim$ *will naturally be of all possible derivatives of curves at* p, *i.e., all possible velocities at* p. *This is exactly what we want. And it is obvious that all possible* $r'(0)$ *would form a 2-dim subspace* T_p *of* $\mathbb{R}^3$. *(You can translate this subspace so that the origin is at* p. *It would look prettier. Or you can be more practical and just treat* T_p *as some abstract 2-dim vector space (even independent of* $\mathbb{R}^3$*), which is the more useful view point.) This* T_p *is defined to be the tangent plane at* p. *You can feel free to generalize this to arbitrary dimension.*

Example 4.35.

1. *(1-dim obj in* $\mathbb{R}^2$*) For the unit circle on the plane, its tangent lines at* $(\sqrt{2}/2, \sqrt{2}/2)^T$ *is spanned by* $(1,-1)^T$.
2. *(2-dim obj in* $\mathbb{R}^3$*) For the unit sphere, its tangent plane at* $(1,0,0)^T$ *is spanned by* $(0,1,0)^T$ *and* $(0,0,1)^T$.
3. *(3-dim obj in* $\mathbb{R}^4$*) For the unit 3-sphere in* $\mathbb{R}^4$, *defined as* $w^2 + x^2 + y^2 + z^2 = 1$, *then its tangent space at* $(1,0,0,0)^T$ *is spanned by* $(0,1,0,0)^T$, $(0,0,1,0)^T$ *and* $(0,0,0,1)^T$.
4. *(1-dim obj in* $\mathbb{R}^3$*) For the curve* $t \mapsto (t, t^2, t^3)^T$, *then it has tangent lines. Its tangent line at* $(1,1,1)^T$ *is spanned by* $(1,2,3)^T$.
5. *(2-dim obj in* $\mathbb{R}^2$*) For the plane* $\mathbb{R}^2$ *itself, the for any point* $p \in \mathbb{R}^2$, *then the tangent plane is the whole* $\mathbb{R}^2$. *However, we prefer to think of these as distinct vector spaces. Even though they all have the same linear structure as* $\mathbb{R}^2$, *their MEANINGS are different! They represent possible movement directions at DIFFERENT points, and therefore they are different. So do NOT write them as* $\mathbb{R}^2$. *It is always*

preferable to write them simply as T_p for various p, and write $T_p \simeq \mathbb{R}^2$ which says that the two has the same linear structure.

As you can see, even though we say T_p is spanned by something blahblah, you should keep in mind that T_p is NOT living in the underlying space $\mathbb{R}^n$. It lives in a different world. So geometrically and intuitively, how can one envision this?

(I live on earth, and we all know that the earth is pretty much a sphere. However, assume that I am a caveman. I don't know the global structure of the earth and what not. All I know is that the ground around me seems to be flat. So I would guess that the world is completely flat and infinite, like a 2-dim vector space. This understanding is in fact quite useful, as long as I only ever walk aroung my cave.)

Think about a small area around p (called a neighborhood) on the sphere S. As I take smaller and smaller neighborhood, inside the neighborhood, the sphere would look flatter and flatter! You can imagine that, taking limit of this process, an infinitesimally tiny neighborhood around p would be basically a flat disk! I take this tiny flat disc and stretch it to infinity, and now you have another construction of the tangent plane! Recall that we originally describe tangent vectors as the "intention" to move. The idea here is that the "intention" to move in some direction is interpreted as a super tiny actual movement inside my infinitesimal neighborhood.

Now we move on to cotangent vectors.

Definition 4.36. At a point p on some geometric object S, let T_p be the tangent space. Then a cotangent vector (or covector for short) is an element of T_p^*, the dual space of the tangent space. (If we write elements of T_p as column vectors, then elements of T_p^* should be row vectors.)

So we see the definition, but what does it mean? A covector obviously is something that evaluate tangent vectors. Recall now that the tangent space can be thought of as the infinitesimal neighborhood around the point p. So a covector should be something that evaluate the infinitesimal neighborhood around the point p.

Example 4.37 (Induced covectors)**.** *Pick any differentiable function f from some geometric object S to $\mathbb{R}$, i.e., it evaluates everything on S. Then obviously it should evaluate the infinitesimal neighborhood of each $p \in S$ as well. In particular, it should induces a covector at each point of S.*

Fix any $p \in S$. Then how does f induce a covector? Well, suppose we input a tangent vector v, say achieved by some differentiable curve $r : \mathbb{R} \to S$ with $r(0) = p$ and $r'(0) = v$. Then since f is differentiable as well, we can compose and get a differentiable function $(f \circ r) : \mathbb{R} \to \mathbb{R}$, and we can take derivative as usual. Then we get a value $(f \circ r)'(0)$. So the covector at p induced by f should send v to $(f \circ r)'(0)$.

But what is this value $(f \circ r)'(0)$? It means that if we stand at p, and move a little bit along the curve r, then how much would the value of f change? So in particular, this is the DIRECTIONAL DERIVATIVE of f in the direction of v, i.e., $D_v f$. So this covector sends each v to the directional derivative $D_v f$.

Example 4.38 (Meaning of infinitesimal equations)**.** *From calculus, I think you should know the formula for directional derivatives. Say $S = \mathbb{R}^2$. Then for any $f : \mathbb{R}^2 \to \mathbb{R}$, we see that $D_v f = \begin{pmatrix} \frac{\partial f}{\partial x}(p) \\ \frac{\partial f}{\partial y}(p) \end{pmatrix} \cdot v = (\frac{\partial f}{\partial x}(p), \frac{\partial f}{\partial y}(p))v$. So the linear functional is identical with the matrix $(\frac{\partial f}{\partial x}(p), \frac{\partial f}{\partial y}(p))$. Huh, it is the gradient transposed!*

As you can see, the gradient should NOT really be a vector. It is suppose to be a CO-vector and a row vector. We usually write the covector as $df = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$.

Now consider the function $x : \mathbb{R}^2 \to \mathbb{R}$ and $y : \mathbb{R}^2 \to \mathbb{R}$ that sends points to their x-coordinates and y-coordinates. You can check that $dx = (1, 0) \in T_p^*$ at all points and $dy = (0, 1) \in T_p^*$ at all points $p \in \mathbb{R}^2$. So in particular, we see that $df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$ is true for all points $p \in \mathbb{R}^2$. Seems familiar?

Also note the following special thing: the coordinates of df are exactly provided by $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$! So these are the "dual" of the dual vectors, and hence they are the original tangent vectors.

In geometry, tangent vectors can also be defined as "derivatives". For each tangent vector v, you can identify it with the directional derivative D_v. These for the tangent space. The covectors are "differentials" like dx, dy, dz, and they represents coordinates for tangent vectors.

Example 4.39 (Covector fields)**.** *A vector field on a geometric object S is an assignment of a tangent vector to each $p \in S$. This is familair. However, above examples shows a COvector field. For each p, we assigned a covector $df_p \in T_p^*$ induced from $f : \mathbb{R}^2 \to \mathbb{R}$.*

When the geometric object you study is in fact some entire $\mathbb{R}^n$, then the covectors induced by the coordinate functions in fact span all possible covectors! (As shown by the formula $df = f_x dx + f_y dy$, which is true for ALL functions.)

And indeed, the tangent spaces T_p for $\mathbb{R}^n$ are all n-dim, and thus the cotangent spaces are all n-dim as well. They SHOULD be spanned by n linearly independent vectors.

The case is more complicated if, say, the geometric domain is NOT $\mathbb{R}^n$, but, say, some sphere. Say S is a unit sphere in $\mathbb{R}^3$. Then there are three possible coordinate functions to take, i.e., $x, y, z : S \to \mathbb{R}$, and they induce covector fields dx, dy, dz. At all points of S, dx, dy, dz are NOT linearly independent. For example, at $p = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})^T$, the tangent space T_p is the plane $x+y+z = 0$ (you can easily see that $x+y+z = 1$ touches the sphere at this point in a tangential way, and then we move this plane to the origin to get a vector space). Now for each tangent vector $(x, y, z)^T$, then the coordinate functions should evaluate the change in the corresponding coordinates along this tangent vector. So $dx[(x, y, z)^T] = x, dy[(x, y, z)^T] = y, dz[(x, y, z)^T] = z$, and they sum up to zero for all input tangent vectors. In particular, we have a relation $dx + dy + dz = 0$, a linear dependence!

In fact, from the sphere equation $x^2 + y^2 + z^2 = 1$, you see that $2xdx + 2ydy + 2zdy = 0$, so in fact you always have $xdx + ydy + zdz = 0$! So the three covectors are never linearly independent.

On the other hand, any two of the three is NOT enough! Say we take dx and dy. Then consider $p = (1, 0, 0)^T$. The tangent space T_p is spanned by $(0, 1, 0)^T$ and $(0, 0, 1)^T$. Note that we have $dx_p = 0$! So at p, dx_p and dy_p would FAIL to span the space T_p^*. In total there are four such failures, at $p = (\pm 1, 0, 0)^T$ or $(0, \pm 1, 0)^T$. We are safe at other points though.

It is not always this bad though. Try the case when S is a paraboloid, and you will see that again dx, dy span the covector space at each point.

Example 4.40. *In particular, you see that $\int f dx + g dy = (\int f dx) + (\int g dy)$ and so on. These operations are admissible because all evaluations involved are linear. You can also verify that $d(xy) = xdy + ydx$ and so on.*

Example 4.41 (Covector fields are not necessarily closed)**.** *A covector field on S is closed if it is df for some $f : S \to \mathbb{R}$. This is NOT always the case, even for $S = \mathbb{R}^2$. Consider $ydx - xdy$. If it is df, then $y = f_x$ and $-x = f_y$. Then $f_{xy} \neq f_{yx}$, contradiction.*

Example 4.42 (Covector fields in physics)**.** *Imagine a gravitational field. Is it a vector field? Well, to be precise, It is actually a covector field. Say you move some mass against your gravitational field, then the field would send your movements to the work you need to do (to overcome gravity). This is an evaluation of tangent vectors.*

And at each point, the gravitational force induced by your field, is in fact a covector. In fact, since all forces are needed to take "dot" products with velocities to calculate work, a scalar. You see that all forces are covectors, not vectors.

Funny thing. If forces are covectors, then what are accelerations? Well, being the derivative of velocity, we see that accelerations are natrually vectors, not covectors. But then what about the formula $F = ma$? A moment of truth. Mass is not just a number. In classical physics, mass is like a "scaled" transpose operation, that sends vectors to covectors. In particular, momentum, mv, are covectors as well, and applying this covector to velocity you get double the kinetic energy $\frac{1}{2}mv^2$, which is a scalar number.

(Actually, the formula $F = ma$ is technically incorrect. The correct formula is $F_i = \sum_j m g_{ij} a^j$, where a^j means the j-th coordinate of a, as per Einstein's notation. In classical physics, we have $g_{ij} = \delta_{ij}$, and thus one might simplify this to $F = ma$. In relativity, g_{ij} might be distorted in various ways. As you can see, mg_{ij} for various i, j would actually form a matrix like object.)

Now we move on to integrations and derivatives.

Example 4.43 (Onion structure of covector fields)**.** *Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$. Then it has level curves, curves obtained by equations $f(x, y) = a$ for fixed a. Say $f(x, y) = x^2 + y^2$, then the level curves are concentric circles around the origin. Now consider $f : \mathbb{R}^3 \to \mathbb{R}$ with $f(x, y, z) = x^2 + y^2 - z$, then we have level surfaces that are various paraboloids. Now the claim is this: the level curves/surfaces/hypersurfaces of f is describing the kernel of df. (Hypersurfaces in an n-dim object are $(n-1)$-dim objects.)*

Indeed, if you move in a tangent direction of the level hypersurface, you are moving in a tangent direction v where f does NOT change. Then obviously $df(v) = 0$. (Recall that $df(v)$ is the directional derivative in the direction of v.)

In your mind, you should see this picture: with these level hypersurfaces, the domain would now look like an onion, with these hyper surfaces as onion layers. In fact, for ANY continuous covector fields, its kernel at various points must change continuously, so they "connect" to some continuous hypersurfaces, and give your domain an onion structure.

Furthermore, consider this. If you know the kernel of a linear functional $\alpha \in V^*$, then you know its "direction", and all you need to know now is the length. So for a covector fields in the domain, the onion layer structure gives you the directions of all covectors involved, and you can interpret the length of these covectors as "densities" of these onion layers. For example, consider xdy on $\mathbb{R}^2$. Then all onion layers are horizontal, but the more to the right, the denser these layers become.

What are line integrals? Given a curve, you ask yourself: how many onion layers did you puncture through? This is the definition line integral. In particular, if we have $f : S \to \mathbb{R}$ and we integrate df from $p \in S$ to $q \in S$ along some curves, then how many layers of df did we punctured through? Well, the layers for df are the level curves, and from p to q you punctured a total of $f(q) - f(p)$ units of level curves. So we have proven that $\int_p^q df = f(q) - f(p)$, the fundamental theorem of calculus.

(Of course, you also need to be careful about orientations. Consider $f(x, y) = x^2 + y^2$ and the closed curve that is the unit square through the covector field df. For everything you punctured, you later "un-punctured" them, so they add up to zero.)

Example 4.44 (Path integral)**.** *The path integral $\int f ds$ is different from above. Given some curve C in $\mathbb{R}^2$ and $f : \mathbb{R}^2 \to \mathbb{R}$, we will NOT treat $\mathbb{R}^2$ as our geometric object to study. Instead, we focus on the curve, and only consider the restriction $f : C \to \mathbb{R}$.*

The curve C is some 1-dim object. Its tangent space and cotangent space are both 1-dim, i.e., they are tangent lines and cotangent lines. So pick the unit covector at each cotangent line T_p^*, this form a unit covector field ds on C.

Recall that for vector-valued functions $\begin{pmatrix} f \\ g \end{pmatrix}$, we treat it as the covector field $fdx + gdy$, and evaluate this on tangent vectors along some curve. Similarly, for our $f : C \to \mathbb{R}$, we treat it as the covector field fds where ds is the unit covector field on C, and we evaluate this on tangent vectors along the curve C itself. This gives $\int f ds$. (Since C is uglier than a flat Euclidean space like $\mathbb{R}^n$, therefore path integrals are uglier to compute, as per your experience might suggest.)

We hereby note some properties to DISTINGUISH tangent vectors and covectors. It starts with push-forwards and pullbacks.

Example 4.45 (Motivating example)**.** *As a motivating example, Curves are pushed forward while functions are pulled back. Consider a continuous map $f : S \to M$ between geometric objects. Then for any curve $r : \mathbb{R} \to S$, it is PUSHed to a curve $(f \circ r) : \mathbb{R} \to M$. We also write this as $f_*(r)$, the pushforward of r.*

(Let S be the Tsinghua University and let M be a hand-drawn map of Tsinghua university. There is a correspondence $f : S \to M$, where some spots (inside of a building) are collapsed to the same image and some distances are distorted and so on. Then if I walk around physically along some curve r in S, it corresponds to some curves on the map. This curves on the map is $f_*(r)$.)

Conversely, functions are pulled back. Consider a function $g : M \to \mathbb{R}$, then it is pulled back to $g \circ f : S \to \mathbb{R}$. We usually write $f^*(g)$ for the pull-back of g via f.

(Say you are single and want to meet cute boys/girls. You come to me for help, and I say don't worry, I have a secret magic map. Then I give you the map M where each point p is marked with the chance $g(p)$ to

meet cute boys/girls. Then if you are physically somewhere in S, you can check the marks on the map M and therefore see your chance. Via this process, you are pulling g back to $f^*(g) = g \circ f$.)

Example 4.46 (Tangents are pushed and Cotangents are pulled). *If $f : S \to M$ is a function between smooth geometric objects, then for any curve $r : \mathbb{R} \to S$, we have a curve $(f \circ r)\mathbb{R} \to M$. In particular, tangent vectors on S would be mapped to tangent vectors of M. So we have a map $Df : T_pS \to T_{f(p)}M$ for each $p \in S$. This is the definition of derivative! (Indeed, at a higher level, a function is called differentiable if it induces these linear maps Df between all tangent spaces.)*

The slogan is this: Tangent vectors are pushed forward, and the push-forward map is called the derivative. Think about $f : \mathbb{R} \to \mathbb{R}$. If you have a tangent vector Δx at p in the domain, which is some tendency to move, then after mapped by f, your tendency to move is indeed $f'(p)\Delta x$.

In short, if you move in a smaller space, and the smaller space is mapped to be a part of a larger space, then you automatically moved in the larger space. Such induced movement is the push-forward of tangent vectors.

Conversely, covectors are pulled back. For $f : S \to M$ and some covector $\alpha \in (T_{f(p)}M)^*$, and for any tangent vector $v \in T_pS$, we can define $f^*(\alpha)$ as the evaluation $f^*(\alpha)(v) = \alpha(Df(v))$. If you pushed something over, then you pulled its evaluation back. In particular, suppose f is mapping some smaller object into some larger object. Then any evaluation of the larger object is automatically an evaluation of the smaller object.

Let us see chain rule as an example. Suppose we have $r : \mathbb{R} \to \mathbb{R}^3$ and $f : \mathbb{R}^3 \to \mathbb{R}$. Then $df = f_x dx + f_y dy + f_z dz$ at east point $p \in \mathbb{R}^3$. Note that at each point p, $f_x(p), f_y(p), f_z(p)$ are constants while dx_p, dy_p, dz_p are the covectors. So in particular, $r^*(df) = f_x r^*(dx) + f_y r^*(dy) + f_z r^*(dz)$. Now let dt be the unit covector field on $\mathbb{R}$, the domain of r. Then for any input unit tangent vectors v at $t \in \mathbb{R}$, then Dr maps v to $r'(t)$ which is evaluated by dx, dy, dz to the corresponding coordinates of $r'(t)$. But since $r = (x(t), y(t), z(t))^T$ by definition of coordinate functions, we see that $r'(t) = (x'(t), y'(t), z'(t))^T$. So $r^*(dx) = \frac{dx}{dt}dt$ and so on. So we have proven the chain rule for multivariable functions.

As a result of the above phenomena, we have the following interesting criteria to see if something is a tangent vector or a cotangent vector.

Example 4.47 (Contravariant vectors and Covariant vectors.). *Take a vector space $V = \mathbb{R}^3$. I have the standard basis, which is fine. Now suppose I pick a new basis $v_1, v_2, v_3 \in V$. I express them in standard basis and let $B = (v_1, v_2, v_3)$ be the matrix. Then what is B? It sends $(1,0,0)^T, (0,1,0)^T, (0,0,1)^T$ to v_1, v_2, v_3, so it is the change of basis matrix from the new basis to the standard basis. In particular, if you want to switch to the new basis, you apply B^{-1}. So if $v = (a_1, a_2, a_3)^T$ in the standard basis, then $v = B^{-1}(a_1, a_2, a_3)^T$ in the new basis. Since the matrix applied is the inverse, these are also called contravariant vectors. Tangent vectors are contravariant.*

In comparison, take $\alpha \in V^*$. V^* is natually endowed with the standard basis $(1,0,0), (0,1,0), (0,0,1)$, i.e., the dual basis of the standard basis. Now suppose in V I change the standard basis to $v_1, v_2, v_3 \in V$. Then in the new dual basis v_1^*, v_2^*, v_3^*, what is α?

Again express v_1, v_2, v_3 in standard basis and let $B = (v_1, v_2, v_3)$ be the matrix. Then $\alpha_{st}(v_{st}) = \alpha_{new}(v_{new}) = \alpha_{new}(B^{-1}v_{st})$. In particular, $\alpha_{new}B^{-1} = \alpha_{st}$, and thus $\alpha_{new} = \alpha_{st}B$. As you can see, if you have α in the standard dual basis, you merely apply B afterwards to get α in the new dual basis. Convenient, yes? Since we now only apply B directly, dual vectors are also called covariant vectors. Cotangent vectors are covariant.

Given a vector, say in physics, you can always perform a change of basis of the physical space, and see whether B or B^{-1} is involved, and you will be able to tell if you have a vector or covector.

What about force? When you do $F = ma$, what you really mean is that $F^T = ma$. Then an ordinary change of basis B would change the acceleration contravariantly, and thus the new F^T should be $B^{-1}F^T$. On the other hand, since force is a covector, the new F should be FB. So $B^{-1}F^T = (FB)^T = B^TF^T$. So if you perform an ORTHOGONAL change of basis B, then taking transpose, you can treat F^T as a regular vector and treat it contravariantly, or treat F as a covector and treat it covariantly. However, if you perform a non-orthogonal change of basis, you will see that the force is innately covariant rather than contravariant. $B^{-1}F^T$ will NOT be the new force coordinates, and only B^TF^T would work.

Say I'm in $\mathbb{R}^2$ and I'm falling down (minus y direction) due to gravity alone, with a current velocity $(1,-1)^T$. Then $a=(0,-g)^T$ and $F=(0,-g)$, and the current power done by gravity is $Fv=g$, and the derivative of the power is $Fa=g^2$ at this moment.

Take a new basis $B=(v_1,v_2)=\begin{pmatrix}1 & 1\\ 0 & -1\end{pmatrix}$. Now my current velocity is $(0,1)^T$, and my current acceleration is $(-g,g)^T$. The current power and its derivative should NOT change after a change of basis, so if the new F is (a,b), then we have $Fv=b=g$ and $Fa=-ag+bg=g^2$. You can solve and see that the new force HAS to be $F=(0,g)$.

Now check that $B^{-1}F^T=(-g,g)^T$, while $B^TF^T=(0,g)^T$. We see that only the second one is correct, and that force is covariant rather than contravariant.

4.5 Tangent space and exponential map

Example 4.48. *Consider O_n, the collection of all n by n real orthogonal matrices. This is a subset of $\mathbb{R}^{n\times n}$, and thus it is a geometric object. What is its tangent spaces?*

Let us first find out about $T_I(O_n)$, its tangent space at I. Then you are asking the following question: infinitesimally speaking, when is $I+tA$ still orthogonal for very tiny t?

Suppose t is infinitescimally small. Then we want $(I+tA)(I+tA)^T=I$. In particular, we want $I+t(A+A^T)+t^2AA^T=I$. Note that t^2 is insignificant compared to t, so we treat t^2 as zero, and we see that we want $A+A^T=0$, i.e., skew-symmetric matrices. So the tangent space to O_n at I is the space of skew-symmetric matrices. In particular, we see that O_n is some $\frac{n(n-1)}{2}$ dimensional surface, since its tangent space is $\frac{n(n-1)}{2}$ dimensional.

In general, the tangent space at $P\in O_n$ should have $(P+tX)(P+tX)^T=I$, whcih gives $XP^T+PX^T=0$, i.e., $X=AP$ for some skew-symmetric A. So the tangent space is still $\frac{n(n-1)}{2}$ dimensional. In particular, "right multiply by the matrix P" would in some sense identify $T_I(O_n)$ with $T_P(O_n)$. We say the tangent vector AP at P is a parallel translate of the tangent vector A at I.

(Say I walk north. At a different location, I can also walk north. Then the two north are parallel translates of each other. I can parallel translate "north" to everywhere on $\mathbb{R}^2$, and think of all these tangent vectors as the same DIRECTION.)

Example 4.49. *Now let us move on th exponential maps. Recall that $e^{A+B}=e^Ae^B$ if $AB=BA$. So in particular, for a skew symmetric matrix A, we have $I=e^{A+A^T}=e^Ae^{A^T}=e^A(e^A)^T$. So the exponential map sends a skew-symetric matrix to an orthogonal matrix.*

Now the key is this: consider the curve e^{At} on O_n. At each point P on the curve, if $P=e^{At}$, then the corresponding tangent vector here is $\frac{d}{dt}e^{At}=Ae^{At}=AP$. So this is a curve that keeps moving along parallel translated directions. In a sense, you picked direction, and you go in that direction forever.

In geometry, the exponential map is defined like this: it always sends some tangent space T_p of S to S. For each tangent vector v at p, $exp(tv)$ means you walk in the direction of v for time t, and you end up at $exp(tv)\in S$.

Example 4.50. *For matrices, we can define length as $||A||=\sqrt{\mathrm{tr}(A^TA)}$. Then distances is defined as $d(A,B)=||A-B||=\sqrt{\mathrm{tr}((A-B)^T(A-B))}$.*

What is this length? Well, the (i,i) entry of A^TA is the dot product of the i-th column of A with itself, which is the sum of squares of entries in the i-th column. You add this up for all i, and you will have the sum of squares of ALL entries of A. And then you take square root. Huh. In fact $||A||$ is the Euclidean length of A in the Euclidean space $\mathbb{R}^{n\times n}$. So our "definition" here merely points out that this length is invariant under any change of basis.

This also induces an inner product on the space of matrices, i.e., $\langle A,B\rangle=\mathrm{tr}(A^TB)$. It also equals to the sum of all products of corresponding entries, like an ordinary dot product on $\mathbb{R}^{n\times n}$.

Now consider the curve $r(t)=e^{At}$ on the geometric object O_n. Then $r'(t)=Ae^{At}$ and $r''(t)=A^2e^{At}$. I claim that $r''(t)$ is always normal to the geometric object O_n, so that the curve $r(t)$ is indeed a "straight line" on O_n, that it goes in one direction on O_n and never change direction.

To see this, for each t, pick any tangent vector Be^{At} at $e^{At} \in O_n$ for some skew-symmetric B. Then $\langle r''(t), Be^{At}\rangle = \operatorname{tr}(e^{-At}A^TA^TBe^{At}) = \operatorname{tr}(A^TA^TB) = \operatorname{tr}(A^2B)$. On the other hand, $\operatorname{tr}(A^2B) = \operatorname{tr}(B^TA^TA^T) = -\operatorname{tr}(BA^2) = -\operatorname{tr}(A^2B)$, where the last equality follows from the basic fact that $\operatorname{tr}(XY) = \operatorname{tr}(YX)$. As a result, $\operatorname{tr}(A^2B) = -\operatorname{tr}(A^2B)$, so both must be 0. So indeed $r''(t)$ is perpendicular to all tangent vectors, and thus perpendicular to O_n.

Now consider some reasonably small (but not infinitesimal) $d > 0$, and the curve e^{tA} from I to e^{dA}. The tangent vectors are Ae^{tA} at each e^{tA}. So the length of this curve is $\int_0^d ||Ae^{tA}||dt = \int_0^d \sqrt{\operatorname{tr}(e^{-tA}A^TAe^{tA})}dt = \int_0^d \sqrt{\operatorname{tr}(A^TA)}dt = \int_0^d ||A||dt = d||A||$. So the distance of from I to e^{dA} ON THE "SURFACE" O_n is $d||A||$. Note that this is NOT the same as $d(I, e^{dA})$ inside $\mathbb{R}^{n\times n}$.

5 Tensor

5.1 Motivation and Definition

Example 5.1 (3D Matrices?)**.** *Let us first see some example of tensors that you already know. To start, what is a 0-tensor? It is just a number, say elements of $\mathbb{R}$. You just have some number a.*

What is a 0-tensor? This is just a scalar $\mathbb{R}$.

What is a 1-tensor? That would be a vector, like elements of $\mathbb{R}^n$. Note that for 1-tensors, each element is a list of numbers, or a 1-dim array of numbers. For each index i you would need an entry for your 1-tensor, and your 1-tensor can be said as $(a_i)_{1\leq i\leq n}$.

What is a 2-tensor? Well, that is what we've been studying all along. They are matrices, like M_n, and they also form a vector space of dimension n^2. Note that this is a 2-dim array of numbers. For each pair of indeces i, j you would need an entry for your 2-tensor, and your 2-tensor can be said as $(a_{ij})_{1\leq i,j\leq n}$.

Now, the following thought probably have already appeared in your mind: What if we have a 3-dim array of numbers? Think about it. By looking at 2-tensors (i.e., matrices), we already have a vast number of theories and applications. It makes sense that a 3-dim array of numbers are probably more complicated and more powerful to study. These are 3-tensors. For any triple of indices i, j, k, you would need an entry for your 3-tensor, and your 3-tensor can be described as $(a_{ijk})_{1\leq i,j,k\leq n}$.

You can similarly define n-tensors. They are n-dim array of numbers.

As you can see, our whole linear algebra 1 and most of linear algebra 2 can be though of the study of 2-tensors, i.e., matrices.

Example 5.2 (Tensor evaluation)**.** *Consider a matrix A, a 2-tensor. If you input a vector v, you get another vector Av. In particular, if $A = (v_1, ..., v_n)$ and $v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$, then $Av = \sum a_iv_i$. As you can see, if you apply a column vector from the right, you have collapsed all columns of A into a single column.*

Similarly, you can apply a row vector to the left. For any vector v, then v^TA is essentially collapsing all rows of A into a single row.

Finally, for any pair of vectors v, w, then v^TAw will end up collapsing all rows and columns of A, and you end up with a 0-tensor. In short, a 2-tensor could eat two vectors and spit out a scalar.

In general, if you start with an n-tensor, i.e., an n-dim array of numbers, then you can collapse k of the n dimensions by applying k vectors in the right way. For example, you may think of a 3-tensor as $T = (A_1, ..., A_n)$, a list of matrices. Think of these A_i as slices of this cubic array of numbers. Then for any vector $v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$, you can collapse T according to v as $\sum a_iA_i$. Note that there are three different ways to "slice up" the cubic array T, therefore there are three ways to "collapse". If you input vectors $u = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ and

v, w, then you can collapse all of T as $\sum a_i v^T A_i w$. These collapsing action are called tensor evaluations. We "contract" its dimensions by inputing vectors.

Now note that $v^T A w$ defines a bilinear map from the pair (v, w) to any real number. And bilinear maps and matrices have one-to-one correspondence. (Bilinear corresponds to matrices, Symmetric bilinear corresponds to symmetric matrices, and so on.) Similarly, n-tensors, via the tensor evaluations, are in one-to-one correspondence with n-linear maps.

For example, an inner product on $\mathbb{R}^n$ is a 2-tensor, because the inner product is trying to eat two vectors.

Example 5.3 (Tensor evaluated partially)**.** *Say you have an n-tensor. Then it is waiting to eat n vectors in total. But if you feed it k vectors, then it is still waiting to eat $n-k$ vectors. So feeding k vectors to an n tensor would result in an $n-k$ tensor.*

For example, say we have a 3-tensor, a trilinear map $\langle -, -, - \rangle : V^n \to \mathbb{R}$. Now I feed it a vector $v_1 \in V$. Then we have a bilinear map $\langle v_1, -, - \rangle : V^n \to \mathbb{R}$, which is a 2 tensor, waiting to eat two more vectors in order to spit out a real number.

For example, a matrix A could eat a vector v and spit out another vector Av. So it eats a vector and leaves a 1-tensor. This is a partial tensor evaluation, and we see that the rank of A minus 1 is 1, i.e., it must be a 2-tensor.

Example 5.4 (Stress tensor)**.** *The following is an application of tensor in engineering.*

What is stress? Well say we have an iron cylindrical bar. Let S dentoe a cross section surface in the middle of it with area A. If I pull both end of the sylinder with force F. (So that I am trying to stretch it.) Now since the cylinder is NOT broken in half, the surface S, if treated as a surface to the left part, is pulled to the right by force F. But if S is treated as a surface to the right part, it is pulled to the left by force F. So S has a total stress of F, and each point inside S is subject to a stress of $\frac{F}{A}$.

Obviously,the direction of the stress force is normal to our surface S. However, S subject to a force in both directions. So to simplify the matter, we say that if S is oriented with normal vector to the right, then the stress force is to the right, and S is oriented with normal vector to the left, then the stress force is to the left. (As a tradition, we orient a closed surface with normal vectors in the outward direction. In our case, if S is oriented to the right, it is treated as part of the surface of the left parts of the cylinder.) This is called a uniaxial normal stress.

But now consider the next example. We pull both ends of the cylinder DOWN by force F and pull the middle up by force $2F$. (So that I am trying to bend it.) Let S be cross sections at the quarter and three quarter location of the cylinder. What are the stress felt by S?

Consider S oriented with normal vector to the right. Then we treat it the surface of the left part of the cylinder, i.e., the shorter part. Since it does not move, it turned out that S must feld a total force of F UPWARD! So again the total stress is F, and each point felt a stress of $\frac{F}{A}$, but the direction is NOT normal to the surface.

So what is stress at a point P? It is something that eats a vector and left with another vector, i.e., a matrix or a 2-tensor. It takes a vector n to denote a surface S through P with normal vector n, and it spits out a vector F to denote the force felt by the stress. They are NOT always in the same direction.

At each point there might be a different stress tensor. So in fact we have a tensor field in this case.

Example 5.5. *The cross product eats two vectors and spit out a third. So it is a 3-tensor.*

Example 5.6 (Trifocal tensor)**.** *Say we have three cameras A, B, C taking pictures of the same object. Say A, B, C produce pictures U, V, W respectively. Now suppose a black dot in the actual 3D space is captured on the picture U, say at some location $u \in U$. (Think of the pictures U, V, W as 2D spaces) Do you know the actual location of this black dot in the actual 3D space? No. You only know the direction of the black dot from your camera A, but you do not know HOW FAR away this black is from your camera A. Everything in that direction would appear at the same location on your camera. So the actual location of this black dot is unknow, and you have a line of possible locations for this black dot in the 3D space.*

Now suppose you also see that the very same black dot appear at some location $v \in V$ on the picture V taken by camera B. This gives you another line of possible locations of this black dot in 3D space. Intersect *this line with the previous one, then you will obtain the ACTUAL location of this black dot! Yay! Then once you know the actual location of the black dot, you can easily DEDUCE where this black dot SHOULD appear on the picture W taken by camera C.*

Think about this. For two vectors $u \in U$ and $v \in V$, we can somehow use them to deduce a location $w \in W$. This is a process that eats two vectors and spit out a third. This is a 3-tensor called the trifocal tensor. Given the set up of the three cameras, you can calculate this tensor first, and then any two pictures form two cameras would tell you what the third picture looks like, even if you haven't taken the third picture yet.

Example 5.7 (Riemann Curvature Tensor)**.** *What is a 4-tensor? A 3-tensor is a list of matrices, something like $(A_1, ..., A_n)$. So a 4-tensor is like a matrix of matrices, something like (A_{ij}). But keep in mind that such representations are basis-dependent, so it is only useful in computations, not in understanding.*

One of the most important 4-tensor is the Riemann curvature tensor. And since general relativity uses curvature to describe gravity and such, this 4-tensor is a very fundamental tool. The question is the following: How to describe curvature?

Imagine that you are on a tannis court, holding your racket directly to the north. Now you walk around this tannis court, and make sure that your racket stays in the same direction, i.e., at each step, your racket should be parallel to its direction in the previous step. After you walked around and returned to your original position, your racket should still point to the north, yes?

Now imagine that you are on the unit sphere in $\mathbb{R}^3$. Say we start at the point $p_x = (1,0,0)$ with the racket in the tangent direction $v_x = (0,0,1)$. You walk straight to the point $p_z = (0,0,1)$, and your tangent vector is now $v_z = (-1,0,0)$. Now you walk straight to the point $p_y = (0,1,0)$, and your tangent vector is now still $v_y = (-1,0,0)$. Finally, you walk back to p_x. But now your tangent vector is $(0,1,0)$, different from before. This indicates that the sphere is curved.

The idea can be said as this: on the flat surface, parallel transport around a loop should bring a tangent vector back to itself. The curvature should measure the failure of this. Of course, if you want the curvature at a single POINT, then you cannot take a big look. Instead, you take limit and only do an infinitesimal loop. The infinitesimal picture is this: on a tangent plane, take any tangent vector v (note that in higher dimensions, this vector is not necessarily on the plane), and consider the parallel transport of v along some infinitesimal loop on this tangent plane. You will get resulting vector v_t. Then $\frac{v-v_t}{t}$ is the curvature. Note that this is a vector.

On the time space, you have 3 spatial dimension and time, so 4 dimension in total. For a 4-dimensional object, the tangent spaces are also 4 dimensional. In particular, at each point, you may have MANY tangent planes. So you need two tangent vectors w_1, w_2 to fix a tangent plane, and you want to study the parallel transport of a tangent vector v, and the resulting difference of v before and after the parallel transport is your output, some vector u. So you eat three vectors w_1, w_2, v and spit out a forth vector u. This is a 4-tensor.

Example 5.8 (Determinant)**.** *Determinant is an n-tensor on $V = \mathbb{R}^n$, because $det(v_1, ..., v_n)$ depends linearly on each inputting column vector. It eats n vectors and spit out a number. And as a result, taking signed volume is an n-tensor as well, because it is what determinant means geometrically.*

Example 5.9 (Tensors from tensors)**.** *Tensors can add. Take any σ_1, σ_2, if they have the same rank, then they add to $\sigma_1 + \sigma_2$ with $(\sigma_1 + \sigma_2)(v_1, ..., v_k) = \sigma_1(v_1, ..., v_k) + \sigma_2(v_1, ..., v_k)$. Similarly, you can scale a tensor by $(k\sigma)(v_1, ..., v_k) = k\sigma(v_1, ..., v_k)$. These are easy. In particular, k tensors form a vector space!*

However, you can also do products of tensors, called tensor products. Say you have an a-tensor σ and a b-tensor τ. Then you can produce an $(a+b)$-tensor $\sigma \otimes \tau$ such that $\sigma \otimes \tau(v_1, ..., v_a, w_1, ..., w_b) = \sigma(v_1, ..., v_a)\tau(w_1, ..., w_b)$.

For example, given two matrices A, B, say $A = (a_{ij})$. Then $A \otimes B$ should be a 4-tensor. Recall that a 4-tensor is a matrix of matrices. It is in fact $(a_{ij}B)$ under the right choice of basis. Try verifying this yourself. (And if you cannot, you know you have not yet mastered tensors.) As you can see, 4-tensors obviously combined the info of BOTH A and B.

Example 5.10 (Covariance and contravariance)**.** *We have been very careless here. Technically, there are two kinds of vectors: column vectors and row vectors, or vectors and covectors. A tensor must specify which* *kinds of vector you eat as well. For example, a covector is waiting to eat a vector, while a vector is waiting to eat a covector. If you have a covector and feed it another covector, there is nothing you can really do. You MAY try to take dot product, but dot products are DEPENDENT on basis! And most of the time, people uses tensors, rather than $3D$-matrices or the like, because we want to calculate INdependent of basis. Using dot product would defeat the purpose.*

So in fact we usually say things like (a, b)-tensors, which is waiting to eat a covectors and b vectors. Any inner product structure is a $(0, 2)$-tensor. A vector is a $(1, 0)$-tensor. A linear map, which eats a row vector and a column vector, would be a $(1, 1)$-tensor.

This is very important. Say we have a change of basis matrix B that sends coordinates in the new basis to the coordinates of the old basis (in particular, it is $B = (v_1, ..., v_n)$ where $v_1, .., v_n$ are the expression of the new basis in coordinates of the old basis). Then for a $(0, 2)$-tensor, say some inner product represented by a matrix A, then a change of basis looks like $A_{new} = B^T A_{old} B$. In comparison, for a linear map represented by the matrix A, a change of basis looks like $A_{new} = B^{-1} A_{old} B$.

As you can see, for an (a, b) tensor, the change of basis formula would multiply B to b arguments and B^{-1} to a arguments. For this reason, $(0, k)$-tensors are called covariant tensor, while $(k, 0)$-tensors are called contravariant tensors, and we may talk about whether an argument is covariant or contravariant and so on.

5.2 Definitions and basic properties

The key to a good definition is the tensor product. Say we have vector spaces V, W. Then their vectors are actually $(1, 0)$-tensors of themselves, waiting to eat covectors. For example, $v \in V$ is waiting to eat some $\alpha \in V^*$ and produce a number $\alpha(v)$. Then for any pair $v \in V$ and $w \in W$, we could form a $(2, 0)$-tensor $v \otimes w : V^* \times W^* \to \mathbb{R}$, where (α, β) is sent to $\alpha(v)\beta(w)$.

Definition 5.11. For two vector spaces V, W, we define their tensor product $V \otimes W$ to be the vector space space of bilinear maps from their dual spaces $V^* \times W^*$ to $\mathbb{R}$. (Or to complex numbers if you choose it.) We have a canonical function $\otimes : V \times W \to V \otimes W$, that sends (v, w) to $v \otimes w$ as described above. (Note that $\otimes$ here is NOT linear in general.)

Proposition 5.12 (Dimension of tensor product)**.** *If V is spanned by $v_1, ..., v_n$ and W is spanned by $w_1, ..., w_m$, then $V \otimes W$ is spanned by $v_i \otimes w_j$. In particular, $V \otimes W$ has dimension* $(\dim V)(\dim W)$.

Proof. Take dual basis $v_1^*, ..., v_n^*$ for V^* and $w_1^*, ..., w_m^*$ for W^*. Take any bilinear $B : V^* \times W^* \to \mathbb{R}$. Then $B(\alpha, \beta) = B(\sum a_i v_i^*, \sum b_i w_i^*) = \sum a_i b_j B(v_i^*, w_j^*)$. Now note that $B(v_i^*, w_j^*)$ are constants independent of α or β, while $a_i b_j = v_i \otimes w_j(\alpha, \beta)$. So $B(\alpha, \beta) = \sum B(v_i^*, w_j^*) v_i \otimes w_j(\alpha, \beta)$. In particular, $B = \sum B(v_i^*, w_j^*) v_i \otimes w_j$. So $V \otimes W$ is indeed spanned by these $v_i \otimes w_j$.

Now we need to show that they are linearly independent. Suppose $B = \sum a_{ij} v_i \otimes w_j = 0$. Then $0 = B(v_k^*, w_\ell^*) = \sum a_{ij} v_k^*(v_i) w_\ell^*(w_j) = a_{k\ell}$ for all k, ℓ. So all coefficients are 0. ☐

Remark 5.13. *In particular, for any $B \in V \times W$, if it is $\sum b_{ij} v_i \otimes w_j$, then we may write the matrix (b_{ij}) as a representation of B. As a bilinear map, it sends $v \in V$ and $w \in W$ to $v^T B w$, where the matrix multiplication is done in terms of coordinates. Check this yourself.*

Proposition 5.14 (The universal bilinear space)**.** *$\otimes : V \times W \to V \otimes W$ is bilinear. Furthermore, any bilinear map $B : V \times W \to U$ to any vector space U factors uniquely as a composition of $\otimes : V \times W \to V \otimes W$ and a linear map $L : V \otimes W \to U$. (So $V \otimes W$ is the space that characterize ALL bilinear property of the pair (V, W).)*

Proof. For the first statement, you just need to check all definitions. Say consider $(kv) \otimes w$. Then for any $(\alpha, \beta) \in V^* \times W^*$, then $((kv) \otimes w)(\alpha, \beta) = \alpha(kv)\beta(w) = k\alpha(v)\beta(w) = k(v \otimes w)(\alpha, \beta)$. So $(kv) \otimes w = k(v \otimes w)$. The rest are similar.

Now suppose you have a map $B : V \times W \to U$. Pick any basis for everything so that $V \times W$ is spanned by $v_i \otimes w_j$, and define $L(v_i \otimes w_j) = B(v_i, w_j)$. And we are done. ☐

Example 5.15. *The map $\otimes$ is usually neither surjective nor injective. For example, $(kv) \otimes w = v \otimes (kw)$, so injectivity is always screwed.*

Now consider vectors written in basis $v = a_i v_i$ and $w = b_i w_i$, then $v \otimes w = \sum a_i b_j v_i \otimes w_j$. So the corresponding matrix for this bilinear map is $(a_i b_j)$, a rank 1 matrix.

In general, tensors like $v \otimes w$ correspond to rank 1 matrices. But if you have $v_1 \otimes w_1 + v_2 \otimes w_2$, then you CANNOT write it as a tensor product of two vectors, as you now have a rank 2 matrix.

Example 5.16. *Some place write $V \times W$ as $V \oplus W$, the direct sum. Then you should try to verify that $U \otimes (V \oplus W) = (U \otimes V) \oplus (U \otimes W)$, and $V \otimes W = W \otimes V$, and so on. All these laws for addition and multiplication should hold as you expected.*

In particular, by associativity, I don't need to write $(U \otimes V) \otimes W$ or $U \otimes (V \otimes W)$. They are simply the same thing, so I simply write $U \otimes V \otimes W$. And we have a short hand notation $V^{\otimes n}$ denoting n copies of V tensored together.

You can easily verify that $U \otimes V \otimes W$ is spanned by $u_i \otimes v_j \otimes w_k$ and so on.

Definition 5.17. An (a, b)-tensor is a multilinear map $T : (V^*)^a \times V^b \to \mathbb{R}$. They live in the vector space $V^{\otimes a} \otimes (V^*)^{\otimes b}$. We also denote this as $T^a_b(V)$ to save space.

Remark 5.18. *In general, $V \otimes W$ is like the space of rectangular matrices, whereas $V \otimes V$ are like the square matrices. The latter has a much richer structure.*

Also note that we claim that $V \otimes W = W \otimes V$. This does NOT mean that $v \otimes w = w \otimes v$, which is FALSE! What we mean is that there is a canonical one-to-one correspondence between elements of $V \otimes W$ and elements of $W \otimes V$. And if you think in terms of rectangular matrices, this canonical correspondence is taking the transpose. However, this does NOT imply that $A = A^T$ for all matrices!

In particular, say you are looking at $T^a(V)$ and thinking of $v_1 \otimes ... \otimes v_a$. Then the ORDER of these vectors MATTERS!

Obviously we have $T^a_b(V) \otimes T^c_d(V) = T^{a+c}_{b+d}(V)$. But we also have a converse of this statement.

Proposition 5.19. *The following three things are in canonical one-to-one correspondence.*

1. *A tensor $\tau \in T^a_b(V)$.*
2. *A linear map $\tau : T^c_d(V) \to T^{a-c}_{b-d}(V)$.*
3. *A multilinear map $\tau : V^c \times (V^*)^d \to T^{a-c}_{b-d}(V)$.*

Proof. This is the formal statement of the informal intuition, that if you want to eat n vectors, and you have already eaten k vectors, then you are still waiting to eat $n - k$ vectors. Thus you are now an $(n - k)$-tensor. The proof is trivial. $\square$

Finally, a very beautiful observation.

Proposition 5.20. $T^a_b(V)^* = T^b_a(V)$.

Proof. They obviously evaluate each other. Just build the corresponding dual basis, and you will see that they are the dual space of each other. $\square$

5.3 Alternating Tensors

By itself, there is not such thing as a "tensor-theory". I mean, 2-tensors already gives you a whole year of linear algebra to learn about. The true theory of 3-tensors can only be more overwhelming. It is next to hopeless to build some grand important theory that is both insightful and generic.

However, for a special kinds of tensor, not only do they form a extremely complex and beautiful theory, they also gives the most rigorous formulation to the mathematical tool withe the MOST applications: calculus.

Example 5.21. *To start, integrations on $\mathbb{R}$ are based on signed length. Integration on $\mathbb{R}^2$ are based on signed area. Integration on $\mathbb{R}^3$ are based on signed volume. As you can see, all integrations come from determinants.*

And determinant is funny. It is a tensor $\det(v_1, ..., v_n)$ with the special property that, if you switch two inputs, then you get a negative sign. It is skew-symmetric!

Definition 5.22. A tensor $\tau \in T_k(V)$ is alternating if whenever two inputs are the same, it evaluates to 0. We use $\bigwedge_k(V)$ to denote the space of alternating $(0, k)$-tensors. Dually, we use $\bigwedge^k(V)$ to denote the space of alternating $(k, 0)$-tensors.

Proposition 5.23. *For an alternating tensor, if you switch the order of two inputs, then you will negate the evaluation. (Some places use this as the definition. Well it does not matter either way.)*

Before we proceed, we need a tool. If you just have two tensors, you can multiply them to get a higher rank tensor. However, if you have two alternating tensors, after you multiply them, they might no longer be alternating any more! We need a new product that puts two alternating tensor together, and gets us a new alternating tensor. To do this, there are some technicalities.

Definition 5.24. Say $\tau \in T_k(V)$. For any permutation σ of k objects, then let τ_σ denote τ with its order of inputs permuted by σ. Then $Alt(\tau) = \frac{1}{k!}\sum_{\sigma \in S_k} \operatorname{sign}(\sigma)\tau_\sigma$ is callted the alternization of τ.

Proposition 5.25. *The alternization is always an alternating tensor. The map $Alt : T_k(V) \to \bigwedge_k(V)$ is linear and surjective.*

Proof. If you switch two inputs, say applying some permutation s that swaps a pair, then $Alt(\tau)_s = \frac{1}{k!}\sum_{\sigma \in S_k} \operatorname{sign}(\sigma)(\tau_\sigma)_s = \frac{1}{k!}\sum_{\sigma \in S_k} \operatorname{sign}(\sigma)\tau_{s\sigma} = \frac{1}{k!}\sum_{\sigma \in S_k} \operatorname{sign}(s^{-1}\sigma)\tau_\sigma$, where the last step is a change of variable where we treat $s\sigma$ as the new σ to sum up. But $\operatorname{sign}(s^{-1}\sigma) = -\operatorname{sign}(\sigma)$ by definition, and thus it is negated.

Now Alt is obviously linear, so we only need to show that it is surjective. Well, for any alternating tensor τ, then $\tau = Alt(\tau)$. So we are done. $\square$

Example 5.26. *$Alt(v \otimes w) = \frac{1}{2}(v \otimes w - w \otimes v)$, and $Alt(u \otimes v \otimes w) = \frac{1}{3!}(u \otimes v \otimes w + v \otimes w \otimes u + w \otimes u \otimes v - w \otimes v \otimes u - v \otimes u \otimes w - u \otimes w \otimes v)$. These are quite similar to the determinant formula, yes?*

Definition 5.27. For any two alternating tensor $\sigma \in \bigwedge_a(V), \tau \in \bigwedge_b(V)$, then we can define an exterior product or wedge product $\sigma \wedge \tau \in \bigwedge_{a+b}(V)$, as $\sigma \wedge \tau = \frac{(a+b)!}{a!b!} Alt(\sigma \otimes \tau)$.

Why the constants? Because we want the following:

Proposition 5.28. *For any $\sigma \in T_a(V)$ and $\tau \in T_b(V)$, we have $(a!)Alt(\sigma) \wedge (b!)Alt(\tau) = (a+b)!Alt(\sigma \otimes \tau)$.*

Proof. $(a!)Alt(\sigma) \wedge (b!)Alt(\tau) = (a+b)!Alt(Alt(\sigma) \otimes Alt(\tau)) = \sum_{s \in S_{a+b}} \operatorname{sign}(s)(Alt(\sigma) \otimes Alt(\tau))_s$.

Note that $(a+b)!Alt(\sigma \otimes \tau) = \sum_{s \in S_{a+b}} \operatorname{sign}(s)(\sigma \otimes \tau)_s$. Now for each fixed $s_a \in S_a$ and $s_b \in S_b$, by manipulating indices, it is easy to see that $\sum_{s \in S_{a+b}} \operatorname{sign}(s)(\sigma_{s_a} \otimes \tau_{s_b})_s = \operatorname{sign}(s_a)\operatorname{sign}(s_b)(a+b)!Alt(\sigma \otimes \tau)$. So we are looking at $\frac{(a+b)!}{a!b!}\sum_{s_a \in S_a, s_b \in S_b} \operatorname{sign}(s_a)^2 \operatorname{sign}(s_b)^2 Alt(\sigma \otimes \tau) = (a+b)!Alt(\sigma \otimes \tau)$. $\square$

The constant for wedge product here makes the following beautiful. If you pick other constants for wedge product and alternization, then the following formula will have an ugly constant.

Corollary 5.29. *For any 1-tensors $\tau_1, ..., \tau_k$, then $\tau_1 \wedge ... \wedge \tau_k = \sum_{s \in S_k} \tau_{s(1)} \otimes ... \otimes \tau_{s(k)}$.*

Proof. Recall that $(a!)Alt(\sigma) \wedge (b!)Alt(\tau) = (a+b)!Alt(\sigma \otimes \tau)$. Set $a = b = 1$ and you have the $k = 2$ version of the statement. Similarly, $(a!)Alt(\sigma) \wedge (b!)Alt(\tau) \wedge (c!)Alt(\delta) = (a+b+c)!Alt(\sigma \otimes \tau \otimes \delta)$ and so on. (You can prove this by induction if you need to.) And our statement easily follows. $\square$

Now above concepts might look to be purely algebraic, but in fact they are totally geometric. We don't have the necessary knowledge and time to get into them, but here let me vaguely explain the ideas behind alternization and wedge product. Say you have two perpendicular line segments. Then the wedge product is like a product of these two line segments, and it is therefore a rectangle. The alternizaiton of their tensor product is like the "enclosure" of the two line segment, and it is therefore an isosceles right triangle. As you can see, the rectangle is twice the triangle.

Similarly, with three line segments, their wedge product would gives you the cube, whereas the alternization of their tensor products gives you a right tegrahedron. And the cube has volume (3!) times the tetrahedron.

Intuitively, the wedge product is designed to keep track of high dimensional parallelotopes, while the alternization is going for maximal symmetry, so it results in high dimensional analogues of triangles, tetrahedrons and so on.

Now that we have established wedge product, I have to repeat that you do NOT need to remember any of the above. The set up is not important. From now on, we write everything in terms of wedge product, and will never again use tensor products, so the above messy conversion formulas are not important to us.

Proposition 5.30. *The following are true:*

1. *The map $\wedge : \bigwedge_a(V) \times \bigwedge_b(V) \to \bigwedge_{a+b}(V)$ is bilinear and associative, and $\sigma \wedge \tau = (-1)^{ab}\tau \wedge \sigma$. In particular, $v \wedge v = 0$ for all 1-form v. (Compare this with the formula for $\det(A, B)$ and $\det(B, A)$, where A, B are blocks so that (A, B) and (B, A) are square matrices.)*

2. *For a basis $v_1, ..., v_n$ for V, then alternating tensors $v_{i_1} \wedge ... \wedge v_{i_k}$ with $1 \leq i_1 < ... < i_k \leq n$ would be a basis for $\bigwedge^k(V)$. In particular, $\dim(\bigwedge^k(V)) = C_n^k = \frac{n!}{k!(n-k)!}$.*

Example 5.31. *What is $\bigwedge^k(V)$? Its elements are the true generalizations of vectors into higher dimensions.*

Think about a vector. It has the three following ingredients: It has a magnitude (its signed length), a direction (a line that it lies on), and an orientation (The way in which it goes alone the line). A bivector should be the following: It has a magnitude (a signed area), a direction (a plane it lies on), and an orientation (clockwise or counter-clockwise). For this bivector, we should ONLY care about these three things, and NOT care about its shape or what not. It could represent a triangle, or a square, or a circle, I don't care. As long as it represent the same magnitude on the same plane with the same orientation, then I treat all shapes of such as the same bivector.

Then bivectors are elements of $\bigwedge^2(V)$. Think about it. For any bivector, pick any parallelogram (v, w) it represents, and pick another parallelogram (v', w') it represents, then I claim that $v \wedge w = v' \wedge w'$. Indeed, since v, w and v', w' span the same 2D space, there is a 2 by 2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $(v, w) = (v', w')A$. And since the two parallelogram has the same area, A must fix volume, so $\det(A) = ad - bc = 1$.

Then $v \wedge w = (av' + cw') \wedge (bv' + dw')$. Note that $v' \wedge v' = w' \wedge w' = 0$, and $v' \wedge w' = -w' \wedge v'$. So expand terms and compute, we have $v \wedge w = (ad - bc)v' \wedge w' = v' \wedge w'$.

Conversely, any wedge product $v \wedge w$ clearly has a plane that it lies on, has a signed area, and it is orientated by first goint along v and then along w, which could be a clockwise or counter-clockwise motion around this parallelogram.

You should think of $\bigwedge^2(V)$ as the set of all LINEAR COMBINATIONS of bivectors. Say $v_1 \wedge v_2 + v_3 \wedge v_4$ for four linearly independent vectors cannot really be represented by any single parallelogram. You can only think of it as a linear combination of two parallelograms.

Now you see that $\bigwedge^2(V)$ are linear combinations of "parallelograms". Then $\bigwedge_2(V)$, its dual space, would now corresponds to a way to assign each parallelogram a real number in a linear way. So it is like you defined some new exotic "area" that you want to apply to parallelograms.

You can easily generalize this idea to rank k (covariant or contravariant) alternating tensors.

Example 5.32. *Think of x, y, z-axis as three vectors to span $\mathbb{R}^3$. Then the xy, yz, zx-planes corresponds to $x \wedge y, y \wedge z, z \wedge x$ that span $\bigwedge^2(\mathbb{R}^3)$.*

Example 5.33. *Consider the cross product. Say the basis is e_1, e_2, e_3. Then $(ae_1 + be_2 + ce_3) \wedge (a'e_1 + b'e_2 + c'e_3) = (bc' - cb')e_2 \wedge e_3 + (ca' - ac')e_3 \wedge e_1 + (ab' - ba')e_1 \wedge e_2$, exactly the formula for cross product!*

If you think about it the result of a cross product is almost always the normal vector to something. So in fact, we should NOT treat the result of a cross product as a vector. Rather, we should treat its result as a bivector, and take the wedge product instead.

Another mystery of cross product is the symmetry and anti-symmetry of the triple product, $u \cdot (v \times w)$, which produces the volume of the parallelopipe made by (u, v, w). In particular, $u \cdot (v \times w) = v \cdot (w \times u) = w \cdot (u \times v) = \det(u, v, w) = -u \cdot (w \times v) = -w \cdot (v \times u) = -v \cdot (u \times w)$, which is amazing and magical and quite hard to imagine. Now I claim that $u \wedge v \wedge w = \det(u, v, w)e_1 \wedge e_2 \wedge e_3$. Prove this yourself. So the triple product is nice because it is actually the wedge product of the three vectors.

Example 5.34. *Here is a computational example. Say we are in $\mathbb{R}^2$. Let dx, dy be the covectors that send all vectors to their x, y-coordinates respectively. Then what is $dx \wedge dy$? I claim that this is the determinant. Note that it is a 2-tensor, so it is waiting to eat two vectors. Then $dx \wedge dy(v, w) = (dx \otimes dy - dy \otimes dx)(v, w) = dx(v)dy(w) - dy(v)dx(w)$. This is the $ad - bc$ formula for the determinant, and it is the area of the parallelogram made by (v, w). (In contrast, $dx \otimes dy(v, w) = dx(v)dy(w)$, which is not really anything.)*

In particular, $dx \wedge dy \wedge dz$ is also the determinant in $\mathbb{R}^3$, and it corresponds to measuring the volumn. Recall that determinants on $\mathbb{R}^n$ means we measure the n-volume.

Here you need to be extra careful. To measure the area of a parallelogram made by edges v, w, you input v and w as vectors, or you can input $v \otimes w$ as well, which is equivalent as (0,2)-tensors evaluate (2,0)-tensors naturally. However, you should NOT enter $v \wedge w = v \otimes w - w \otimes v$. Why? Because evaluating at $v \otimes w$ already gives me the area, and evaluating at $-w \otimes v$ gives the negation of the area of the parallelogram at the opposite orientation. Negation of oppositie orientation is like a double negative, and they cancel away. So you end up with twice the folumn instead. So if you insist on entering alternating tensors, you should use $Alt(v \otimes w)$ instead.

Example 5.35. *Note that $\bigwedge_n(\mathbb{R}^n)$ is always 1 dimensional. And we know that taking determinant is in it. So any alternating n-tensor on an n-dimensional space must be a multiple of the determinant, i.e., a multiple of taking n-volume.*

Also note that $\bigwedge^m(\mathbb{R}^n)$ is zero-dimensional if $m > n$. This corresponds to the fact that there are no higher-dimensional stuff in a lower-dimensional space. It also corresponds to the fact that if you take more than n socks from a drawer and there are only n colors, then you must have a pair of socks of the same color (and hence the corresponding alternative tensor is zero.) In contrast, there are certainly higher dimensional tensors in lower dimensional space. Just take $v \otimes v \otimes ... \otimes v$ enough times and you can get a tensor of arbitrarily large rank.

Example 5.36. *Note that $\bigwedge^k(\mathbb{R}^n)$ and $\bigwedge^{n-k}(\mathbb{R}^n)$ has the same dimension! This duality is our old intuition that "k-dim subspaces are in correspondence with $(n-k)$-dim subspaces". One such correspondence would be taking orthogonal complements. This is NOT a canonical correspondence though, because there are many ways to define "orthogonality", if your basis is not fixed.*

Example 5.37. *Take a right tetrahedron with mutually perpendicular edges u, v, w, and mutually perpendicular faces S_{uv}, S_{vw}, S_{wu} and a slant face F. Then I claim that $Area(F)^2 = Area(S_{uv})^2 + Area(S_{vw})^2 + Area(S_{wu}^2)$. This is an analogue of the Pythagorean theorem. (We also proved this in linear algebra one class.)*

How to prove this? Think of $S_{uv}, S_{vw}, S_{wu}, F$ as alternating tensors, say $S_{uv} = \frac{1}{2}u \wedge v$ and so on. If u, v, w are mutally orthogonal, then the inner product structure of V induces an inner product structure on $\bigwedge^2(V)$ where $u \wedge v, v \wedge w, w \wedge u$ are mutually orthogonal. Finally, note that $F = \frac{1}{2}(u - v) \wedge (v - w) = \frac{1}{2}u \wedge v + \frac{1}{2}v \wedge w + \frac{1}{2}w \wedge u$, which is a decomposition of F into mutually orthogonal components. So the square of the magnitude of F is exactly the sum of squares of the maginitude of these mutually orthogonal components. So we are done.

5.4 Differential Forms and Boundaries

For a geometric object S, at each point p, we can find tangent space T_p and cotangent space T_p^*. A vector field is an assignment of a tangent vector in T_p at each point p. Similarly, a (a,b)-tensor field is an assignment of an (a,b)-tensor in $T_b^a(T_p)$ at each point p. In this spirit, a k-form is just an alternating $(0,k)$-tensor field, i.e., we are assigning an alternating tensor in $\bigwedge_k(T_p)$ to each point p.

Think about it like this: The tangent space V is the infinitesimal neighborhood around p. By taking this form (alternating tensor field) at each point p, we are now ready to measure some k-dim volume at the infinitesimal neighborhood around p. So given any k-dim sub-object $T \subseteq S$, we add up the infinitesimal k-volume of T at each point, and we will get the total volume of T. (Note that here we are defining volume not in the traditional way, but rather using the differential form.)

So given a k-form ω on S and a k-dim sub-object $T \subseteq S$, we have an evaluation $\int_T \omega$. This is the essense of integration.

We would not go that abstract yet. For us, it is enough to just focus on $S = \mathbb{R}^n$.

Definition 5.38. On $\mathbb{R}^n$, a k-form ω is an assignment of an alternating tensor $\omega_p \in \bigwedge_k(T_p)$ at each point $p \in \mathbb{R}^n$.

Example 5.39. *What are 0-forms? Note the zero tensors are just scalars, so the zero alternating tensors are also scalars. So a differential 0-form assigns a value in $\mathbb{R}$ to each point p. Huh, this is just the definition of a function $f : \mathbb{R}^n \to \mathbb{R}$. So differential* 0*-form are just your familiar functions. You integrate f along a point, i.e., a 0-dim object. This integration here is also called evaluation of f at that point.*

What about 1-forms? Note that $\bigwedge_1(T_p) = T_1(T_p) = T_p^*$, so this is just a covector field. Say in $\mathbb{R}^3$, then at each point p, the covectors $dx|_p, dy|_p, dz|_p$ span T_p^*. So our differential 1-form is just $fdx + gdy + hdz$ where f, g, h take real values and vary from points to points, i.e., they are functions $f, g, h : \mathbb{R}^3 \to \mathbb{R}$! So things like $xdx + ydz$ and such are all 1-forms. You integrate 1-forms along curves, i.e., 1-dim geometric objects.

What about 2-forms? Well since $\bigwedge_1(T_p)$ is spanned by dx, dy, dz, we see that $\bigwedge_2(T_p)$ is spanned by $dx \wedge dy, dy \wedge dz, dz \wedge dx$. So a generic 2-form on $\mathbb{R}^3$ is $fdxdy + gdydz + hdzdx$, where we obmitted the wedge because we are lazy, and $f, g, h : \mathbb{R}^3 \to \mathbb{R}$ are functions. So things like $\sin(x+z)dxdy + z^x dydz + zdxdy$ *are all 2-forms. You integrate 2-forms along surfaces, i.e., 2-dim geometric objects.*

What about 3-forms? Note that T_p is 3-dimensional, so $\bigwedge_3(T_p)$ is 1 dimensional, and spanned by $dxdydz$. So 3-forms are $fdxdydz$ for some function $f : \mathbb{R}^3 \to \mathbb{R}$. Things like $xyzdxdydz$ are 3-forms. You integrate them on some 3D domain.

Example 5.40 (Integrations)**.** *How are integrations done? First of all, what is a valid domain of integration? Well, say you are in $\mathbb{R}^3$ and you are trying to integrate $dxdydz$. Then $\int_U dxdydz$ is defined to be the signed volume of U (the sign depends on the orientation of U). Then for a generic $\int_U f(x,y,z)dxdydz$, you first break down U into tiny pieces. Then at each piece P, since it is so tiny and f is continuous, $f(x,y,z)$ is pretty much constant inside P. Take any $p \in P$, then $\int_P f(x,y,z)dxdydz$ is approximately $f(p)vol(P)$. Then you add these up, and this would approximately be $\int_U f(x,y,z)dxdydz$. Taking limit to send each P to be infinitesimally small, then the limit gives you the true value of $\int_U f(x,y,z)dxdydz$.*

Now say in $\mathbb{R}^3$ you are trying to integrate some 2-form $fdxdy + gdydz + hdzdx$ along some smooth surface S. First you break down S into super tiny pieces such that each piece is approximately flat. Now take an almost-flat piece P, it approximately has some tiny area and lies on some tangent plane of the surface, and it has an orientation coming from the orientation of S. So it is approximately some bivector at some point. Then you can evaluate this bivector, i.e., something in $\bigwedge^2(T_p)$, using our two form, i.e., something in $\bigwedge_2(T_p) = (\bigwedge^2(T_p))^*$. Add these up. Take limits to send each P to be infinitesimally small. You know the drill.

Definition 5.41. We call a k-form on $\mathbb{R}^n$ smooth if all corresponding functions are all smooth, i.e., infinitely differentialbe. (In particular, it means that the alternating tensor that get assigned to each point vary smoothly from point to point.) We usually use $\Omega_k(\mathbb{R}^n)$ to denote the vector space of smooth differential k-forms. You sometimes also see people talking about differential forms, that is when all involved functions are differentiable.

Example 5.42 (Integral-forms duality). *Note that Ω_k is necessarily infinite dimensional. For example, $\Omega_0(\mathbb{R}^n)$ is the space of ALL smooth functions from $\mathbb{R}^n \to \mathbb{R}$.*

Even though it is infinite dimensional, it still have a dual space. In particular, think about this: what is integration? An integration along some k-dim smooth object S is a LINEAR map $\int_S : \Omega_k(\mathbb{R}^n) \to \mathbb{R}$, so integrations are DUAL vectors to smooth forms.

In practice, most dual vectors to smooth forms will be linear combinations of integrations along various objects. There are weird exceptions, but don't think about them too much because no one ever use them. Also integrations already provides you with enough dual vectors, in the following way: If $\int_S \omega = 0$ for all smooth S, then $\omega = 0$.

For a proof, take S to be super super small so that it is almost in the infinitesimal neighborhood of some point p. Then S is practially inside T_p. It is k-dimensional with magnitude, direction and orientation, so it is some element of $\bigwedge^k(T_p)$. But ω integrates to 0 on it, so ω evaluate this element of $\bigwedge^k(T_p)$ to zero. But since S is arbitrary, ω evaluates ALL elements of $\bigwedge^k(T_p)$ at ALL points p to zero. So $\omega = 0$.

An interesting converse is also true: If $\int_S \omega = 0$ for all k-form $\omega \in \Omega_k$, then S is equivalent to being empty, at least from a k-dim point of view.

Now intuitively, k-forms are defining some exotic oriented k-volumes. A very solid intuition of ours is the following: length times length is area, length times area is volume, and so forth. This intuition is completely captrued by the wedge product, or exterior product.

Proposition 5.43. *Given an a-form ω and a b-form η, we can take their wedge product at each point and get an $(a+b)$-form $\omega \wedge \eta$. Then we have the following property:*

1. *The exterior product is bilinear and associative;*
2. $\omega \wedge \eta = (-1)^{ab}\eta \wedge \omega$.

Now the next natural concept is to take derivatives, i.e., say $df = f_x dx + f_y dy + f_z dz$. As you can see and guess, this derivative process send k-forms to $(k+1)$-forms. There are various ways to define this. Here are some examples.

Example 5.44. *Take a 0-form $f : \mathbb{R}^3 \to \mathbb{R}$. Then what should be the meaning of df? First of all, it is a 1-form and should evaluate tangent vectors. Note that df feels like it is evaluating the change in f. So for each tangent vector v, $df(v)$ should mean how much f should change in the direction of this tangent vector v, i.e., the directional derivative! In particular, $df\begin{pmatrix}a\\b\\c\end{pmatrix}$ should be $af_x + bf_y + cf_z$ or $(f_x, f_y, f_z)\begin{pmatrix}a\\b\\c\end{pmatrix}$. So df should be a row vector (f_x, f_y, f_z).*

Now say $f = x$, i.e., the function of taking x-coordinates of each point. Then you can easily see that $dx = (1,0,0)$. Similarly, $dy = (0,1,0)$ and $dz = (1,0,0)$, so we see that $df = f_x dx + f_y dy + f_z dz$. Here you should interpret dx, dy, dz as the changes in x, y, z-coordinates, and their linear combinations gives you the change in f-value.

Now for a generic k-form ω on $\mathbb{R}^n$, it eats k vectors. On the other hand, $d\omega$ should be a $(k+1)$-form, and thus wants to eat one more vector than before. So here is a (WRONG) way to define $d\omega$. Fix tangent vectors $v_1, ..., v_k$ and an extra tangent vector w at a point p. If you move from p in the direction of w by a tiny bit, then you get to some point $q = p + tw$ where eventually we will send t to a limit of 0. Then you can define $d\omega(w, v_1, ..., v_k)|_p = \lim_{t\to 0} \frac{\omega_{p+tw}(v_1,...,v_k) - \omega_p(v_1,...,v_k)}{t}$, the directional derivative of ω in the direction of w.

This sounds good in theory, but it is WRONG. If you define $d\omega$ like this, it may NO longer be alternating! For example, take a one-form $\omega = f dx$ on $\mathbb{R}^2$. (A generic 1-form looks like $f dx + g dy$, and here I'm just studying one of the summand.) Then $d\omega(w, v)$ would be (WRONGLY) defined as $d(xdy)(w,v) = \lim_{t\to 0} \frac{(fdx)_{p+tw}(v) - (fdx)_p(v)}{t} = \lim_{t\to 0} \frac{f(p+tw)dx(v) - f(p)dx(v)}{t} = dx(v)\lim_{t\to 0}\frac{f(p+tw)-f(p)}{t} = dx(v)df(w) = (df \otimes dx)(w, v)$. But $df \otimes dx$ is NOT alternating!

Here is a very important philosophy: you MUST define things with meanings. Say you are in $\mathbb{R}^3$ with $fdx \wedge dy$ for some function f. Then $d(fdx \wedge dy)$ should be seeking volumes. To find the volume according to $d(fdx \wedge dy)$, you first measure the base area with $dx \wedge dy$, and then you measure the extra edge sticking out of the base with df, and then you use these to to find volumn, i.e., taking wedge product. So we have $d(fdx \wedge dy) = df \wedge dx \wedge dy$.

Definition 5.45. Recall that we write $dx_{i_1} \wedge dx_{i_2} \wedge ... \wedge dx_{i_k}$ as dx_I where I is the multi-index $(i_1, ..., i_k)$. Then elements in Ω_k are $\sum f_I dx_I$. We define the exterior derivative as $d(\sum f_I dx_I) = \sum df_I \wedge dx_I$.

Example 5.46 (Taking Boundary and Exterior Derivatives)**.** *Now think about a typical dual vector to k-forms, i.e., some integration $\int_S$ for some k-dim object S. There is a very obvious transformation here on S, called taking boundary, that sends each S to its boundary ∂S. This extends to a linear map from $\partial : \Omega_k^* \to \Omega_{k-1}^*$ that sends $\int_S$ to $\int_{\partial S}$. Here note that the boundary of some k-dim object is $(k-1)$-dim.*

Now recall, if there is a linear map between two vector spaces, then there is a corresponding DUAL linear map between the two dual spaces in the other direction, called the adjoint. So there is a linear map $d : \Omega_{k-1} \to \Omega_k$ which is the adjoint of ∂. I claim that this is exactly the exterior derivative we just defined! In particular, we have a diagram below where the evaluation along each column is the same:

$$\Omega_k^* \xleftarrow{\partial} \Omega_{k+1}^*$$

$$\Omega_k \xrightarrow{d} \Omega_{k+1}$$

More precisely, by definition we have $\int_{\partial S} \omega = \int_S d\omega$. This is Stoke's Theorem. (Someplace define exterior derivative FIRST as the dual of taking boundary, THEN derive its formula. It does not really matter which way you do it.)

For example, take a curve C with orientation from one point p to another q in $\mathbb{R}^2$. Then ∂C are the points p, q, except that there is an induced orientation on these points. q will have positive orientaiton and p negative orientation. (Sort of like $q - p$.)

Suppose you have a function $f : \mathbb{R}^2 \to \mathbb{R}$. What is df? It is something such that, for ALL curves C, $\int_C df = \int_{\partial C} f = f(q) - f(p)$, as integrations along points are just evaluations. As you probably know from your calculus class, the differential form $df = f_x dx + f_y dy + f_z dz$ has exactly such a property.

In general, given a form ω, then by $\int_S d\omega = \int_{\partial S} \omega$, all the integrations of $d\omega$ are already known, so we should be able to deduce what $d\omega$ is. Suppose you have a 2-form $fdxdy$ in $\mathbb{R}^3$. Take a cube U. Then $\int_U d(fdxdy) = \int_{\partial U} fdxdy$. Now what does $dxdy$ mean? Recall our previous example of decomposing a slant surface F into orthogonal components $F = a(u \wedge v) + b(v \wedge w) + c(w \wedge u)$ for orthonormal basis u, v, w, then $a(u \wedge v)$ is the "shadow" of F on the uv-plane. So $dxdy$ is evaluating bivectors by first project them to the xy-plane and then take the area of this shadow.

So for $\int_{\partial U} fdxdy$, note that ∂U has six faces, but four of them are perpendicular to the xy-plane, so $fdxdy = 0$ on these faces. The two other faces are parallel to the xy-plane, the top one is oriented with normal vector upward, and the bottom one is oriented with normal vector downward. So $\int_{\partial U} fdxd$ essentially means taking the integration of f on the top face, and then minuse the integration of f on the bottom face.Say the top face is $z = a$ and the bottom face is $z = b$, and the shadow of both on the xy-plane is $S \in \mathbb{R}^2$. Then $\int_{\partial U} fdxdy = \int_S [f(a) - f(b)]dxdy$, where the latter integration is taken inside $\mathbb{R}^2$.

But $f(a) - f(b) = \int_{L_{a,b}} df$ where $L_{a,b}$ is the straight line from the point (x, y, b) on the bottom face to the corresponding point (x, y, a) on the top face. So $\int_S [f(a) - f(b)]dxdy = \int_{S \times [a,b]} dfdxdy = \int_U dfdxdy$. Now this is true for all small cube U, and any 3D integration can be approximated by integrations on small cubes, so $\int_{\partial U} fdxdy = \int_U dfdxdy$ for ALL 3D domain U. In particular, we see that $d(fdxdy) = dfdxdy$ as predicted by our formula. We have proven Stoke's Theorem

Proposition 5.47. *Let ω, η be smooth a-form and smooth b-form. Then we have*

1. *d is linear and adjoint to taking the boundary.*

2. $d(\omega \wedge \eta) = (d\omega) \wedge \eta + (-1)^k \omega \wedge (d\eta)$*.*

3. $d(d\omega) = 0$*. We also write this as* $d^2\omega = 0$*.*

Proof. The first one is already done. The second one is just computation. For the third one, note that the boundary of boundary is always empty, so $d^2 = 0$. □

Corollary 5.48. *Mixed derivatives are the same (for twice differentiable functions defined everywhere).*

Proof. Say we are on $\mathbb{R}^2$. Consider $0 = d(df) = d(f_x dx + f_y dy) = f_{xy} dydx + f_{yx} dxdy = (f_{yx} - f_{xy}) dxdy$. So $f_{yx} - f_{xy} = 0$. In particular, that mixed derivatives are the same is a consequence of the fact that boundary of the boundary is empty! □

Corollary 5.49 (Green Stoke Divergence)**.** *For any 2D domain A in* $\mathbb{R}^2$*, we have* $\int_A (g_x - f_y) dxdy = \int_{\partial A} f dx + g dy$*, or also written as* $\int_A curl(F) dxdy = \int_{\partial A} F \cdot dr$*, where F is vector valued function, i.e., a vector field. Similarly, for any 2D surface S in* $\mathbb{R}^3$*, we have* $\int_A (h_y - g_z) dydz + (f_z - h_x) dzdx + (g_x - f_y) dxdy = \int_{\partial A} f dx + g dy + h dz$*, or also written as* $\int_A curl(F) \cdot n dS = \int_{\partial A} F \cdot dr$*, where F is again a vector field and n is the normal vector to A. Finally, for any 3D domain A in* $\mathbb{R}^3$*, we have* $\int_A (f_x + g_y + h_z) dxdydz = \int_{\partial A} f dydz + g dzdx + h dxdy$*, or also written as* $\int_A div(F) dxdydz = \int_{\partial A} F \cdot n dS$*, where F is again a vector field and n is the normal vector to the surface* ∂A*.*

Proof. The only nontrivial portion is that for any curve C, when doing $\int_C$, then $(dx, dy, dz)^T$ would evaluate into the velocity of C. But this is obvious by definition. $\int_C$ means breaking it down into infinitesimal portion and then each small piece is approximately some tangent vector v, the velocity. And then we feed this tangent vector to the integrand, so of course $(dx, dy, dz)^T$ would evaluate into v as it literally lists out the three coordinates of v.

Similarly, let S be any 2D surface, I want to show that $(dydz, dzdx, dxdy)^T$ would evaluate into the normal vector to S. For any two input tangent vectors v, w of S, then recall things like $dydz$ are WEDGE products, not tensor products. In particular, $dy \wedge dz = dy \otimes dz - dz \otimes dy$. So $(dydz, dzdx, dxdy)^T$ is evaluated into $(v_y w_z - w_y v_z, v_z w_x - w_z v_x, v_x w_y - w_x v_y)^T = v \times w$ the cross porduct. So of course it is normal to S and its lenght is the area of the parallelogram (v, w). □

Finally, let us try to visualize the exterior derivative process a bit.

Example 5.50. *Recall the onion structure for covector fields. In general, for a k-form, think of it as endowing your space* $\mathbb{R}^n$ *with an onion structure, where the layers are* $(n-k)$*-dim surfaces (whose tangent vectors are exactly the kernel of your k-form), and the densities of these surfaces at each* $p \in \mathbb{R}^n$ *gives the magnitude of your k-form at each point. Integration over some k-surface simply asks how many* $(n-k)$*-surfaces have you punctured through. This gives you a way to draw some k-forms explicitly.*

For example, $xdx + ydy$ *on* $\mathbb{R}^2$ *can be thought of as a folliation of curves, and you can see that these are all concentric circles around the origin. Since its magnitude gets larger the further out, therefore we are dense away from zero, and sparse close to zero. Say we have a curve C from somewhere on the circle of radius r to another point on a circle with radius q, then* $\int_C xdx + ydy = \int_C d(\frac{x^2+y^2}{2}) = \int_{\partial C} \frac{x^2+y^2}{2} = \frac{q^2}{2} - \frac{r^2}{2}$*, which corresponds to how many circles you have cut through (counting densities).*

A 2-form $f dxdy$ *on* $\mathbb{R}^2$ *would now be a folliation of points. So all we need is the density, and it is described by f. This is simple enough.*

Now consider xdy*. All kernels are in the x-direction, so all folliating curves here are horizontal. However, the densities is larger to the right (when x is large), and smaller to the left (when x is small). HOW is that possible? It means that we have NEW lines that popped up as you move to the right. Since the coefficient x of* xdy *increases in unit speed as we move to the right in unit speed, the number of lines increases at a unit rate as we move to the right.*

When we do $d(xdy)$*, we simply takes the boundary of all these folliating curves to get the folliating points. Only horizontal lines that started half way could have boundary (i.e., endpoints), and since new lines appears with steady rate, their foundaries are points with steady unit density, which is exactly* $dxdy$*, the standard area evaluation.*

In contrast, you see that $d(df) = 0$. Why? Because the folliating curves for df are level curves, and they are all closed loops without boundaries. So $d(df)$ have nothing.

Note that when we say "how many layers are punctured", we mean to count these punctures with signs. Say we have a closed curve C. I go in and then go out. Then these two punctures should cancel away! So if I integrate $\int_A d(df) = \int_{\partial A} df$ on any 2D domain A, I am cutting level curves of f by ∂A. Since ∂A is a loop, it ends up where it started, and since level curves are loops as well, anything you cut, you must uncut them later. So you integrate to 0. So $d(df) = 0$.

Example 5.51. *For a 3D example, let us consider the Stoke's theorem. On one hand, we have a 1-form ω that we want to integrate on the boundary ∂S to some surface S. 1-forms on $\mathbb{R}^3$ should be a folliation by surfaces (possibly not closed up), and the integration is asking how many surface is punctured by ∂S.*

On the other hand, we have a 2-form $d\omega$ that we want to integrate on S. Now the 2-forms on a 3D space should be a folliation by curves with densities, which corresponds to the boundaries of the previous folliation by surfaces. Integrating $d\omega$ over S means how many of these curves are cut by S. But for any two surface S, T, S cut ∂T iff T cut ∂S (counting signs). So we are done, and we have established Stoke's theorem.

5.5 de Rham cohomology

Example 5.52. *Consider the following chains.*

$$0 \xrightarrow{d} \Omega_0(\mathbb{R}^2) \xrightarrow{d} \Omega_1(\mathbb{R}^2) \xrightarrow{d} \Omega_2(\mathbb{R}^2) \xrightarrow{d} 0$$

$$0 \xrightarrow{d} \Omega_0(\mathbb{R}^3) \xrightarrow{d} \Omega_1(\mathbb{R}^3) \xrightarrow{d} \Omega_2(\mathbb{R}^3) \xrightarrow{d} \Omega_3(\mathbb{R}^3) \xrightarrow{d} 0$$

Here note that $\Omega_k(\mathbb{R}^n) = 0$ the zero vector space if $k > n$, so in fact you can continue to the right forever by keep adding zero spaces. And we also add in the first zero in the chains just because it is pretty. Note that $d^2 = 0$, so in these chains, if you go two step then you must die. Chains of vector spaces like these, where you go two step you die, are called chain complexes.

Also note that the Ω_0 and Ω_n can be identified as functions, while the Ω_1 and Ω_{n-1} are vector fields. Then d_0 here is always taking the gradient, and d_1 here is always taking the curl. For $\mathbb{R}^3$, d_{n-1} here is taking the divergence. In particular, $d(F \cdot \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix}) = curl(F) \cdot \begin{pmatrix} dydz \\ dzdx \\ dxdy \end{pmatrix}$, and $d(F \cdot \begin{pmatrix} dydz \\ dzdx \\ dxdy \end{pmatrix}) = div(F) dxdydz$.

Our chains here has the extra nice property that it is EXACT. If you recall, an exact sequence is a chain complex such that the image of each linear map is exactly the kernel of the next one. To be more explicit, from now on we use d_n to denote the exterior derivative from Ω_n to Ω_{n+1}. Then over $\mathbb{R}^n$, we always have $\mathrm{Ker}(d_{n+1}) = \mathrm{Ran}(d_n)$.

Definition 5.53. A form ω is closed if $d\omega = 0$. A form ω is exact if it is in $\mathrm{Ran}(d)$.

We are only going to prove this informally. We want to show that closed form are the same thing as exact form. Since $d^2 = 0$, obviously exact forms are closed. We now need to show that closed form are exact.

Think about the dual process, where now we are thinking about geometric shapes and the "taking boundary" map. Then we want to prove that if something has no boundary (closed up), then it is the boundary of something.

Imagine we have a closed loop L in $\mathbb{R}^3$, which has no boundary. Is L itself the boundary of something? Well, imagine that I am a cowboy holding a rope loop, and L is the loop. We tighten the rope little by little, until the whole loop shinks to one point. So in this process, L deforms continuously into a single point. Now L sweeps through a surface S via this process, and obviously $L = \partial S$.

To sum up, since everything in $\mathbb{R}^n$ is "contradtible", i.e., has a continuous deformation to a single point, all closed things are boundaries. So closed forms are exact.

Example 5.54. *For a vector field F, $F = \nabla f$ for some f iff $curl(F) = 0$.*

However, if we replace $\mathbb{R}^n$ by some other geometric objects, then closed form may no longer exact. Consider $\mathbb{R}^2 - \{0\}$, the punctured plane. This is pretty much the same as $\mathbb{R}^2$, but we can define a 2-form on it by $\frac{-ydx+xdy}{x^2+y^2}$. This 2-form is WELL-defined on $\mathbb{R}^2 - \{0\}$, but ILL-defined on $\mathbb{R}^2$, because the denominator CANNOT be zero. Interesting, yes?

Furthermore, $d(\frac{-ydx+xdy}{x^2+y^2}) = 0$. You can verify this by direct calculation, or you can see that the folliation by curves for $\frac{-ydx+xdy}{x^2+y^2}$ are exactly the rays from origin to infinity, and there are NO rays that start midway. So all end points are either at the origin, or at infinity, so there are NO end points on $\mathbb{R}^2 - \{0\}$, so its exterior derivative is zero.

Finally, $\frac{-ydx+xdy}{x^2+y^2}$ is NOT exact. If it is exact, say it is df, then as $\int_C df = \int_{\partial C} f$, we see that the 1-form should integrate to zero on ALL closed curves. However, integrate it on the unit circle and you get 2π. So it is NOT exact.

What happened? Again imagin the cow boy and rope loop example. We want to find S so that ∂S is our loop, the unit circle. We tighten the rope loops little by little, try to contract it into a single point, and OOPS, since we now live in $\mathbb{R}^2 - \{0\}$, the origin is now a hole, and our loop get stuck at the hole. There is no way around it. (In particular, any S with unit circle as boundary must contain the origin, and hence it does not exist in $\mathbb{R}^2 - \{0\}$.)

So the sequence below is NOT exact. The exactness failed at Ω_1, where $\operatorname{Ker}(d_1) \neq \operatorname{Ran}(d_0)$.

$$0 \xrightarrow{d} \Omega_0(\mathbb{R}^2 - \{0\}) \xrightarrow{d_0} \Omega_1(\mathbb{R}^2 - \{0\}) \xrightarrow{d_1} \Omega_2(\mathbb{R}^2 - \{0\}) \xrightarrow{d} 0$$

Now how much have we failed? Here is an interesting result:

Lemma 5.55. *For any closed 1-form ω on $\mathbb{R}^2 - \{0\}$, let C be the unit circle (oriented counter-clockwise), and let L be any loop that winds around the origin k-times (positive k for counterclockwise and negative k for clockwise). Then $\int_L \omega = k \int_C \omega$.*

Proof. For any rectangle R that does NOT contain the origin, then $\int_{\partial R} \omega = \int_R d\omega = 0$. Now using tiny rectangles, I can approximate almost all loops. So for all loops NOT containing the origin, ω integrate to zero.

Now take any loop L, and suppose it winds around the origin k-times (positive k for counterclockwise and negative k for clockwise). Consider $L - kC$, which would then be a loop NOT containing the origin, so $\int_{L-kC} \omega = 0$. In particular, $\int_L \omega = k \int_C \omega$. ☐

Corollary 5.56. *For any loop L on $\mathbb{R}^2$ that does NOT go through the origin, the value $\frac{1}{2\pi} \int_L \frac{-ydx+xdy}{x^2+y^2}$ is always an integer, and it is the number of times L winds around the origin (positive being counter-clockwise), called the winding number.*

Proposition 5.57. $\dim(\operatorname{Ker}(d_1)/\operatorname{Ran}(d_0)) = 1$ *in* $\Omega_1(\mathbb{R}^2 - \{0\})$.

Proof. I claim that if you have a closed form that integrate to zero on the unit circle C, then it is exact. With this claim, you see that the integration map $\int_C : \operatorname{Ker}(d_1) \to \mathbb{R}$ has kernel exactly $\operatorname{Ran}(d_0)$, and hence we have our result.

Now say ω is a closed 1-form with $\int_C \omega = 0$. Then for any closed loop L, $\int_L \omega = k \int_C \omega = 0$. So ω integrate to zero on ALL loops.

Pick any point $p \in \mathbb{R}^2 - \{0\}$ and FIX it. For any other point $q \in \mathbb{R}^2 - \{0\}$, pick any path P from p to q, and define $f(q) = \int_P \omega$. I claim that this is in fact independent of the path P you choose, so that the function here $f(q)$ is unambiguously defined. Then obviously $\omega = df$ by Stock's theorem.

For my claim, let P, Q be two paths. Then $P - Q$ might be interpreted as a loop. Then $\int_{P-Q} \omega = 0$, so $\int_P \omega = \int_Q \omega$. ☐

Definition 5.58. The de Rham cohomology at dimension k is the quotient $H_n = \operatorname{Ker}(d_k)/\operatorname{Ran}(d_{k-1}))$.

Intuitively, de Rham cohomology "count" the number of k-dim holes. (We are not proving this. You should just trust me on this. If you are curious, take topology class in the future.) Let us say that a k-dim hole is a hole whose boundary is k-dim. A punctured plane has a 1-dim hole. A sphere has NO 1-dim hole, but it has a 2-dim hole inside of it. What about a punctured sphere? Does it have a 1-dim hole and a 2-dim hole? Think about it a bit. The answer is that it would have no hole whatsoever. If a geometric object has a k-dim holes, then its de Rham cohomology at dimension k would be a-dimensional. Good geometers could usually tell de Rham cohomology by sight.

Apart from the amazing geometric interpretation of the dimension of de Rham cohomology, its contents and structures also provide us with a tangible way to compute stuff. The space Ω_k are all infinite dimensional, but H_k are almost always finite dimensional, and hence much easier to compute.

Another amazing thing about de Rham cohomology is that sometimes we have geometric objects that we want to cut and puncture and glue and manipulate. When you do these things, it is VERY unclear what would happen to Ω_k. However, the number of holes behaves much better in this aspect, so H_k are much easier to see.

Example 5.59. *I claim that for $m \neq n$, there are NO continuous bijection between $\mathbb{R}^n$ and $\mathbb{R}^m$. Proof: say f is such a map. Now we remove a point from $\mathbb{R}^n$, then $f : \mathbb{R}^n - \{p\} \to \mathbb{R}^m - \{f(p)\}$ would also be a continuous bijection, and hence it must perserves the hole-structure, i.e., they should have identical de Rham cohomology. However, we know that the de Rham cohomology of $\mathbb{R}^n - \{p\}$ and $\mathbb{R}^m - \{f(p)\}$ are different. So we are done.*

(Note that the $1 + \dim H_0$ counts the number of connected components. You see that after removing a point, $\mathbb{R}$ snaps into two, while $\mathbb{R}^2$ is still connected. This is reflected by their difference in $\dim H_0$.)

In comparison, it is POSSIBLE to have a continuous surjection from $\mathbb{R}$ to $\mathbb{R}^2$! If you are curious, search online for space-filling curves.

5.6 Hodge Dual and Maxwell's equation

Again recall that $\dim \Omega_k(\mathbb{R}^n) = \dim \Omega_{n-k}(\mathbb{R}^n)$. Is there a way to formalize this into some duality? The answer is yes. Say you are looking at $v \wedge w$, which is a bivector. Now you really really want to turn this into a vector, say $v \times w$. The the hodge star operator is something with $*(v \wedge w) = v \times w$. In terms of coordinates, in $\mathbb{R}^3$, we want $*(dx \wedge dy) = dz$ and so on. It sends "stuff" to the "rest of the stuff". Equivalently, it sends a k-dim stuff to something like its orthogonal complement, but with identical magnitude. Now note that we immediately have $*(dy \wedge dx) = -dz$, so you need to be careful about the orders of things and the minus signs.

Note that this depends on the inner product structure.

Example 5.60 (Induced inner product structure)**.** *Given an inner product structure on V, it is equivalent to a way to identify V with V^*. But then this is equivalent to an inner product structure on V^*!*

Now given two inner product spaces V, W, then we have ways to identify V with V^* and W with W^*. Then this immediately gives a way to identify $V \otimes W$ with $V^* \otimes W^*$, which is an inner product structure on $V \otimes W$!

Therefore, given any inner product structure on V, $T^a_b(V)$, as a tensor product of a bunch of V and V^*, has a canonical induced inner product structure from V.

In particular, consider $\bigwedge_n(V)$, which is 1 dimensional and has an inner product structure. Take the only unit vector ω in it, then this corresponds to the traditionally "taking n-volume" operation. It maps things to their n-volumn. We call this the standard volumn form. (Note that without inner product, there is no length, and thus no volumn. Volumns on V must depends on inner product structure on V.)

Example 5.61. *Now you can define hodge star like this: let ω be the sdandard volume form. For any $\beta \in \Omega_k$, then $*\beta$ is the $(n-k)$-form such that $\alpha \wedge (*\beta) = \langle \alpha, \beta \rangle \omega$. In particular, for differential forms of unit magnitude, $\beta \wedge (*\beta) = \omega$, i.e., $*\beta$ is the thing that "complete" β into the standard volume form.*

Computational wise, given basis $x_1, ..., x_n$, then we have $*(dx_1 \wedge ... \wedge dx_k) = dx_{k+1} \wedge ... \wedge dx_n$. And for any permutation σ, we have $*(dx_{\sigma(1)} \wedge ... \wedge dx_{\sigma(k)}) = sign(\sigma) dx_{\sigma(k+1)} \wedge ... \wedge dx_{\sigma(n)}$. It does NOT interfere with coefficients, i.e., $*(f dx dy) = f dz$.

Also note that $**\beta = (-1)^{k(n-k)}\beta$ *for any k-form* β*. So this is NOT a real duality. Nevertheless, we call this the Hodge dual.*

Example 5.62. *In time space, the "inner product" is defined as* $(t,x,y,z)\cdot(t',x',y',z') = -tt'+xx'+yy'+zz'$*. In particular, it is NOT really an inner product, as it is NOT positive-definite. Nevertheless, you can define hodge dual accordingly. Then* $*(dtdx) = -dydz, *(dtdy) = -dzdx, *(dtdz) = -dxdy, *(dxdy) = dtdz, *(dydz) = dtdx, *(dzdx) = dtdy$*. Here the first three has an extra negative sign than expected, because of the negative sign in our time-space "inner product".*

Example 5.63. *The full Maxwell's equation in vacuum is the following: we can describe the electromagnetic structure using a 2-form, the so-called electromagnetic tensor. Note that the space-time is 4-dim, and thus a 2-tensor can be represented as a 4 by 4 matrix, give as* $F = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix}$*, where* E_i *and* B_i *are describing the electric field and the magnetic fields. (Note that the entries related to time is electrical, whereas the purely spatial entries are all about the magnetics.) Note that as a matrix it is skew-symmetric, so as a tensor it is alternating, and thus a 2-form.*

The Maxwell's equation in vacuum says that any such F *must have* $dF = d(*F) = 0$*. No proof is required, since all you need in physics are experiments. However, note that* $dF = 0$ *implies that* F *is exact. In fact* $F = d(A^\alpha)$ *where* A^α *is the electromagnetic 4-potential, defined as* (ϕ, A) *where* ϕ *is the electric potential and* A *is the magnetic potential (a vector potential), such that* $curl(A) = B$ *the magnetic field, and* $-\nabla\phi - \frac{\partial}{\partial t}A = E$*, the electic field.*

Let us see how this acts out. Consider $dF = 0$*. Recall that as a (0,2)-tensor, we write a matrix for* F *as* $\begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix}$*. (Btw, I am taking the speed of light as the unit here.) This means that* $F = dt \wedge (E_x dx + E_y dy + E_z dz) - B_z dxdy - B_y dzdx - B_x dydz$*. Taking exterior derivative, the electric portion leaves* $-dt \wedge d(E_x dx + E_y dy + E_z dz) = -dt \wedge (curl(E) \cdot \begin{pmatrix} dydz \\ dzdx \\ dxdy \end{pmatrix})$*. The magnetic portion leaves* $-dt \wedge (\frac{\partial B}{\partial t} \cdot \begin{pmatrix} dydz \\ dzdx \\ dxdy \end{pmatrix}) - div(B)dxdydz$*. Now, the coefficients for the basis* $dtdxdy, dtdydz, dtdzdx, dxdydz$ *must be zero. So in particular, we obtained the Gauss's law for magnetism* $div(B) = 0$ *and the Maxwell-Faraday equation* $\frac{\partial B}{\partial t} + curl(E) = 0$*.*

Here the Gauss's law says that, if you treat $B_z dxdy + B_y dzdx + B_x dydz$ *as a 2-form in* $\mathbb{R}^3$*, then its exterior derivative is zero. In particular, the magnetic lines are all closed loops, since they cannot have boundaries! And the Maxwell-Faraday equation says that a changing magnetic field is related to the exterior deriavtive of the 1-form* $E_x dx + E_y dy + E_z dz$ *in* $\mathbb{R}^3$*. Note that boundaries are where things start. So changing magnetic field would GENERATE electric field.*

What about the Hodge dual? The hodge dual sends, say, $dtdx$ *to* $dydz$*. So the* $(1,2)$ *entry of the matrix for* F *is sent to the* $(3,4)$*-entry. In general, there might be negative signs involved. You can verify that* $*F$ *can be represented by the matrix* $\begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & -E_z & E_y \\ B_y & E_z & 0 & -E_x \\ B_z & -E_y & E_x & 0 \end{pmatrix}$*. And as a 2-form we have* $*F = -dt \wedge (B_x dx + B_y dy + B_z dz) + E_z dxdy + E_y dzdx + E_x dydz$*. As you can see, the hodge dual does have a feeling of duality, yes? Now in vacuum, we have* $d * F = 0$*, which dually implies that* $div(E) = 0$ *and* $\frac{\partial E}{\partial t} - curl(B) = 0$*. The first one is Gauss's law for electricity, which states that electrical circuits are closed, as they have no boundary. The second one is the Ampère's circuit law, which states that a change in electric field would GENERATE magnetic field.*

Note that the term "boundary" here has multiple meanings. For a vector field, say B, you may think of it as $B_x dx + B_y dy + B_z dz$ and take curl. Then you are looking at a folliation of $\mathbb{R}^3$ by surfaces, orthogonal to the magnetic lines everywhere. If you think of it as $B_x dydz + B_y dzdx + B_z dxdy$, then you are looking at a folliation of $\mathbb{R}^3$ by curves, i.e., the familiar magnetic lines. The latter view usually feels more natural for us, whereas for the electric field, the first view usually feels more natural.

The above situations are only true in vacuum. In general, $d * F = 0$ may fail due to the so-called charge density and current density, together they form a 3-form $J = \phi dxdydz + j_1 dtdydz + j_2 dtdzdx + j_3 dtdxdy$. As you can see, the "time" coordinate, the first one, is actually completely spatial, where as the other three coordinates must form a spatial 2-form that is also related to time, i.e., it has field lines and they are flowing. Then the Maxwell's equation becomes $d * F = J$.

Example 5.64. *What does it mean to have $dF = d * F = 0$? Consider the case of a 1-form in $\mathbb{R}^2$ with $dF = d * F = 0$. Then F and $*F$ are both folliations of curves in $\mathbb{R}^2$, and by definition of the Hodge star, the curves for F are orthogonal to the curves of $*F$, and they should have the same density everywhere. Furthermore, none of the curves involved could have boundary, because $dF = d * F = 0$.*

One example of solution is $F = dx$, where all curves are verticle, and then $*F = dy$, where all curves are horizontal. Another example with a singularity is defined on $\mathbb{R}^2$, with $F = dr/r$. Then F corresponds to circles with greater density as you get closer to the origin, and $*F$ are rays from the origin to infinity.

6 Linear Algebra Over Finite Field

6.1 Rubik's cube

We hereby tries to solve the Rubik's cube. Note that we seek to understand its structure, rather than to be fast.

The first thing you should notice is that there are twelve possible basic moves. For each face, you can turn it clockwise or counterclockwise.