Linear Algebra Lecture Notes

jongman@gmail.com

January 19, 2015

This lecture note summarizes my takeaways from taking Gilbert Strang's Linear Algebra course online.

1 Solving Linear Systems

1.1 Interpreting Linear Systems

Say we have the following linear system:

$$\begin{bmatrix} 2 & 4 & 1 \\ 1 & 3 & 7 \\ 0 & 5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 35 \\ 14 \end{bmatrix}$$

There are two complementing ways to interpret this.

1.1.1 The Row-wise Interpretation

In the classical row-wise picture, each equation becomes a hyperplane (or a line) in a hyperspace (or space). For example, $2x_1 + 4x_2 + x_3 = 14$. The solution is where the three hyperplanes meet.

1.1.2 The Column-wise Interpretation

In the column-wise picture, we think in terms of column vectors. We want to represent the right hand side as a linear combination of column vectors of A.

1.2 Elimination

Elimination is a series of row operations that will change your given matrix into a upper-triangular matrix. The allowed operations are as of follows.

- Adding a multiple of a row to another row
- Changing the order of rows

Elimination, combined with back substitution, is how software packages solve systems of linear equations.

1.2.1 Row Operations and Column Operations

Say you are multiplying a 1×3 row vector with a 3×3 matrix. What is the result? It is a linear combination of rows of the matrix.

$$\begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = 1 \times [a, b, c] + 3 \times [d, e, f] + 2 \times [g, h, i]$$

Similarly, multiplying a matrix with a column vector on its right side gives us a liner combination of columns of the matrix. We conclude that multiplying on the left will give us row operations; multiplying on the right gives us column operations.

1.2.2 Representing Elimination With A Series of Matrices

Since elimination is a purely row-wise operation, we can represent it with a series of multiplication operations on the left of the matrix. The matrices that are multiplied to do the elimination are called the elementary matrices or the permutation matrices, depending on what they are trying to do. Now, the elimination process of a 3×3 matrix A can be represented as:

$$E_{23}E_{12}A = U$$

Now, keep an eye on $E_{23}E_{12}$: if you multiply these together, it will be a single matrix that will do the entire elimination by itself!

1.2.3 Side: Multiple Interpretations of Matrix Multiplications

Say we are multiplying two matices $A \times B = C$. Multiple ways to "interpret" this operation:

- Dot product approach: $C_{ij} = \sum A_{ik}B_{kj}$ (all indices are row first)
- Column-wise approach: $C^i = A \times B^i$. Columns of C are linear combinations of columns in A.
- Row-wise approach: $C_i = A_i \times B$. Rows of C are linear combinations of rows in B.
- Column multiplied by rows: Note that a column vector multiplied by a row vector is a full matrix. Now, we can think of C as a sum of products between ith column of A and ith row of B!
- Blockwise: If we split up A and B into multiple blocks where the sizes would match, we can do regular multiplication using those blocks! If A and B were both split into 2×2 chunks, each block being a square. Then, $C_{11} = A_{11}B_{11} + A_{12}B_{21}$!

1.3 Finding Inverses with Gauss-Jordan

Say we want to find an inverse of A. We have the following equation:

$$A\,[c_1 c_2 \cdots c_n] = I_n$$

where c_i is the i-th column of A^{-1}. Now, each column in I_n are linear combinations of columns of A - namely, ith column of I_n is Ac_i. So each column in I_n gives us a system of linear equations, that can be solved by Gauss elimination. The way of solving n linear systems at once is called the Gauss-Jordan method. We work with an augmented matrix of form $[A|I_n]$ and we eliminate A to be I_n. We can say:

$$\prod E_j \, [A|I_n] = [I_n|?]$$

Say we found a set of Es that make the above equation hold. Then, we got:

$$\prod E_j A = I_n \iff \prod E_j = A^{-1} = \prod E_j I_n$$

the last equality telling us that the right half of the augmented matrix after the elimination is A^{-1}, thus proving the validity of the algorithm.

1.3.1 Inverses of Products and Transposes

What is $(AB)^{-1}$? We can easily see $B^{-1}A^{-1}$ is the answer because:

$$B^{-1}A^{-1}AB = I = ABB^{-1}A^{-1}$$

Now, $(A^T)^{-1}$? We can start from

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

and transpose both sides. Now we get

$$\left(A^{-1}\right)^T A^T = I^T = I$$

So $\left(A^{-1}\right)^T$ is the inverse of A^T.

1.4 Elimination = Factorization; LU Decomposition

Doing a Gaussian elimination to a matrix A will reduce it to a upper triangular matrix U. How are A and U related? LU decomposition tells us that there is a matrix L that connects the two matrices. (Also, note we ignore row exchanges for now)

We take all the elimination matrices used to transform A to E. Since row exchanges are not allowed, all of these elements are lower triangular. (We take an upper row, and add an multiple of it to a lower row).

Example: let

$$A = \begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix}$$

We can represent the elimination with a elimination matrix:

$$\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} A = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$$

Thus

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 4 & 7 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}}_{E^{-1}} \underbrace{\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}}_{U}$$

We can factor out a diagonal matrix so U will only have ones on the diagonal:

$$A = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

1.4.1 A case for 3×3 matrices

Say $n = 3$. Then we can represent the elimination is

$$E_{32}E_{31}E_{21}A = U$$

Now, the following holds:

$$A = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}U = LU$$

and we call the product of elimination matrices L. Question: how is L always lower triangular? Let's start with $E_{32}E_{31}E_{21}$. Since each elimination matrix subtracts an upper from from lower rows - so everything is moving "downwards". A nonzero number cannot move "up".

Now how do we calculate L? We go through an example, and we make the following claim: if you are not using row exchanges, the multipliers will go into L! That means, the multiplier we used to make U_{21} zero will go into L_{21}. This is checked with an example.

1.4.2 Time Complexity for Gaussian elimination

Of course, the naive way gives $O\left(n^3\right)$. To eliminate using the i-th (0-based) row, we would have to change $(n-i)^2$ cells. So, $n^2 + (n-1)^2 + \cdots + 1^2$. Integrating n^2 gives $\frac{1}{3}n^3$ – which is the ballpark range for this sum.

1.5 Permutation Matrices

Say we have a list of all possible permutation matrices for a matrix of size $n \times n$. There are $n!$ possibilities: we have $n!$ different permutations.

What if we multiply two of those? The result is another permutation, so this set is closed under multiplication.

What if we invert one of those? The result is also a permutation - so this is closed under inversion as well.

1.5.1 On Transpose and Symmetric Matrices

Trivial result: A^TA is always symmetric for any A. Proof?

$$\left(A^TA\right)^T = A^TA^{TT} = A^TA$$

1.5.2 Permutation and LU matrices

How do we account for row exchanges in LU decomposition? We exchange rows before we start!

$$PA = LU$$

2 Vector Spaces and Subspaces

2.1 Definitions and Examples

- Closed under scalar multiplication
- Closed under vector addition

Some things are obvious:

- Closed under linear combination
- Contains 0

2.2 Subspaces

A subspace is a subset of a vector space, which is a vector space itself. For example, $\mathbf{R}^2$ has three kinds of subsets: the plane itself, any line that goes through $(0,0)$, and $\{(0,0)\}$.

If you take a union of two subspaces, in general, it is not a subspace. However, their intersection is still a subspace.

2.3 Spaces of a matrix

Given a vector A, the set of all possible linear combinations of its vectors is called the column space of A: $C(A)$. A column space is inherently related to the solvability of linear systems. Say we want to solve $Ax = b$. Any possible value of Ax is in the column space by definition; so $Ax = b$ is solvable iff when $b \in C(A)$.

The Null space is defined by the solution space of $Ax = 0$. Null and column spaces can have different sizes: if $A \in R^{m \times n}$, $N(A) \in R^n$ and $C(A) \in R^m$. How do we find column and null spaces? Once again, elimination.

2.4 Finding the Null Space: Solving $Ax = 0$

Say we eliminated A to get a "row echelon form" matrix U. Here are some definitions.

Pivot columns The columns that contain the pivots used in the elimination.

Free columns The rest of the columns.

Rank of a matrix Number of pivot columns.

Why are free columns called "free"? In solving $Ax = 0$, we can assign arbitrary values to the variables associated with free columns. The rest of the variables will be uniquely defined from those values.

To find the entire null space, we construct a "particular" solution for each of the free variables. We can set its value 1, with rest of the free variables 0. Now we can get a particular solution. We repeat for all the $n - r$ free variables, and take their linear combination. We now know this set "spans" all possible values for the free variables.

2.4.1 Reduced-row Echelon Form

Reduced-row echelon form does one more elimination upwards, and make the pivots 1. Making pivots won't change the solution since we are solving $Ax = 0$. Also, note that since every entry (except the pivot itself) of the pivot column is eliminated, if we take the pivot rows and columns we get an identity matrix of size r.

The typical form of an rref can be shown as a block matrix:

$$R = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$$

where I is the identity matrix, the pivot part, and F is the free part. Note that you can read off the particular solutions directly off the matrix: now each row of the equation $Rx = 0$ takes the following form:

$$x_p + ax_{f_1} + bx_{f_2} + cx_{f_3} + \cdots = 0$$

where x_p is a pivot variable, and x_{f_i}s are the free variables. And now getting the value of x_p is extremely easy.

We can abstract even further; think of a null space matrix $\mathcal{N}$ such that $R\mathcal{N} = 0$. Each column of this matrix is the particular solution. And we can set:

$$\mathcal{N} = \begin{bmatrix} -F \\ I \end{bmatrix}$$

From the block multiplication, we know $R\mathcal{N} = 0$ and how each column of $\mathcal{N}$ looks like.

2.5 Solving $Ax = b$

Now we solve a generic linear system. First, some solvability condition: we can solve it if $b \in \mathcal{C}(A)$.

Finding the solution space is pretty simple:

- Find a particular solution by setting all free vars to 0, and solving for pivot variables.
- Add it to the null space!

Is the solution space going to be a subspace? No, unless it goes through origin.

2.5.1 Rank and Number of Solutions

The key takeaway is that you can predict the number of solutions by only looking at the rank of the matrix.

Say A is a $m \times n$ matrix. What is the rank r?

- Full rank square matrix: When $r = n = m$, the rref becomes I and we have exactly one solution for any b in $Ax = b$.
- Full column rank: This happens in "tall" matrices, where $r = n < m$. The rref looks like $\begin{bmatrix} I \\ 0 \end{bmatrix}$ and we have no free variables, so $\mathcal{N}(A) = \left\{\vec{0}\right\}$. Also, for any b, there might be 0 solutions (when the zero row should equal a nonzero b_i) or exactly one solution.
- Full row rank: This happens in "wide" matrices, where $r = m < n$. The rref looks like $\begin{bmatrix} I & F \end{bmatrix}$ (in practice, the columns of I and F are intertwined.). Since we have no zero rows, the number of solutions is not going to be 0. Also, since there is a free variable, we are always getting a ∞ number of solutions.
- Not-full column/row rank: The rref looks like $\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$. We get 0 or ∞ solutions depending on b.

3 Linear Independence, Basis and Dimension

First, a lemma: when $A \in R_{m \times n}$ where $m < n$, there is a nonzero solution to $Ax = 0$ since we always get a free variable.

3.1 Linear Independence

Linear independence A set of vectors $v_1, v_2, \cdots, v_n$ is linearly independent when no linear combination of them (except for the 0 combination) result in a 0 vector.

Equivalently, say $A = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$. The column vectors are linearly independent:

- $\mathcal{N}(A) = \{0\}$
- A is full column rank.

A corollary of the above lemma: 3 vectors in a 2-dimensional space cannot be independent.

3.2 Spanning, Basis, and Dimensions

Definition: A set of vectors $\{v_1, v_{2,} \cdots, v_l\}$ span a space iff the space consists of all linear combinations of those vectors.

A basis for a space is an independent set of vectors $\{v_1, v_2, \cdots, v_d\}$ which span the space. The number of vectors d is called the dimension of the space. Here are some facts:

- A set of vectors in $\mathbf{R}^n$ $\{v_1, \cdots, v_n\}$ gives a basis iff the $n \times n$ matrix with those as columns gives an invertible matrix.
- Every basis has the same number of vectors, the number being the dimension of the space.

3.2.1 Relationship Between Rank and Dimension

The rank of matrix A is the number of pivot columns. At the same time, it is the dimension of the column space of A - $\mathcal{C}(A)$.

OTOH, what is the dimension of $\mathcal{N}(A)$? For each free variable, we get a special solution with that free variable set to 1 and other free variable to 0. Each of these special solutions are independent, and span the null space! So the dimension of $\mathcal{N}(A) = n - r$.

3.3 Four Fundamental Spaces of A

Given a $A \in \mathbf{R}^{n \times m}$, here are the four fundamental spaces.

- Column space $\mathcal{C}(A) \in \mathbf{R}^n$
- Null space $\mathcal{N}(A) \in \mathbf{R}^m$
- Row space $\mathcal{R}(A)$ = all linear combinations of rows of A = $\mathcal{C}\left(A^T\right) \in \mathbf{R}^m$
- Null space of A^T - often called the "left null space of A": $\mathcal{N}\left(A^T\right) \in \mathbf{R}^n$

  This is called the left null space because, $\mathcal{N}\left(A^T\right) = \left\{y | A^T y = 0\right\} = \left\{y | y^T A = 0\right\}$ the latter equality derived from taking the transpose.

We should understand what dimensions and basis do they have; and also how do they relate with each other. First, the dimension and the basis:

- $\mathcal{C}(A)$ has dimension r, and the pivot columns are the basis.
- $\mathcal{N}(A)$ has dimension $n - r$, and the special solutions are the basis.
- $\mathcal{R}(A)$ has dimension r, and the first r rows in the rref are the basis.
- $\mathcal{N}\left(A^T\right)$ has dimension $m - r$, and the last $m - r$ rows in the elementary matrix E s.t. $EA = R$. IOW, the E that comes out of Gauss-Jordan.

Note that summing the dimension of $\mathcal{C}(A)$ and $\mathcal{N}(A)$ gives n, where summing dimensions of $\mathcal{R}(A)$ and $\mathcal{N}\left(A^T\right)$ gives m!

3.3.1 Elimination and Spaces

What does taking rref do to the matrix's spaces? We do know $\mathcal{C}(A) \neq \mathcal{C}(R)$ since the last row of R can potentially be zero. However, elimination does not change the row space; and its basis is the first r rows in R.

3.3.2 Sets of Matrices as Vector Space

Suppose the following set: all sets of 3×3 matrices. It is a vector space because it is closed under addition, multiplication by scalar!

What are some possible subspaces? All upper triangular matrices, all symmetric matrices, all diagonal matrices, multiples of I, ..

We discuss an interesting properties of these different subspaces:

$$\dim(A+B) = \dim A + \dim B - \dim(A \cap B)$$

3.3.3 Solutions As Vector Space

What are the solutions of the following differential equation?

$$\frac{\partial^2 y}{\partial x^2} + y = 0$$

We can say: $y = a\sin x + b\cos x$. Now, we can think of the solution as a vector space with $\sin x$ and $\cos x$ as their basis. Note these are functions, not vectors! So this is a good example why the idea of basis and vector spaces plays a large role outside the world of vectors and matrices.

3.3.4 Rank One Matrices

Rank one matrices are special, as we can decompose it into a product of column vector and row vector $A = uv^T$. They are "building blocks" of other higher-rank matrices. A four-rank matrix can be constructed by adding four rank-one matrices together.

4 Applications

4.1 Graphs and Networks

"The" most important model in applied math! A graph with m edges and n nodes can be represented by an incidence matrix of size $m \times n$, each row representing an edge. The entry $A_{ij} = 1$ if edge i is coming into node j, -1 if edge i leaves node j, and 0 otherwise. Note this is different from the adjacency matrix form I am used to.

Some remarks about how notions in electric network and linear algebra concepts relate to each other.

- Loops: When 2 or more edges form a loop (not necessarily a cycle), those rows are _not_ independent and vice versa.
- Sparsity: This is a very sparse matrix! However, in applied linear algebra, it is way more common to have structured matrices.
- Null space: Say x is a vector of potentials for each nodes - then Ax means differences between potentials of nodes. So $Ax = 0$ gives you pairs of nodes for which the potentials are the same. If the graph is connected, the null space is a single dimensional space - $c\vec{1}$.
- $\mathcal{N}(A^T)$: What does $A^T y = 0$ mean? Kirchhoff's circuit law is like the flow preservation property in electrical networks; net incoming/outgoing current is zero for any node. If y is a vector of currents, $A^T y$ is a vector of net incoming currents.
  - The basis of this null spaces are related to the loops in the graph. Suppose we pick a loop and send a unit current along it. This gives us a basis.

  - Repeat, and we can take all loops one by one and take all the basis!
  - Say there are two loops: $a-b-c-a$ and $a-b-e-a$. Is $a-c-b-e-a$ a valid loop? No, it is the sum (or difference) of two loops and the special solution will be dependent of the previous two special solutions.
  - Now, realize that the number of the loop is $\dim \mathcal{N}\left(A^T\right) = m - r$!

- $\mathcal{R}\left(A^T\right)$: What do the pivot columns represent? *It is a fucking spanning tree*! Whoa If there is a cycle those rows are not going to be independent!

Taking all of this together, we can derive Euler's formula!

- $\dim \mathcal{N}\left(A^T\right) = m - r$
- $r = n - 1$ (since the pivot columns represent a spanning tree which always have $n-1$ edges!)
- Then

$$\text{Number of loops} = \text{Number of edges} - (\text{Number of nodes} - 1)$$

  Now, in graph theory speak:

$$F = E - V + 1 \iff V - E + F = 1$$

  Ladies and gentlemen, let me introduce Euler's formula. Holy crap.

Also, we can merge everything in a big equation. So far, we know:

- $e = Ax$ (potential differences)
- $y = Ce$ (Ohm's law)
- $A^T y = f$ (Kirchhoff's law)

So we get:

$$A^T C A x = f$$

whatever that means haha.

5 Orthogonality

5.1 Definition

What does it mean for two subspaces/vectors/basis to be orthogonal?

Vector orthogonality: Two vectors x and y are orthogonal iff $x^T y = 0$. We can connect this to Pythagorean theorem - $\|x\|^2 + \|y\|^2 = \|x+y\|^2$ iff x and y are orthogonal.

$$\|x\|^2 + \|y\|^2 = x^T x + y^T y = \|x+y\|^2 = x^T x + y^T y + 2x^T y \iff x^T y = 0$$

Subspace orthogonality: two subspace S is orthogonal to subspace T when every vector in S is orthogonal to every vector in T. Examples:

- Are xz and yz planes orthogonal in $\mathbf{R}^3$? No they aren't: they have a nonzero intersection! The vectors in that intersection cannot be orthogonal to themselves.

Facts:

- Row space orthogonal to null space. Why? $x \in \mathcal{N}(A)$ iff

$$Ax = \begin{bmatrix} r_1 \cdot x \\ \vdots \\ r_n \cdot x \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

  So x is orthogonal to all rows. And of course, it will be orthogonal to all linear combinations of the rows.

- Row space and null spaces are orthogonal complements of $\mathbf{R}^3$: nullspace contains *all* vectors orthogonal to the row space.

5.2 Projections

5.2.1 Why Project?

If $Ax = b$ cannot be solved in general. So what do we do? We find the closest vector in $\mathcal{C}(A)$ that is closest to b, which is a projection!

5.2.2 2D Case

Suppose we project vector b onto a subspace that is multiple of a. Say that the projected point is ax. We know the vector from ax to b is orthogonal to a. We have:

$$0 = a^T(ax - b) = a^T ax - a^T b \iff x = \frac{a^T b}{a^T a}$$

Look at the projected point ax:

$$ax = a\frac{a^T b}{a^T a} = \left(\frac{aa^T}{a^T a}\right) b$$

Note the last formula - that rank one matrix is the "projection" matrix P! It has the following properties:

$$P^2 = P \text{ and } P^T = P$$

5.2.3 General Case

It's the same derivation! If p is the projected point, we can write it as $A\hat{x}$. Then the error vector $b - A\hat{x}$ is perpendicular to the *column* space of A. So:

$$0 = A^T(b - A\hat{x}) = A^T b - A^T A\hat{x} \iff \hat{x} = \left(A^T A\right)^{-1} A^T b$$

Welcome pseudoinverse! Now, get p by

$$p = A\hat{x} = \underbrace{A\left(A^TA\right)^{-1}A^T}_{P}b$$

We got the projection matrix!

5.3 Least Squares

When A is tall, $Ax = b$ is not generally solvable exactly. We multiply A^T to both sides of the equation to get

$$A^TAx = A^Tb$$

Hoping A^TA is invertible and we can solve this exactly. When is A^TA invertible? It turns out rank of A^TA is equal to rank of A - so A^TA is invertible only when A is *full column rank*.

5.3.1 Invertibility of A^TA

Assume A is full column rank. Let's prove the following set of equivalent statements:

$$A^TAx = 0 \iff x = 0 \iff \mathcal{N}\left(A^TA\right) = \{0\} \iff A^TA \text{ is invertible}$$

Take the first equation, and calculate dot products of each side with x:

$$x^TA^TAx = 0 = (Ax)^T Ax \iff Ax = 0$$

Since A is full column rank, $\mathcal{N}(A) = \{0\}$. So x must be 0.

5.3.2 Least Squares as A Decomposition

If least square decomposes a vector b into $p+e$, $p \in \mathcal{C}(A)$ and $e \in \mathcal{N}(A)$. Now, if $p = Pb$, what is e? Of course, $e = (I - P)b$. We get $p + e = (P + I - P)b = b$.

5.4 Orthonormal Basis

A set of orthonormal vectors $\{q_1, q_2, \cdots\}$ is a set of unit vectors where every pair is perpendicular. We can write this with a matrix Q:

$$Q = [q_1 q_2, \cdots]$$

where q_is are column vectors. Now, the above requirement can be written as:

$$Q^TQ = I$$

This matrix Q is called a orthonormal matrix. (Historically, we only call it orthogonal when it's a square...) What happens when Q is square? Since $Q^TQ = I$, $Q = Q^{-1}$.

5.4.1 Rationale

The projection matrix can be found as

$$P = Q\left(Q^TQ\right)^{-1}Q^T = QQ^T$$

5.4.2 Gram-Schmidt

Given a set of vectors, how can we make them orthonormal? Well... I sorta do know... tedious process to do by hand.

- Take an arbitrary vector, and normalize it. Include it in the result set.
- For each other vector v,
  - For each vector u in the result set, subtract the projection of v onto u from v: $v = v - \left(u^Tv\right)\cdot u$
  - Normalize the resulting v and include it in the result set.

5.4.3 QR Decomposition

How do we write Gram-Schmidt in terms of matrices? We could write Gaussian elimination by

$$PA = LU$$

We write this as:

$$A = QR$$

Note that, because of how Gram-Schmidt works, R is going to be a lower triangular matrix! The first column of Q is going to be the first column of A, scaled! Also, the second column is going to be a linear combination of first two columns of A and etc.

6 Determinants

6.1 Properties

1. Identity matrix has determinant 1.
2. When you swap two rows, the sign of the determinant will change.
3. The determinant is a "linear function" of the matrix: (I'm not saying $\det A + \det B = \det(A+B)$ – this is not true)
   (a) If you multiply a row by a scalar t, the determinant will be multiplied by t.
   (b) $\det\begin{bmatrix} a & \cdots & x+y & \cdots & b \end{bmatrix}^T = \det\begin{bmatrix} a & \cdots & x & \cdots & b \end{bmatrix}^T + \det\begin{bmatrix} a & \cdots & y & \cdots & b \end{bmatrix}^T$

The following can be derived from the above points:

- If two rows are equal, determinant is 0.
- If we subtract a multiple of a row from another row, determinant doesn't change. (This can be proved from 3b and 4)
- If there's a row of zeros, determinant is 0.
- For a upper triangular matrix, determinant is the product of the diagonal products. (Proof: do reverse elimination to get a diagonal matrix with same determinant, use rule 3a repeatedly until we get I)
- $\det A \iff A$ is singular
- Determinant is multiplicative: $\det AB = \det A \cdot \det B$
  - So: $\det A^{-1} = \frac{1}{\det A}$
- $\det A = \det A^T$ (this means, swapping columns will flip the sign)
  - Proof: let $A = LU$. $\det A = \det(LU) = \det L \det U$. $\det L = 1$, and $\det U = \det U^T$ since it is upper triangular.

6.2 Calculation

6.2.1 Big Formula

$$\det A = \sum_{p \in \text{permute}(1 \cdots n)} \pm \prod a_{ip_i}$$

Argh.. messy. Whatever.

6.2.2 Cofactor

Taking the big formula, and collecting terms by the number in the first row gives the cofactor expansion formula. Say the matrix is 3×3:

$$\begin{aligned} \det A = \quad & a_{11}(a_{22}a_{33} - a_{23}a_{32}) \\ + \quad & a_{12}(-a_{21}a_{33} + a_{23}a_{31}) \\ + \quad & a_{13}(a_{21}a_{32} - a_{23}a_{31}) \end{aligned}$$

Notice the quantities in parenthesis are either determinants of the 2×2 matrices, or their negatives. Formally, C_{ij} is defined to be a cofactor

$$C_{ij} = \pm \det(\text{smaller matrix with row } i \text{and col } j \text{removed})$$

where the sign is + when $i + j$ is even, - when $i + j$ is even. This follows the checkerboard pattern. (Formally, cofactors with the sign are called *minors*.) The resulting cofactor formula for the determinant is:

$$\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

6.3 Applications

6.3.1 Formula for A^{-1}

$$A^{-1} = \frac{1}{\det A} C^T$$

where C is the matrix of cofactors in A. How do we verify it? Let's check if

$$AC^T = (\det A)\, I$$

Expand the elements of the matrix - we actually see the diagonals are the determinants, from the cofactor formula!

But what about the off-diagonal elements? We claim those calculations are actually using cofactor formula for a matrix with two equal rows! Aaaaah.....

6.3.2 $Ax = b$ and Cramer's Rule

Of course, we know

$$x = \frac{1}{\det A} C^T b$$

What are the entries of x in this formula? Since $c_i b$ is always a determinant of some matrix, we can write it as:

$$x_i = \frac{\det B_i}{\det A}$$

for some matrix B_i. Cramer realized that B_i is A with column 1 replaced by b. This are beautiful, but not practical ways of calculating stuff.

6.3.3 Determinant and Volume

The determinant of a $n \times n$ matrix A is the volume of a box in a n dimensional space with n sides of it coinciding with the column/row vectors of A. Let's look at some example cases to get ourselves convinced:

- $A = I$: it's going to be a (hyper)cube with volume 1.
- $A = Q$ (orthonormal): another cube, rotated. From here we can rediscover $Q^T Q = I$ - take determinants from both sides: $(\det Q)^2 = 1$. So $\det Q = \pm 1$.

Also, we revisited many determinant properties and made sure they hold in this context as well.

7 Eigenvalues and Eigenvectors

7.1 Definition

You can think of a matrix as a linear mapping, using $f(x) = Ax$. For a matrix, we can find vector(s) x such that $f(x)$ is parallel to x. Formally:

$$Ax = \lambda x$$

where λ is a scalar, called the eigenvalue. The xs are called the eigenvectors.

7.1.1 Example: Projection Matrix

Say there is a projection matrix, which projects stuff into a plane: when a vector is already in the plane, it won't change. So they will be eigenvector with eigenvalue 1. We can say there are two perpendicular eigenvectors that span the plane. Are there any other eigenvalue? (Intuitively, we expect to find one, since we are in a 3D space) Yes, find the normal vector that goes through the origin. This will become a 0 vector - so this one is an eigenvector with eigenvalue 0.

7.1.2 Example 2: Permutation Matrix

Let

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

Trivially, $[1, 1]^T$ is an eigenvector with eigenvalue 1. Also, $[-1, 1]^T$ is an eigenvector with eigenvalue -1.

7.1.3 Properties

- $n \times n$ matrices will have n eigenvalues.
- Sum of the eigenvalues equal the trace of the matrix (sum of the diagonal entries).
- The determinant is the product of the eigenvalues.

7.2 Calculation

Rewrite the equation as:

$$(A - \lambda I)\, x = 0$$

Now, for this to be true for nontrivial x, $A - \lambda I$ has to be singular: this is equvalent to:

$$\det (A - \lambda I) = 0$$

We solve this to find λs. After that, we can use elimination to find x.

7.2.1 Properties

If we add aI to a matrix, each of its eigenvalues will increase by a. The eigenvectors will stay the same! See: if $Ax = \lambda x$,

$$(A + aI)\, x = \lambda x + ax = (\lambda + a)\, x$$

What do we know about general addition? If we know eigenvalues and eigenvectors of A and B, what do we know about $A + B$? First guess would be their eigenvalues be added, but false. Because they can have different eigenvectors.

7.2.2 When Things Are Not So Rosy

Think of a rotation matrix; what vector is parallel to itself after rotating 90 degrees? None! When we carry out the calculation, we get complex eigenvalues. Also, for some matrices we can have duplicate eigenvalues, and no independent eigenvectors. Why is this important? We will see.

7.3 Diagonalization of A Matrix

Let us assume that A has n linearly independent eigenvectors, x_1 to x_n, each associated with eigenvalues λ_1 to λ_n. We can put them in columns of a $n \times n$ matrix, and call this S. Then, we can write what we know as follows:

$$AS = A\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_n x_n \end{bmatrix}$$

How do we write the latter with a matrix representation? We can multiply S with a diagonal matrix with eigenvalues along the diagonal:

$$\begin{bmatrix} \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_n x_n \end{bmatrix} = S \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = S\Lambda$$

We call the diagonal matrix Λ. Now we have:

$$AS = S\Lambda \iff S^{-1}AS = \Lambda \iff A = S\Lambda S^{-1}$$

Note since we assumed the eigenvectors are linearly independent, S is invertible.

From this, we can infer interesting properties of eigenvalues. What are the eigenvectors and eigenvalues of A^2? Intuitively, of course, the eigenvectors are the same - for an eigenvector x, Ax is a scalar multiplication of x. So A^2x will still be a scalar multiplication, with a factor of λ^2. However, we can see it from diagonalization as well:

$$A^2 = \left(S\Lambda S^{-1}\right)^2 = S\Lambda S^{-1} S\Lambda S^{-1} = S\Lambda^2 S^{-1}$$

In general, if you take powers of a matrix, it will have the same eigenvectors, but the eigenvalues will get powered.

7.3.1 Understanding Powers of Matrix Via Eigenvalues

What does

$$\lim_{n\to\infty} A^n$$

look like? Does it ever go to 0? We can find out, by looking at its eigenvalues:

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

For A^n to go to 0, Λ^n has to go to zero - so the absolute value of all eigenvalues are less than 1!

7.3.2 Understanding Diagonalizability

A is sure to have n independent eigenvectors (and thus be diagonalizable) if all the eigenvalues are different. Note there are cases where there are repeated eigenvalues and there are independent eigenvectors (well, take I for an example).

7.4 Applications

7.4.1 Solving Recurrences

Let's solve a *difference* equation:

$$u_{k+1} = Au_k$$

with a given u_0. How to find $u_k = A^k u_0$, without actually powering the matrix? The idea is to rewrite u_0 as a linear combination of normalized eigenvectors of A:

$$u_0 = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

Now, Au_0 is:

$$Au_0 = Ac_1 x_1 + Ac_2 x_2 + \cdots + Ac_n x_n = c_1\lambda_1 x_1 + c_2\lambda_2 x_2 + \cdots$$

Yep, we got a pattern! Now we know:

$$A^k u_0 = \sum_i c_i \lambda_i^k x_i$$

The idea of eigenvalues kind of sank in after looking at this example; they, sort of, decompose the linear mapping represented by A into orthogonal basis. After you represent a random vector in this space, the effects of A can be isolated in each direction. So it actually describes A pretty well! The name "eigen" now kind of makes sense.

7.4.2 Deriving Closed Form Solution of Fibonacci Sequence

Let $f_0 = 0$, $f_1 = 1$, $f_n = f_{n-1} + f_{n-2}$. I know that I can write

$$u_n = \begin{bmatrix} f_n \\ f_{n-1} \end{bmatrix} \text{ and } u_{n+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} u_n = Au_n$$

What are the eigenvalues of A? The characteristic polynomial comes out to be

$$(1-\lambda)\cdot(-\lambda) - 1 = \lambda^2 - \lambda - 1 = 0$$

Plugging this into the quadratic formula, we get:

$$\lambda = \frac{1 \pm \sqrt{5}}{2} \colon \lambda_1 = \frac{1+\sqrt{5}}{2},\ \lambda_2 = \frac{1-\sqrt{5}}{2}$$

Since $\lambda_1 > 1$, $\lambda_2 < 1$, when n grows, λ_1 will dominate the growth of the fibonacci number. So we know:

$$f_{100} \approx c \cdot \left(\frac{1+\sqrt{5}}{2}\right)^{100}$$

Why? Rewrite u_0 as

$$u_0 = c_1 x_1 + c_2 x_2$$

and we know

$$u_{100} = c_1 \lambda_1^{100} x_1 + c_2 \lambda_2^{100} x_2$$

and with $\lambda_2 < 1$, the second term is meaningless when k is large. Whoa whoa.... nice...
For the sake of completeness, let's calculate the eigenvectors as well. The eigenvectors are:

$$x_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix},\ x_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$$

Solve for c_1 and c_2 by equating:

$$u_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = c_1 x_1 + c_2 x_2 = \begin{bmatrix} \frac{c_1 + c_2 + (c_1 - c_2)\sqrt{5}}{2} \\ c_1 + c_2 \end{bmatrix}$$

Now we have the formula for u_k - thus a closed form solution for f_k.

7.4.3 Differential Equations

Consider the following system of differential equations:

$$\frac{du_1}{dt} = -u_1 + 2u_2$$
$$\frac{du_2}{dt} = u_1 - 2u_2$$

with initial value $u_1(0) = 1$, $u_2(0) = 0$. Let $u(t) = [u_1(t), u_2(t)]^T$ and we can write above as:

$$\frac{du}{dt} = Au$$

with

$$A = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix}$$

The eigenvalues of A are 0 and -3, and the associated eigenvectors are $x_1 = [2, 1]^T$ and $x_2 = [1, -1]^T$, respectively.

Now, the solution is a linear combination of two special solutions, each corresponding with an eigenvalue:

$$u(t) = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2$$

which we can check by plugging into the above representation. We can solve for c_1 and c_2 using the initial condition - and we now know everything. So eigenvalues still give us insights about which parts of the solution blows up, or goes to 0, etc. Of course, since eigenvalues now sit on the exponent, it has to be negative to go to 0. If the eigenvalue is 0, the corresponding portion will stay constant.

Note: the eigenvalues might be imaginary - in that case, only the real portion counts in terms of asymptotic behavior. For example,

$$e^{(-3+6i)t} = e^{-3t} e^{6it} = e^{-3t} (\cos t + i \sin t)^6$$

and the latter part's magnitude is 1.

7.4.4 Thinking In Terms of S and Λ

In

$$\frac{du}{dt} = Au$$

A is a non-diagonal matrix, and represents the interaction, relation between the variables. This is *coupling*; we can *decouple* variables by using eigenvalues. Now, decompose u into a linear combination of eigenvectors by setting $u = Sv$. We get:

$$\frac{du}{dt} = S\frac{dv}{dt} = ASv \iff \frac{dv}{dt} = S^{-1}ASv = \Lambda v$$

Wow, we now have a set of equations like: $v_1' = \lambda_1 v_1$, $v_2' = \lambda_2 v_2$, and so on. Same as in the difference equation example. Now, how do we express the solution in this? Since

$$\frac{dv}{dt} = \Lambda v$$

we would like to express v as

$$v(t) = e^{\Lambda t} v(0)$$

which gives

$$u(t) = S e^{\Lambda t} S^{-1} u(0)$$

which gives the "decoupling" effect we are looking for. But WTF is that $e^{\Lambda t}$? First let's define e^{At} as:

$$e^{At} = I + At + \frac{(At)^2}{2} + \cdots + = \sum_{n=0}^{\infty} \frac{(At)^n}{n!}$$

Just like the power series definition for $e^{ax} = 1 + ax + \frac{(ax)^2}{2!} + \frac{(ax)^3}{3!} + \cdots$. Now what is $Se^{\Lambda t}S^{-1}$?

$$e^{At} = e^{(S\Lambda S^{-1})} = \sum \frac{\left(S\Lambda S^{-1}\right)^n t^n}{n!} = \sum \frac{S\Lambda^n S^{-1} t^n}{n!} = S\left(\sum \frac{\Lambda^n t^n}{n!}\right) S^{-1} = Se^{\Lambda t} S^{-1}$$

So it's same as e^{At}!

7.5 Applications: Markov Matrices and Fourier Series

7.5.1 Markov Matrix

What is a Markov Matrix? Nonnegative square matrix, with all columns summing up to 1. Now, we know those processes are never going to blow up. Maybe they will reach a steady state. We already know:

- Any eval cannot be greater than 1, because it will make things blow up.
- If an eval equals 1, the evec for that eval will be the steady state.

We will find out that Markov matrices always have 1 as an evalue. Also, it will never be a repeated eigenvalue.

7.5.2 Proof of 1 Being An Eigenvector

Since A has 1 as column sums, $A - 1I$ has zero sum columns. Now, the rows are dependent: add up all rows and they sum to 0. This leads us to the corollary that A and A^T has same eigenvalues (because $\det(A - \lambda I) = \det\left((A - \lambda I)^T\right) = \det\left(A^T - \lambda I\right)$)

7.5.3 Projections with Orthonormal Basis

If we have a set of orthonormal basis, arbitrary vector v can be represented as:

$$v = Qx$$

Since Q is a projection matrix, $Q^{-1} = Q^T$. So we can solve for x as: $x = Q^T v$
Nothing Fourier related so far. What now?

7.5.4 Fourier Series

We write a given function f as a linear combination of sin and cos:

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots$$

This infinite series is called the Fourier series. We now work in function space; instead of orthogonal vectors we use orthogonal functions: 1, $\cos x$, $\sin x$, $\cos 2x$, and so on. We represent a function with a linear combination of those basis functions.

But what does it mean that two functions are orthogonal? How is dot product defined between two functions? We define:

$$f^T g(x) = \int f(x)\, g(x)\, dx$$

We can calculate this between constants, sines and cosines because they are all periodic.

Now, a_0 is easy to determine - take the average value. What is a_1?

$$\int_0^{2\pi} f(x)\cos x dx = a_1 \int_0^{2\pi} \cos^2 x dx \implies a_1 = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos x dx$$

where the latter equality comes from the fact the basis functions are orthogonal. (Btw, I'm not sure how we can fix the bounds on the above integral. Maybe I should go back to the book.)

7.6 Symmetric Matrices and Positive Definiteness

When $A \in S_n$ where S_n is the set of $n \times n$ symmetric matrices, we state:

- All eigenvalues are real
- We can choose eigenvectors such that they are all orthogonal.

The exact proof is left to the book.

The usual diagonalization, $A = S\Lambda S^{-1}$ now becomes $A = Q\Lambda Q^{-1} = Q\Lambda Q^T$, the latter equality coming from the fact that Q has orthonormal columns, so $QQ^T = I$. This is called the spectral theorem; spectrum means the eigenvalues of the matrix.

7.6.1 Proof

Why real eigenvalues? Let's say

$$Ax = \lambda x$$

We can take conjugates on both sides;

$$\bar{A}\bar{x} = \bar{\lambda}\bar{x}$$

However, since we assume A to be real, we know:

$$\bar{A}\bar{x} = A\bar{x} = \bar{\lambda}\bar{x}$$

Try using symmetry, by taking transposing stuff:

$$\bar{x}^T A^T = \bar{x}^T A = \bar{x}^T \bar{\lambda}$$

The second equality coming from symmetry assumption. Now, multiply $\bar{x}^T$ to the both sides of the first equation:

$$\bar{x}^T Ax = \lambda \bar{x}^T x$$

And we multiply x to the right side of the last equality:

$$\bar{x}^T Ax = \bar{\lambda}\bar{x}^T x$$

Then

$$\lambda \bar{x}^T x = \bar{x}^T A x = \overline{\lambda} \overline{x}^T x$$

Now, we know:

$$\lambda = \bar{\lambda}$$

thus λ is real - if $\bar{x}^T x$ is nonzero.

$$\bar{x}^T x = \sum_i (a + ib)(a - ib) = \sum_i a^2 + b^2 > 0$$

(unless x is 0. But if x is 0, $\lambda = 0$)

7.6.2 When A is Complex

We repeat the above argument, without assuming $A = \bar{A}$. The last equality can work if:

$$A = \overline{A}^T$$

7.6.3 Rank One Decomposition

Recall, if A is symmetric, we can write:

$$A = Q \Lambda Q^T = [q_1 q_2 \cdots] \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \end{bmatrix} = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots$$

So every symmetric matrix can be decomposed as a linear combination of perpendicular projection (rank one) matrix!

7.6.4 Pivots And Eigenvalues

Number of positive/negative eigenvalues for symmetric matrices can be determined from the signs of the pivots. The number of positive eigenvalues is the same as the number of positive pivots.

7.6.5 Positive Definiteness

A PSD matrix is a symmetric matrix. If symmetric matrices are "good" matrices, PSD are "excellent". It is a symmetric matrix with all eigenvalues are positive. Of course, all the pivots are positive.

So, for 2×2 matrices, PSD matrices always have positive determinants and positive trace. However, this is not a sufficient condition for positive definiteness, as can be demonstrated in the following matrix:

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

We state that a matrix is positive definite iff all its subdeterminants are positive; they determinants of submatrices formed by taking a $m \times m$ submatrix from the left top corner.

To summarize:

- All eigenvalues are positive
- All pivots are positive
- All subdeterminants are positive

7.7 Complex Numbers and Examples

Introducing complex numbers and FFT.

7.7.1 Redefining Inner Products for Complex Vectors

If $z \in \mathbf{C}^n$, $z^T z$ is not going to give me the length of the vector squared, as in real space. Because $z^T z = \sum_j (a_j + ib_j)^2 = \sum_j a_j^2 - b_j^2 \neq \|z\|$. As seen in the proof of real eigenvalues of symmetric vectors, we need: $\bar{z}^T z$.

For simplicity, we write:

$$\bar{z}^T z = z^H z = \sum \|z_i\|^2 = \|z\|^2$$

where H stands for Hermitian. So, from now on, let's use Hermitian instead of usual inner product.

7.7.2 Redefining Symmetric Matrices

We also claim our notion of symmetric matrix $A = A^T$ is no good for symmetrix matrix. We want $\bar{A}^T = A$ for obvious reasons (so the diagonal elements of $\bar{A}^T A$ are $\|a_i\|^2$).

We define a *Hermitian matrix* A to satisfy:

$$A^H \triangleq \bar{A}^T = A$$

7.7.3 Orthonormal Basis

Now, for a matrix Q with orthonormal columns, we say:

$$Q^H Q = I$$

Also we call these matrices *unitary*.

7.7.4 Fourier Matrices

$$F_n = \begin{bmatrix} 1 & 1 & 1 & \cdots \\ 1 & w & w^2 & \cdots \\ 1 & w^2 & w^4 & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ 1 & w^{n-1} & w^{2(n-1)} & \cdots \end{bmatrix}$$

to generalize, we have $(F_n)_{ij} = w^{ij}$. Note both indices are 0-based. Also, we want $w^n = 1$ - nth primitive root of unity. So we use

$$w = \cos\frac{2\pi}{n} + \sin\frac{2\pi}{n} = e^{i(2\pi/n)}$$

(Of course I know we can use modular arithmetic to find a different primitive root of unity.. but meh.)

One remarkable thing about Fourier matrices is that their columns are orthonormal. Then the following is true:

$$F_n^H F_n = I \iff F_n^{-1} = \bar{F_n}^T$$

which makes it easy to invert Fourier transforms!

7.7.5 Idea Behind FFT

Of course, FFT is a divide-and-conquer algorithm. In the lecture, larger order Fourier matrices are connected to a smaller order matrice by noticing:

$$F_n = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_{n/2} & 0 \\ 0 & F_{n/2} \end{bmatrix} P$$

for which P is a odd-even permutation matrix. Exploiting the structures of these matrices, this multiplication can be done in linear time.

7.8 Positive Definiteness and Quadratic Forms

7.8.1 Tests of Positive Definiteness

These are equivalence conditions of positive definiteness for a symmetric matrix A:

1. All eigenvalues are positive.
2. All subdeterminants are positive.
3. All pivots are positive.
4. $x^T Ax > 0$ for all x.

(Also, if all the positiveness in above definition is swapped by nonnegativeness, we get a positive semidefinite matrix.)

Now what's special with the new 4th property? Actually, property 4 is the definition of positive definiteness in most texts; property 1-3 are actually just the tests for it. What does the product $x^T Ax$ mean? If we do it by hand we get:

$$x^T Ax = \sum_{i,j} A_{ij} x_i x_j$$

This is called the quadratic form. Now, the question is: is this positive for all x or not?

7.8.2 Graphs of Quadratic Forms

Say $x \in \mathbf{R}^2$ and let $z = x^T Ax$; what's the graph of this function? If x is not positive (semi)definite, we get a saddle point. A saddle point is a maximum in one direction whilest being a minimum in another. (Actually these directions are eigenvalue directions.)

What happens when we have a positive definite matrix? We know z will be 0 at the origin, so this must be the global minimum. Therefore, we want the first derivative to be all 0. However, this is not enough to ensure minimum point.

We want to refactor quadratic form as a sum of squares form ("the completing the squares" trick). So we can ensure that $x^T Ax$ is positive everywhere except 0, given a particular example. But how do we do it in a general case? The course reveals that, actually, Gaussian elimination is equivalent to completing the squares! Holy crap... Positive coefficients on squares mean positive pivots, which means positive definiteness!

And if we try to recall Calculus we were presented with a "magic" formula called the second derivative test - which was just checking if second derivative matrix was positive definite! Niiicee.

7.8.3 Geometry of Positive Definiteness; Ellipsoids

If a matrix is positive definite, we know $x^T Ax > 0$ for except $x = 0$. If we set $x^T Ax = c$ for a constant c, this is an equation of a ellipse - or an ellipsoid.

The major/middle/minor axis of this ellipsoid will be determined by the eigenvectors of A, their lengths being determined by the eigenvalues.

7.8.4 $A^T A$ is Positive Definite!

Yep, $A^T A$ is always positive definite regardless of A. Because covariances.. but here's a proof:

$$x^T A^T Ax = (Ax)^T (Ax) = \|Ax\|^2 > 0 \text{ unless } x = 0$$

7.9 "Similar" Matrices

7.9.1 Definition

Two $n \times n$ matrices A and B are similar, if for some invertible M, $A = MBM^{-1}$. Example: A is similar to Λ because $\Lambda = S^{-1}AS$. Why are they similar? *They have the same eigenvalues*. How do we prove it? Let $Ax = \lambda x$. Then $Ax = MBM^{-1}x = \lambda x \iff BM^{-1}x = \lambda M^{-1}x$. Now, $M^{-1}x$ is an eigenvector of B with the same eigenvalue.

So this measure divides matrices into groups with identical set of eigenvalues. The most "preferable" of them, obviously, are diagonal matrices. They have trivial eigenvectors.

7.9.2 Repeated Eigenvalues And Jordan Form

What if two or more eigenvalues are the same? Then the matrix might not be diagonalizable - it might not have a full set of independent eigenvectors. What happens now? Using above definition of "similar" matrices, there can be two "families" of matrices with same set of eigenvalues, but not similar to each

other. One is the diagonal matrix, in its own family. The other family contains all the other matrices with given eigenvalues.

In the latter family, the matrix that is as close as possible to the diagonal matrix, is called the Jordan Form for the family of matrices.

7.9.3 Updated Definition

Looking at some examples reveals that similar matrices actually have the same number of eigenvalues as well. So when eigenvalues are repeated, diagonal matrices have n eigenvectors - other matrices don't.

7.9.4 Jordan Theorem

Every square matrix A is similar to a Jordan matrix J which is composed of Jordan blocks in the main diagonal. For an eigenvalue λ_i repeated r times, the Jordan block is a $r \times r$ matrix with λ_i on the diagonals, and 1s on the above diagonal with rest of the entries 0.

7.10 Singular Value Decomposition

7.10.1 Introduction

We can factorize symmetric matrices as $A = Q\Lambda Q^T$. We can also factor a general matrix A as $A = S\Lambda S^{-1}$ but this can't be done in some cases, when S is not invertible. SVD is a generalized diagonalization approach which defines a decomposition

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

for any matrix A with a diagonal Σ and orthonormal U and V. In other words, we can say:

$$A = U\Sigma V^T = U\Sigma V^{-1} \iff AV = U\Sigma$$

where V is a orthnormal basis for the row space, U is an orthonormal basis for the column space. Hmm. Do the dimensions even match?

7.10.2 Calculation

How can we find V? Let's think of A^TA.

$$A^TA = V\Sigma^TU^TU\Sigma V^T = V\Sigma^2V^T$$

Yes! We can factorize A^TA (which is posdef) to find its eigenvalues and orthnormal eigenvectors. The eigenvectors will be V, and we can construct Σ by taking the square root of eigenvalues of A^TA (which is possible, since A^TA is posdef).

Similarly, looking at AA^T will let us find the Us.

Of course, our assertion (that U and V are basis of column and row spaces, respectively, and why the eigenvalues of A^TA and AA^T are the same) needs to be proven. (In the lecture, the latter fact is stated as obvious.. where did I miss it?)

8 Linear Transformation

8.1 Linear Transformation and Matrices

A linear transformation is a function T with the following properties:

- $T(u+v) = T(u) + T(v)$
- $T(cu) = cT(u)$

Of course, all linear transformations have an unique matrix representation, and vice versa. How can we find a matrix given a linear transform?

Let's say we have a linear transform T: $\mathbf{R}^n \to \mathbf{R}^m$. We have a basis for the input space $v_1, \cdots, v_n$ and a basis for the output space $w_1, \cdots, w_m$. How do we find the matrix for this transformation? Transform the first basis v_1 to get: $T(v_1) = c_1 w_1 + c_2 w_2 + \cdots + c_m w_m$. And those coefficients $c_1 \cdots c_m$ take up the first column of the matrix.

8.2 Change of Basis

Image compression: say we are compressing a grayscale image. Using standard basis to represent this image does not exploit the fact that neighboring pixels tend to have similar luminousity value. Are there other bases that can give us good sparse approximations?

JPEG uses the Fourier basis. It divides images into 8×8 blocks and changes the basis to Fourier basis, and we can threshold the coefficients. Hmm, actually, I want to try this out. Also, wavelets are a popular choice recently.

8.2.1 Same Linear Transformation from Different Bases

Say we have a linear transformation T. If we put vectors in coordinate system defined by a set of basis $v_1 \cdots v_n$, we will get a matrix for the transformation. Let's call it A. If we use another set of basis $w_1 \cdots w_n$, we get another matrix which we call B.

A and B are similar as in 7.9!

8.2.2 Eigenvector basis

Say we have an eigenvector basis: $T(v_i) = \lambda_i v_i$. Then, the change of basis matrix is diagonal!

8.3 Left, right, pseudoinverse

Let's generalize the idea of an inverse. A 2-sided inverse is the usual inverse:

$$AA^{-1} = I = A^{-1}A$$

A matrix A has a 2-sided inverse when it's square, and has full rank: $n = m = r$.

8.3.1 Left and Right Inverses

Suppose a full column-rank matrix $n = r$. The null space is just $\{0\}$, and for any $Ax = b$ there is either 0 or 1 solution. Since A is full column rank, A^TA is full rank and invertible. Note that we have a left inverse here: $\left(A^TA\right)^{-1}A^T$. If we multiply this inverse and A we get:

$$\left(A^TA\right)^{-1}A^TA = \left(A^TA\right)^{-1}\left(A^TA\right) = I$$

Resulting I is a $n \times n$ matrix.

Similarly, suppose a full row-rank matrix ($m = r$). The null space has dimension $n - r$, so for every $Ax = b$ there are infinitely many solutions. Now we have a right inverse: $A^T\left(AA^T\right)^{-1}$.

8.3.2 Left Inverse and Project

What do we get if we multiply the left inverse on the right?

$$A\left(A^TA\right)^{-1}A^T$$

This is a projection matrix! Similarly, if we multiply the right inverse on the left, we will get a projection matrix that projects onto its row space.

8.3.3 Pseudoinverse

What is the closest thing to an inverse when A is neither full row or column rank? Think of this: take any vector in the row space, and transform it by A, you'll get something from a column space. (In this regard, A is a projection onto the column space.) The two spaces (row and column) have the same dimension, so they should have a 1:1 (or invertible) relationship - so the matrix that undoes this transformation is called the pseudoinverse, A^+.

8.3.4 Pseudoinverse from SVD

How do we find a pseudoinverse? Maybe we can start from SVD:

$$A = U\Sigma V^{-1}$$

where Σ is a (sort-of) diagonal matrix sized $m \times n$ with singular values on the diagonal. What is its pseudoinverse? Well that's easy - we can get a $n \times m$ matrix and put inverse singular values on the diagonal. Then what happens in $\Sigma\Sigma^+$? We get a $m \times m$ matrix with first r diagonal entries 1. $\Sigma^+\Sigma$ works similarly.

How do we go from Σ^+ to A^+? Because U and V are orthonormal, we can just use:

$$A^+ = V\Sigma^+U^T$$