GTDE LinAlg Study Guide

Edwin Trejo Balderas

December, 2022

This living document is a study guide for all the content I learned in Georgia Tech course MATH 1554. It aims to not only present theorems but include their proofs and some major implications, as well as further notes clarifying practice exam questions where necessary.

1 Module 1 - Linear Equations

1.1 Systems of linear equations

- Linear eq - $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$
  - has coeffs (a, b), vars (x), and dimension (n)
- Linear system - more than one lin-eq
- Solution set - all possible values of $x_1, x_2, \ldots, x_n$ that satisfy the system
- Solution - a point in the solution set
- Solution sets can be empty, have one element, or be infinitely large
- Row operations -
  - Addition
  - Interchange
  - Scale
- Systems can be written as an augmented matrix -

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \ldots & \vdots \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} & b_m \end{pmatrix}$$

- Consistent - describes a system w/ at least one solution
- Row Equivalent - Describes matrices that can be transformed into each other via row operations
- Row equivalence $\Leftrightarrow$ same solution set

1.2 Row Reduction and Echelon Form

- Requirements for Echelon Form
  - All zero rows at bottom
  - Leading entry of a row is to the right of all leading entries above
  - All entries below a leading entry are zero
- Requirements for Reduced Row Echelon Form
  - Those of echelon form
  - All leading entries = 1
  - leading entries are the only non-zero entry in their column
- Pivot position - position in a matrix that corresponds to a leading 1 in RREF
- Pivot column - contains a pivot position
- Row Reduction Alg -
  - Swap if necessary to make leftmost non-zero entry in first row
  - Scale row 1 so that leading entry is 1
  - Make appropriate row replacements to make all entries below leading entry 0
  - repeat 2 and 3 for all remaining rows
- Basic variable - variable corresponding to a pivot
- Free variable - variable that is not basic
- Existence + Uniqueness - Consistent iff last column of augmented matrix has no pivot
- Unique solution if no free var, infinitely many otherwise

1.3 Vector Equations

- linear combination - vector $\vec{y}$ is a lin comb of $v_1, v_2, \dots, v_p$ if can be written as

$$\vec{y} = c_1 v_1 + c_2 v_2 + \cdots + c_p v_p$$

  Where c_i are weights

- Span - all linear combinations of a set of vectors, including 0

1.4 The Matrix Equation

- Matrix times a vector - a linear combination of the matrix's columns with the vector as their weights
- $A\vec{x}$ - span of columns of A
- $A\vec{x} = \vec{b}$ only has solution if $\vec{b}$ is a lin comb of vec of A

1.5 Solution sets of lineat systems

- Systems of the form $A\vec{x} = 0$ are homogenous
- If $\vec{b} \neq 0$ then inhomogenous
- Trivial solution to a homogenous system - $\vec{0}$
- Big question is if homogenous system has non-trivial sol
- $A\vec{x} = \vec{0}$ has a nontrivial solution $\Leftrightarrow$ there is a free var $\Leftrightarrow$ A has a column with no pivot
- Parametric vector form - way of representing solution set with a vector equation that represents relationships between free and basic variables

1.6 Linear Independence

- Lin Indep iff

$$c_1 v_1 + c_2 v_2 + \cdots + ckv_k = \begin{pmatrix} v_1 & v_2 & \dots \substack{v \\ k} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_k \end{pmatrix} = 0$$

  has only trivial solution

- If a set of vectors contains $\vec{0}$ then they are linearly dependent
- If we have k vectors in $\mathbf{R}^n$ and $k > n$ then the vectors are linearly dependent
  - Every vector in $A = \begin{pmatrix} v_1 & v_2 & \dots & v_k \end{pmatrix}$ has to be pivotal for vectors to be linearly independent
  - But more vectors than elements $\rightarrow$ not all vectors can have a pivot position $\rightarrow$ there will be some free variables

1.7 Intro to Linear Transforms

- Matrix transformation - Let $A \in \mathbf{R}^{m \times n}$, then we define

$$T : \mathbf{R}^n \rightarrow \mathbf{R}^m, T(\vec{x}) = A\vec{x}$$

  - Domain of T is $\mathbf{R}^n$
  - Dodomain of T is $\mathbf{R}^m$
  - $T(\vec{x})$ is image of $\vec{x}$ under T
  - Range of T is all possible images $T(\vec{x})$

- Ways of representing $A\vec{x} = \vec{b}$
  - linear equations
  - augmented matrix
  - matrix eq
  - vector eq
  - linear transofrmation eq

- Principle of superposition states that if T is linear then

$$T(c_1 v_2 + c_2 v_2 + \cdots + c_k v_k) = c_1 T(v_1) + c_2 T(v_2) + \cdots + c_k T(v_k)$$

- Every matrix tranformation is linear

1.8 Linear Transforms

- Standard vectors in $\mathbf{R}^n$ are $\vec{e_1}, \vec{e_2}, \ldots, \vec{e_n}$ where

$$\vec{e_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \vec{e_2} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \ldots, \vec{e_n} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \cdots \\ 1 \end{pmatrix}$$

- For $A \in \mathbf{R}^{m \times n}$ with columns $(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n})$

$$A\vec{e_i} = \vec{v_i}$$

- For linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$, there is a unique matrix A where

$$T(\vec{x}) = A\vec{x}$$

  called the standard matrix
- Note that $A = \begin{pmatrix} T(\vec{e_1}) & T(\vec{e_2}) & \ldots & T(\vec{e_n}) \end{pmatrix}$
- Rotation Matrix - $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$
- Reflection through n-axis - negate all values except n
- Reflection through $x_2 = x_1$ - $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
- Horizontal stretching - $\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}$
- Vertical stretching - $\begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix}$
- If $T : \mathbf{R}^n \to \mathbf{R}^m$ for all $\vec{b} \in \mathbf{R}^m$ has a $\vec{x} \in \mathbf{R}^n$ then it is onto
- If a transformation is onto, it is always consistent
- Existence property, applies if every row is pivotal, and/or if A's columns span $\mathbf{R}^m$
- If $T : \mathbf{R}^n \to \mathbf{R}^m$ for all $\vec{b} \in \mathbf{R}^m$ has at most one $\vec{x} \in \mathbf{R}^n$ then it is one-to-one
- If a transformation is one-to-one, it always has one or no solutions, and has linearly independent columns
- Uniqueness property, applies if every column is pivotal

2 Module 2 - Matrix Algebra

2.1 Matrix Operations

- Definitions of addition and scalar multiplication to be added if demand arises
- For matrix multiplication, number of columns of right = number of rows of left
- Each column of product = (left)(column of right)
- $AB = (A\vec{b_1}, A\vec{b_2}, \ldots, A\vec{b_p})$
- Matrix multiplication is:
  - Associative
  - Distributive
  - Has an Identity
  - NOT commutative
  - NOT cancellable
  - HAS 0 divisors
  - Equivalent to a composition of linear transforms
- Transpose of a matrix - matrix A^T whose rows are the columns of A
- Properties
  - $(A^T)^T = A$
  - $(A + B)^T = A^T + B^T$
  - $(rA)^T = rA^T$
  - $(AB)^T = B^T A^T$
- Power of a matrix - $A^n = AA \ldots A$ n times

2.2 Inverse of a Matrix

- A matrix A is invertible if there is a matrix B such that
$$AB = BA = I$$
Then we denote B as A^{-1}
- Singular - describes a non-invertable matrix
- A matrix is invertable iff it is both onto and one-to-one
- Algorithm
  - Row reduce the augmented $(A|I_n)$ to RREF
  - If RREF has form $(I_n|B)$ then $A^{-1} = B$, if not, A is singular
  - Alternatively, solving $A\vec{x_i} = \vec{e_i}$ for all elementary vectors gives the columns of A^{-1}
- Inverse properties
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^T)^{-1} = (A^{-1})^T$
- Elementary matrix - one that differs from I_n by one row operation
- Each operation can be represented w/ an elementary matrix multiplication
- A is invertible iff it is row equivalent to the identity

2.3 Invertible Matrices

- Current equivalences in invertible matrix theorem
  - A is invertible
  - A is row equivalent to I_N
  - All columns are pivotal
  - $A\vec{x} = \vec{0}$ has only the trivial solution
  - The columns of A are linearly independent
  - $A\vec{x} = \vec{b}$ has a solution for all $\vec{b} \in \mathbf{R}^n$
  - Columns of A span $\mathbf{R}^n$
  - A has a left or right inverse
  - A^T is invertible

2.4 Partitioned Matrices

- Matrices can be partitioned into sub-matrices with varying dimensions
- Can be used to simplify multiplication if important matrices are found (e.g. $I_n, 0$)
- To find inverse of partitioned matrix, construct partitioned identity and form system of equations for each term in multiplication

2.5 LU Factorization

- LU Fact. - decomposes a matrix into a product
- Upper triangular matrix - One where everything below the main diagonal is 0 ($a_{i,j} = 0 for all i > j$)
- Lower triangular matrix - One where everything above the main diagonal is 0 ($a_{i,j} = 0 for all j > i$)
- If A can be row reduced to echelon form without row exchanges, then $A = LU$ where L is an $m \times m$ lower triangular with 1 in main diagonal, and U is an echelon form of A
- To solve $A\vec{x} = \vec{b}$ with $A = LU$
  - Find LU decomp of A
  - Set $U\vec{x} = \vec{y}$ and solve $L\vec{y} = \vec{b}$
  - Solve $U\vec{x} = \vec{y}$
- To compute $A = LU$
  - Note that if A is LU factorizable, it can be row reduced to U w/o interchanging rows, so
  $$E_p \dots E_1 A = U = L^{-1}A = U$$
  Where E_j is row operations that are NOT exchanging rows, so all are lower triangular and invertible
  - So
  $$L^{-1} = E_p \dots E_1 \Rightarrow L = E_1^{-1} \dots E_p^{-1}$$
  - SO the process is to row reduce A to U and reverse the order of applied row operations to find L

2.6 Leontif IO Model

- An economy with N sectors can have its output measured by $\vec{x} \in \mathbf{R}^n$
- Consumption Matrix - C, describes how units are consumed by sectors to produce output
- Method 1 of defining entries of C - sector i sends some output to sector j, call it $c_{i,j}x_i$
- Method 2 of defining entries of C - sector i needs some output of sector i, call it $c_{i,j}x_i$
- $C\vec{x} =$ units consumed
- $\vec{x} - C\vec{x} =$ units left after internal consumption
- There is also external demand, $\vec{d}$, we ask if there is an $\vec{x}$ that satisfies

$$\vec{x} - C\vec{x} = (I - C)\vec{x} = \vec{d}$$

  Which is the Leontif IO model

2.7 Computer Graphics

- Homogenous coordinates in $\mathbf{R}^3$ - Points x, y in $\mathbf{R}^2$ can be represented with (x, y, H) where $H \neq 0$ which lies on the plane in $\mathbf{R}^3$ above the base plane
- Often, we set $H = 1$
- Matreix eq for a translation $(x, y \to (x + h, y + k))$ -

$$\begin{pmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x + h \\ y + k \\ 1 \end{pmatrix}$$

- Multiple points are stored as multiples columns in the right matrix
- To rotate points around a specific coord,
  - Translate points to the origin
  - Rotate
  - Translate back to original position
- 4d homogenous coords can be used for translations of 3d points similar to above
- 3d rotations around an axis alter everything in the other two axes

2.8 Subspaces of $\mathbf{R}^n$

- Subspace - a subset that is closed under scalar multiplication and vector addition, and HAS THE 0 VECTOR
- ex - is $\{\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbf{R}^2 | ab = 0\}$ a subspace? No, since

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

  Which is not in the space

- Column space of A - span of columns of A
- Null space of A - Span of solution set of $A\vec{x} = \vec{0}$
- For "find a matrix whose column space is spannec by $\vec{a}$ and null space is spanned by $\vec{b}$" problems, set one column to $\vec{a}$ and solve $(\vec{a}\vec{x})\vec{b} = \vec{0}$
- Basis of a subspace - a set of linearly independent vectors in the subspace that span it
- To construct a basis for null space, use parametric vector form on RREF
- To construct basis for col space, use pivotal columns of RREF

2.9 Dimension and Rank

- There are multiple choices for the basis of a subspace
- If **B** is a basis for a subspace H and $\vec{x}$ is in H, the coordinates of $\vec{x}$ relative to **B** are the weights such that
$$\vec{x} = c_1 b_1 + c_2 b_2 + \cdots + c_p b_p$$
and
$$\left[x\right]_{\mathbf{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}$$
Is the coordinate vector of $\vec{x}$
- To find $\left[x\right]_{\mathbf{B}}$, solve $B|\vec{x}$ where B's columns are the basis vectors
- Dimension of a subspace is the number of vectors needed to make a basis
- 0 CANNOT BE A BASIS VECTOR
- dim(Null A) is number of free var
- dim(Col A) is number of basic var / pivot cols
- If H is a p-dimensional subspace, any set of p linearly indep vectors in H is automatically a basis for H
- Rank - dimension of the column space of a Matrix = NUMBER OF PIVOTS
- if A has n columns, Rank$A$+dim(null$A$)=n
- For A to be invertible, dim(Null$A$) MUST BE 0

3 Module 3 - Determinants and Eigenvalues

3.1 Intro to Determinants

- DEFINITION -
  - if A is 1×1 and $A = [a_{1,1}]$then det$A$=$a_{1,1}$
  - for A being $n \times n$ where $n > 1$,
  $$\text{det}A = a_{1,1}\text{det}A_{1,1} - a_{1,2}\text{det}A_{1,2} + \cdots + (-1)^{1+n}a_{1,n}\text{det}A_{1,n}$$
  Where$A_{i,j}$ is A after eliminating row i and col j
- Cofactor - (i, j) cofactor of A is
$$C_{i,j} = (-1)^{i+j}\text{det}A_{i,j}$$
- Determinants can be taken down any row or column
- If A is triangular, then det$A = a_{1,1}a_{2,2} \dots a_{n,n}$

3.2 Properties of Determinants

- For a square matrix A:
  - If a multiple of a row is added to another row to obtain B, then $\det A = \det B$
  - If two rows are swapped to produce B then $\det A = -\det B$
  - If a row is scaled, then $k \det A = \det B$
  - If A is reduced to echelong form by r interchanges of rows, then $|A| = (-1)^r$ if invertible and 0 if singular
  - $|A| = |A|^T$
  - A is invertible iff $|A| \neq 0$
  - $\det AB = \det A \cdot \det B$
  - $\det A^{-1} = \frac{1}{\det A}$

3.3 Volume and Linear Transformations

- $\det A$ of a 2×2 matrix is the area of the parallelogram that the transformed unit vectors creates
- similarly, determinant of a 3×3 is the volume of the parallelopiped they form
- The area/volume of a shape after a linear transformation is the initial area/volume * determinant of the transformation

3.4 Markov Chains

- Probability vector - one with only non-negative elements that sum to 1
- Stochastic Matrix - square matrix whose columns are probability vectors
- Markov chain - a sequence of probability vectors and a stochastic matrix such that

$$\vec{x_{k+1}} = P\vec{x_k}, k = 0, 1, 2, \ldots$$

- Steady state vector - for a stochastic matrix, is one such that

$$P\vec{q} = \vec{q}$$

- To find the steady-state vector, note that

$$P\vec{q} - I\vec{q} = (P - I)\vec{q} = 0$$

  Solve the system, and scale the solution to make $||\vec{q}|| = 1$

- A stochastic matrix is regular if there is some k such that P^k has only positive entries
- If P is regular stochastic, then it has a unique steady-state vector $\vec{q}$ and $\vec{x_{k+1}} = P\vec{x_k}$ converges to $\vec{q}$ as $k \to \infty$
- In other words, in a long time

3.5 Eigenvalues and Eigenvectors

- If A is square and there is some $\vec{v}$ where

$$A\vec{v} = \lambda\vec{x}$$

  Then $\vec{v}$ is an eigenvector of A and λ is the corresponding eigenvalue

- Even if all entries in A are real, λ can be complex (for rotations)
- If λ is real and negative, then $A\vec{v}$ and $\vec{v}$ are parallel but point in the same direction
- The λ-eigenspace for A is the subspace spanned by a given λ
- The eigenspace for A is $\mathrm{Nul}(A - \lambda I)$
- The diagonal elements of a triangulat matrix are its eigenvalues
- If A is singular $\leftrightarrow$ 0 is an eigenvalue of A
- Stochastic matrices have at least one eigenvalue = 1
- If $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ are eigenvectors associated with distinct eigenvalues, then they are linearly independent
- ROW OPERATIONS CHANGE EIGENVALUES $\to$ REDUCED MATRICES CANNOT HAVE THEIR E-VAL DETERMINED

3.6 The Characteristic Equation

- Characteristic polynomial of A -

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

- Characteristic equation -

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

- Roots of characteristic polynomial are the eigenvalues of A
- Trace - sum of diagonal elements of a matrix
- Algebraic multiplicity of an eigenvalue - a_i, multiplicity as the root of the characteristic polynomial
- Geometric multiplicity of an eigenvalue - g_i, $dim(Nul(A - \lambda I))$ Always at least 1, but can be more than 1
- Obey the following - $1 \leq a_i \leq n$, $1 \leq g_i \leq a_i$
- For a regular stochastic matrix, the eigenvector associated with the eigenvalue 1 is the steady state, all others are less than one and converge to 0
- Similar Matrices - A and B such that there is some P where

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

- If A and B are similar then they have the same characteristic polynomial
- HOWEVER, if two have different eigenvalues, they do not necessarliy have to be similar
-

3.7 Diagonalization

- Diagonal matrix - one where non-zero elements are only on the main diagonal
- If A is diagonal, then powers are easy, just raise elements in the main diagonal to the power applied to the matrix
- A is diagonalizable if it is similar to the diagonal matrix D, or

$$A = PDP^{-1} = \begin{pmatrix} \vec{v_1} & \vec{v_2} & \dots & \vec{v_n} \end{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \vdots & \\ & & & \lambda_n \end{pmatrix} \begin{pmatrix} \vec{v_1} & \vec{v_2} & \dots & \vec{v_n} \end{pmatrix}^{-1}$$

- A is diagonalizable $\leftrightarrow$ $A has n$ linearly independent eigenvectors
- If A has n distinct eigenvalues, it is diagonalizable
- SINGULARITY DOES NOT IMPLY A MATRIX IS NOT DIAGONALIZABLE
- A is diagonalizable iff $g_i = a_i$ for all i
- If there are repeated eigenvalues, construct a_i eigenspace basis vectors for each eigenvalue so that $g_1 = a_i$

3.8 Complex Eigenvalues

- If λ is the complex root of a polynomial, its conjugate is also a root (more elaboration on complex number theory is not included due to the author's familirity with the subject)
- A matrix of the form $C = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ will have eigenvalues $\lambda = a \pm ib$ whose argument is the degrees of rotation the marix performs
- A matrix of the form above is called a dilation-rotation because it dilates by $r = \sqrt{a^2 + b^2}$ and rotates by $\phi = \tan^{-1} \frac{b}{a}$
- If A is a real 2×2 matrix with eigenvalue $\lambda = a - bi, b \neq 0$ and associated eigenvector $\vec{v}$, we can decompose it into

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

  where

  $$P = \begin{pmatrix} Re\vec{v} & Im\vec{v} \end{pmatrix}, C = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

3.9 Google Page Rank

- For a $m \times m$ regular stochastic matrix, P where $m \geq 2$, the following is true

  $$\lim_{n \to \infty} P^n \vec{x_0} = \vec{q}$$

  Where $\vec{x_0}$ is any probability vector and $\vec{q}$ is the steady state (unique eigenvalue of P with eigenvalue $\lambda = 1$, the rest are less than 1) of P
- Also, there is a stochastic matrix Π where

  $$\lim_{n \to \infty} P^n = \Pi$$

  And all columns of Π are $\vec{q}$
- Google's page rank make the following assumptions
  - A user is equally likely to go to any webpage the current one links to
  - A user on a linkless page will stay on that page
  - The distribution of users can be modeled with a Markov process with matrix $n \times n$ where n is number of pages on web
- The matrix is called a transition matrix and its steady state contains the **importance** of each page and it *can* characterize the long-term behavior of the web
- A problem is that the transition matrix is not necessarily regular so we do not have a guaranteed steady state
- Another is that linkless pages end up with the largest importance
- Adjustment 1 - users on a linkless page will choose any page on the web with equal probability and move to that, new transition matrix is P_*

- Adjustment 2 - users will move to any page their page links to with probability p and any other page on the web with probability $1-p$ so the transition matrix becomes

$$G = pP_* + (1-p)K$$

where all elements in K are $\frac{1}{n}$

- p is called the damping factor
- This forces G to be regular stochastic when $0 \leq p < 1$
- Google uses a damping factor of 0.85
- Then, we $\lim_{n\to\infty} G^n \vec{x_0}$ converges to a single steady state vector $\vec{q}$

4 Module 4 - Orthogonality

4.1 Inner Product, Length, and Orthogonality

- Dot product - $\vec{v} \cdot \vec{u} = \vec{v}^T \vec{u} = v_1u_1 + v_2u_2 + \cdots + v_nu_n$
- $(\vec{v} + \vec{w}) \cdot \vec{u} = \vec{v} \cdot \vec{u} + \vec{w} \cdot \vec{u}$
- $\vec{u} \cdot \vec{u} \geq 0$
- $\vec{u} \cdot \vec{u} = 0 \Rightarrow \vec{u} = 0$
- Length - $||\vec{u}|| = \sqrt{\vec{u} \cdot \vec{u}}$
- LENGTH IS ALWAYS POSITIVE
- If a vector has length 1, it is a unit vector
- Distance between 2 vectors is $||\vec{u} - \vec{v}||$
- $\vec{a} \cdot \vec{b} = ||\vec{a}||||\vec{b}|| \cos\theta$
- Orthogonal - describes two vectors if $\vec{a} \cdot \vec{b} = 0$ or $||\vec{a} + \vec{b}||^2 = ||\vec{a}||^2 + ||\vec{b}||^2$
- Note similarity of the above to the pythagorean theorem
- Yes $\vec{0}$ is orthogonal to everything but we usually mean non-zero vectors when talking about orthogonality
- The span of vectors orthogonal to another is a subspace
- Orthogonal compliment - of a subspace, the set of all vectors orthogonal to the subspace W, and is itselt a subspace $W^\perp$
- $RankW + RankW^\perp = n$
- $RowA$ - space spanned by rows of A
- Basis for $RowA$ is the pivot columns of A
- $dim(RowA) = dim(ColA)$
- $RowA = ColA^T$
- In general, $RowA, ColA$ are not related
- $NulA = (RowA)^\perp$
- $NulA^T = (ColA)^\perp$

4.2 Orthogonal Sets

- Orthogonal set - one where all vectors are ortho to each other
- If all vectors in an ortho set are non-zero then the set is also linearly independent
- If there is an orthogonal basis for a subspace W then any $\vec{w} \in W$ can be expressed as

$$\vec{w} = c_1\vec{u_1} + c_2\vec{u_2} + \cdots + c_p\vec{u_p}$$

Where

$$c_q = \frac{\vec{w} \cdot \vec{u_q}}{\vec{u_q} \cdot \vec{u_q}}$$

- If all vectors in the basis have unit length, then

$$\vec{w} = (\vec{w} \cdot \vec{u_1})\vec{u_1} + \cdots + (\vec{w} \cdot \vec{u_p})\vec{u_p}$$

and

$$||\vec{w}|| = \sqrt{(\vec{w} \cdot \vec{u_1})^2 + \cdots + (\vec{w} \cdot \vec{u_p})^2}$$

- The closest vector to $\vec{y}$ in the span of $\vec{u}$ is

$$\hat{y} = \frac{\vec{y} \cdot \vec{u}}{\vec{u} \cdot \vec{u}} \vec{u}$$

- The projection of a vector into the span of another is

$$proj_{\vec{u}} \vec{y} = \frac{\vec{y} \cdot \vec{u}}{\vec{u} \cdot \vec{u}} \vec{u}$$

- A vector can be decomposed into $\vec{z} = \vec{y} - \hat{y}$ is orthogonal to the subspace that $\vec{y}$ was projected to
- The projection of the basis of a subspace to the orthogonal compliment of that subspace is $\vec{0}$
- U has orthonormal columns $\Rightarrow U^T U = I_n$
- If U has orthonormal columns, then
  - $||U\vec{x}|| = ||\vec{x}||$
  - $(U\vec{x}) \cdot (U\vec{y}) = \vec{x} \cdot \vec{y}$
- Thus, U preserves lengths and orthogonality
- Orthogonal matrix - a square one whose columns are orthonormal
- If U is orthogonal, then $U^T = U^{-1}$, its columns are linearly independnt, and its determinant is 1 or -1

4.3 Orthogonal Projections

- If W is a subspace and $\hat{y}$ is the projection of $\vec{y}$ onto W, then for any $\vec{v} \in W$ we have

$$||\vec{y} - \hat{y}|| < ||\vec{y} - \vec{v}||$$

  AKA, $\hat{y}$ is the vector in W closest to $\vec{y}$
- Thus, the distance between $\vec{y}$ and a subspace is $||\vec{y} - \hat{y}||$
- Ortho decomp theorem - Let W be a subspace with basis $\vec{u_1}, \vec{u_2}, \ldots, \vec{u_p}$ then all $\vec{y}$ have the unique decomposition

$$\vec{y} = \hat{y} + z$$

  where

$$\hat{y} = \frac{\vec{y} \cdot \vec{u_1}}{\vec{u_1} \cdot \vec{u_1}} \vec{u_1} + \cdots + \frac{\vec{y} \cdot \vec{u_1}}{\vec{u_p} \cdot \vec{u_p}} \vec{u_p}$$

  We call $\hat{y}$ the orthogonal projection of $\vec{y}$ onto W

4.4 Graham Schmidt

- For a set of vectors, $\vec{x_1}, \vec{x_2}$, we can construct an ortho basis $\vec{v_1}, \vec{v_2}$ for the subspace they span with the following

$$\vec{v_1} = \vec{x_1}$$
$$\vec{v_2} = \vec{x_2} - \frac{\vec{x_2} \cdot \vec{v_1}}{\vec{v_1} \cdot \vec{v_1}} \vec{v_1}$$

- For more basis vectors $\vec{x_1}, \ldots, \vec{x_p}$ we can do something similar

$$\vec{v_1} = \vec{x_1}, W_1 = Span\vec{v_1}$$
$$\vec{v_2} = \vec{x_2} - proj_{W_1}\vec{x_2}, W_2 = Span\vec{v_1}, \vec{v_2}$$
$$\vec{v_3} = \vec{x_3} - proj_{W_2}\vec{x_3}, W_3 = Span\vec{v_1}, \vec{v_2}, \vec{v_2}$$
$$\vdots$$

- Orthonormal basis - a set of mutually orthogonal and normal vectors
- Any matrix A w/ linearly independent columns has the QR-factorization

$$A = QR$$

  where Q is $m \times n$ and its columns are an orthonormal basis for $ColA$ and R is $n \times n$, upper triangular, with positive entries in its diagonal
- For QR, we do not consider matrices with linearly dependent columns normally
- Q can be constructed with graham schmidt
- R can be found with $R = Q^T A$

4.5 Least Squares Problems

- To make a $A\vec{x} = \vec{b}$ for a set of univariate data, make

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \end{pmatrix} \begin{pmatrix} x \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \end{pmatrix}$$

- We use least squares to find the best approximate line
- A least squares solutino to $A\vec{x} = \vec{b}$ is $\hat{x}$ where

$$||\vec{b} - A\hat{x}|| \leq ||\vec{b} - A\vec{x}||$$

  for all $\vec{x}$
- The normal equations - $A^T A\hat{x} = A^T\vec{b}$, coincides with least squares solutions to $A\vec{x} = \vec{b}$
- If A's columns are linearly independent, then $A^T A$ is invertible and the least squares solution to $A\vec{x} = \vec{b}$ is

$$\hat{x} = (A^T A)^{-1} A^T\vec{b}$$

- Factorizing $A = QR$, the least squares solution can be found with $R\hat{x} = Q^T\vec{b}$

4.6 Applications to Linear Models

- With the normal equations we can find the $\vec{x}$ that minimized $||A\vec{x} = \vec{y}||$
- The residual vector is defined as $\vec{r} = A\vec{x} - \vec{y}$ and $||A\vec{x} = \vec{y}||^2 = ||\vec{r}||^2$ where entries of $\vec{r}$ are called residuals
- Mean deviation form - converting all data to $x_* = x - \bar{x}$
- By using mean deviation form, columns of A are orthogonal, so $A^T A$ is diagonal, so normal equations are much easier to use
- We can also use least squares for non-linear data, using known functions of the form

$$y = c_0 + c_1 f_1(x) + c_2 f_2(x) + \cdots + c_k f_k(x)$$

- Example is a polynomial of the form $y = c_1 x + c_2 x^2 + c_3 x^3$
- Also, we could do multivariable functions

$$y = c_0 + c_1 f_1(x_1, \ldots, x_p) + c_2 f_2(x_1, \ldots, x_p) + \cdots + c_k f_k(x_1, \ldots, x_p)$$

- Example is the function $z = c_0 + c_1 x + c_2 y$

5 Module 5 - Symmetric Matrices and SVD

5.1 Diagonalization of Symmetric Matrices

- Symmetric matrix - one that satisfies $A = A^T$
- Powers of a symmetric matrix are symmetric
- $(AA^T) = (A^T)^T A^T = AA^T$ is symmetric
- Symmetric matrices must be square
- Square and diagonal matrices are symmetric
- Symmetric matrices have all-real eigenvalues
- If A is symmetric with two distinct eigenvectors with distinct eigenvalues, then the eigenvectors are orthogonal
- Generally, eigenspaces associated with distinct eigenvalues of a symmetic matrix are orthogonal subspaces
- Symmetric matrices can be diagonalized into $A = PDP^T$ since P would be orthogonal given the eigenvectors of A and $P^T = P^{-1}$
- Gram-Schmidt might be needed to construct a fully orthogonal P if eigenvalues are repeated
- Furthermore, if A can be decomposed into $A = PDP^T$ then it is symmetric
- The spectrum of a matrix is the full set of its eigenvalues
- Spectral decomposition of a matrix - if A can be orthogonally diagonalized intro $A = PDP^T$ then it has the spectral decomposition

$$A = \lambda_1 \vec{u_1}\vec{u_1}^T + \cdots + \lambda_n \vec{u_n}\vec{u_n}^T$$

- Ordering eigenvalues from largest to smallest allows us to approximate the matrix itself by choosing how many terms to include in its decomposition

5.2 Quadratic Forms

- Quadratic form - function $Q : \mathbf{R}^n \to \mathbf{R}$ given by

$$Q(\vec{x}) = \vec{x}^T A\vec{x} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{1,2} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \dots & a_{n,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

  Where $\vec{x}$ is variables

- Cross product term - one with multiple variables
- The highest degree of any term is 2
- Coefficients of one-var terms are on the main diagonal and those of cross product terms are on the triangles
- The squared length of a linear transform is a quadratic form

$$||A\vec{x}||^2 = (A\vec{x}) \cdot (A\vec{x}) = \vec{x}^T A^T A\vec{x}$$

- Change of variable is done to remove cross product terms
- Since A is symmetric, it can be decomposed into $A = PDP^T$
- The orthogonal change of variable is $\vec{x} = P\vec{y} \Rightarrow \vec{y} = P^{-1}\vec{x}$
- Thus $\vec{x}^T A\vec{x} = \vec{y}^T D\vec{y} = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2$
- If we set Q to some constraint, we can use change of var to more easily see max and min points, as well as if Q can be positive or negative
- The quadratic form from a 2×2 matrix will give a surface in 3D space
- If Q is _ it is _
  - $Q > 0$, positive definite
  - $Q < 0$, negative definite
  - $Q \geq 0$, positive semidefinite
  - $Q \leq 0$, negative semidefinite
  - Both negative and positive, indefinite
- Q is positive definite when all eigenvalues are positive, negative definite if all values are negative, and indefinite if it has at least one positive and negative eigenvalue

5.3 Quadratic Forms

- Surface of a sphere is given by $1 = x_1^2 + x_2^2 + x_3^2 = ||\vec{x}||^2$
- We may want to solve for the quantity at a point $Q(\vec{x}) = c_1 x_1^2 + c_2 x_2^2 + c_3 x_3^2$
- To find the largest and smallest values, not that

$$Q(\vec{x}) = c_1 x_1^2 + c_2 x_2^2 + c_3 x_3^2 \leq max c_1, c_2, c_3 (x_1^2 + x_2^2 + x_3^2) = max c_1, c_2, c_3 ||\vec{x}||^2$$

- For $Q = \vec{x}^T A\vec{x}$, the max is associated with the eigenvector of the largest eigenvalue of A is A is diagonal, or D from $A = PDP^{-1}$ if it is not
- Similar with the minimum value of Q
- The above is a constrained value problem
- When repeated eigenvalues occur, construct all eigenvectors needed to make an appropriate basis, and ensure other eigenvectors are orthogonal to all
- With repeated eigenvalues and thus multiple bases for an eigenspace, all max/mins are located at unit vectors in the SPAN of the eigenvectors, not just one of them
- If $\vec{x}$ is constrained to be orthogonal to the max/min vector, the new max/min is given by $\vec{u_2}/\vec{U_{n-1}}$

5.4 The SINGULAR VALUE DECOMPOSITION

- Singular value - of a real matrix A, are the square roots of the eigenvalues of A^TA, notated as $\sigma_1 = \sqrt{\lambda_1}$
- If we want to maximize $||A\vec{x}||$ subject to $||\vec{v}|| = 1$, we note that the maximum would occur at the same location where $||A\vec{v}||^2$ is maximized and that

$$||A\vec{v}||^2 = \vec{v}^TA^TA\vec{v}$$

  Therefore the max of $\vec{v}^TA^TA\vec{v}$ occurs at the eigenvectors of A^TA, so the max of $||A\vec{v}||$ also does, and this max is the square root cause unsquare
- The eigenvalues of A^TA are all non-negative
- The n orthogonal eigenvectors of A^TA are ordered such that their eigenvalues are $\lambda \geq \cdots \geq \lambda_n$ and there are r non-zero singular values of A the following is true:

$$\{\vec{v_{r+1}}, \ldots, \vec{v_n}\}$$

  form an orthogonal basis for nul A and

$$\{\vec{v_1}, \ldots, \vec{v_r}\}$$

  form an orthogonal basis for row$A$, and $rankA = r$
- ALSO, the ordered eigenvectors of A^TA form a set $A\vec{v_1}, \ldots, A\vec{v_r}$ that is an orthogonal basis for Col$A$
- Furthermore, we define $\vec{u_i} = \frac{1}{\sigma_i}A\vec{v_i}$
- The set $\{\vec{u_1}, \ldots, \vec{u_r}\}$ forms an orthogonal basis for Col A
- Left singular vector - $\vec{u_i}$
- Right singular vector - $\vec{v_i}$
- SVD - Suppose A is $m \times n$ with $\sigma_1 \geq \cdots \geq \sigma_n$, then it has the decomposition $A = U\Sigma V^T$ where

$$D = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n \end{pmatrix}$$

  - $\Sigma = \begin{pmatrix} D \\ 0_{m-n,n} \end{pmatrix}$ if $m \geq n$
  - $\Sigma = \begin{pmatrix} D & 0_{m,n-m} \end{pmatrix}$ if $m < n$
- In the SVD V is orthogonal and $n \times n$ and its columns are the eigenvectors of A^TA with additional orthogonal columns built with gram-schmidt if necessary
- In the SVD U is orthogonal and $m \times m$ and its columns are $\vec{u_i} = \frac{1}{\sigma_i}A\vec{v_i}$ with additional orthogonal columns built with gram-schmidt if necessary

5.5 Applications of the SINGULAR VALUE DECOMPOSITION

- If A is $n \times n$ and intertible, then $\frac{\sigma_1}{\sigma_n}$ is the condition number of A
- Condition number tells how sensitive methods of solving $A\vec{x} = \vec{b}$ will be to errors in A, bigger condition number means more sensitivity
- SVD can be used to make a spectral decomposition of any matrix with rank r

$$A = \sum_{i=1}^{r} \sigma_i \vec{u}\vec{v}^T$$

- Each resulting term has a rank of 1, and this decomposition can be used to approximate A similar to a Taylor expansion
- SVD AND THE SUBSPACES - for a matrix A where
  - $\vec{v_i}$ are the eigenvectors of A^TA
  - $\vec{u_i}$ are the eigenvectors of AA^T
  - r=rank$A$

  The following is true
  - $\vec{v_1}, \ldots, \vec{v_r}$ form an orthonormal basis for Row$A$=Nul$A^\perp$
  - $\vec{v_{r+1}}, \ldots, \vec{v_n}$ form an orthonormal basis for Nul$A$=Row$A^\perp$
  - $\vec{u_1}, \ldots, \vec{u_r}$ form an orthonormal basis for Col$A$=Nul$(A^T)^\perp$
  - $\vec{u_{r+1}}, \ldots, \vec{u_m}$ form an orthonormal basis for NulA^T=Col$A^\perp$