Math 307 Study Guide

Alisa Liu

April 2018

Chapter 1: Linear equations

1.1 Solving linear equations

- Vector norms
  - p-norm
$$\|x\|_p = \left(|x_1|^p + ... + |x_n|^p\right)^{\frac{1}{p}}$$
  - 1-norm
$$\|x\|_1 = |x_1| + ... + |x_n|$$
  - 2-norm
$$\|x\|_2 = \sqrt{|x_1|^2 + ... + |x_n|^2}$$
  - ∞-norm
$$\|x\|_\infty = \max\{|x_1|, ..., |x_n|\}$$
- $\|\mathbf{x}\|$ is a norm iff the following properties hold
  - $\|\mathbf{x}\| \geq 0$
  - $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$
  - $\|s\mathbf{v}\| = |s| \cdot \|\mathbf{v}\|$
  - $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
- Matrix norms
  - Hilbert-Schmidt (Frobenius) norm
$$\|A\|_{HS} = \left(\sum_{j=1}^{n}\sum_{i=1}^{m}|a_{ij}|^2\right)^{\frac{1}{2}}$$
  - Matrix (operator) norm
$$\|A\| = \max_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2} = \max_{\|z\|=1}\|Az\|$$
    If A is diagonal, then $\|A\| = \max\{|\lambda_1|, ..., |\lambda_n|\}$
- Condition number
$$k(A) = \left\|A^{-1}\right\|\|A\| = \frac{\text{largest stretch factor}}{\text{smallest stretch factor}}$$
  If A is diagonal and invertible, then
$$k(A) = \frac{\max\{|\lambda_1|, ..., |\lambda_n|\}}{\min\{|\lambda_1|, ..., |\lambda_n|\}}$$

1.2 Interpolation

Given a set of data points, we want to fit a function exactly.

- Lagrange interpolation: Fit to $p(x) = a_1x^{n-1} + ... + a_{n-1}x + a_n$.

$$\begin{bmatrix} x_1^{n-1} & ... & x_1 & 1 \\ \vdots & & \vdots & \vdots \\ x_n^{n-1} & ... & x_n & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

- Cubic spline interpolation: Fit to piecewise function

$$f(x) = \begin{cases} p_1(x) & x_1 \le x \le x_2 \\ \vdots & \\ p_{n-1}(x) & x_{n-1} \le x \le x_n \end{cases}$$

$$p_j(x) = A_jx^3 + B_jx^2 + C_jx + D_j$$

We impose conditions so that $f(x)$ is continuous at all points and smooth.

1.3 Finite difference approximations

Given a differential equation $f''(x) + q(x)f(x) = r(x)$ and initial conditions $f(x_0) = A, f(x_N) = B$, we want to find an $N + 1$-point approximation of $f(x)$.

$$\left(\begin{bmatrix} 1 & 0 & 0 & 0 & ... & 0 & 0 \\ 1 & -2 & 1 & 0 & ... & 0 & 0 \\ 0 & 1 & -2 & 1 & ... & 0 & 0 \\ 0 & 0 & 1 & -2 & ... & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & ... & -2 & 1 \\ 0 & 0 & 0 & 0 & ... & 0 & 1 \end{bmatrix} + (\Delta x)^2 \begin{bmatrix} 0 & 0 & 0 & ... & 0 & 0 \\ 0 & q_1 & 0 & ... & 0 & 0 \\ 0 & 0 & q_2 & ... & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & ... & q_{N-1} & 0 \\ 0 & 0 & 0 & ... & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \\ f_N \end{bmatrix} = \begin{bmatrix} A \\ (\Delta x)^2 r_1 \\ (\Delta x)^2 r_2 \\ \vdots \\ (\Delta x)^2 r_{N-1} \\ B \end{bmatrix}$$

We can modify the first and last rows of L in the case that we have different boundary values.

Chapter 2: Subspaces, Bases, and Dimension

2.1 Vector spaces and subspaces

- If V is a vector space, then $S \subseteq V$ is a subspace of V if $\forall u, v \in S$ and a, b scalars, $au + bv \in S$.

2.2 The four fundamental subspaces

- The four subspaces
  - $N(A) \subseteq R^n$
    - Solve the homogenous linear system $Ax = 0$. This is the same as solving $Ux = 0$, where $U = \text{rref}(A)$.
  - $R(A) \subseteq R^m$
    - Find the linearly independent columns of A, which correspond to pivot columns of U.
    - Use $R(A) = R(W^T)$, where $W = \text{rref}(A^T)$

- $N(A^T) \subseteq R^m$
  - Solve the homogenous linear system $A^T x = 0$. This is the same as solving $Wx = 0$.
- $R(A^T) \subseteq R^n$
  - Find the linearly independent columns of A^T, which correspond to linearly independent columns of W
  - Use $R(A^T) = R(U^T)$

- Dimensions
  - $\text{rank}(A^T) = \text{rank}(A)$
  - $\text{rank}(A) + \dim(N(A)) = n$
  - $\text{rank}(A^T) + \dim(N(A^T)) = m$

- Properties of the inner product
  - $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$
  - $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
  - $\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a \langle \mathbf{x}, \mathbf{z} \rangle + b \langle \mathbf{y}, \mathbf{z} \rangle$
  - $\langle \mathbf{x}, \mathbf{x} \rangle = \|x\|_2^2$

- Two subspaces are orthogonal iff their basis vectors are pairwise orthogonal. If B is a basis for S_1 and C is a basis for S_2, then $B^T C = 0$.

- Relation among $N(A), N(A^T), R(A), R(A^T)$
  - $N(A) = [R(A^T)]^\perp \quad N(A^T) = [R(A)]^\perp$
  - In order to determine whether $\mathbf{x}$ is in $R(A)$, we can check if $\mathbf{x} \perp N(A^T)$

2.3 Graphs and networks

- D is the $m \times n$ incidence matrix where $m = \#$ edges, $n = \#$ nodes

- $\mathbf{v} \in R^n$ assigns the voltage to each node. $D\mathbf{v} \in R^m$ assigns the voltage difference across each edge. $\mathbf{y} \in R^m$ assigns a current to each edge. $D^T \mathbf{y} \in R^n$ assigns the current accumulation at each node.

- The four subspaces
  - $N(D)$ is the set of voltages on nodes such that the voltage difference across each edge is 0. Since the voltages of any connected nodes have to be identical, $N(D) = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ for any connected graph.

$$\dim(N(D)) = \#\text{ connected components}$$

  - $R(D)$ is the set of possible voltage differences across edges. Since the sum of differences around a closed loop is 0, we have one constraining equation for every linearly independent loop in the graph.

$$\dim(R(D)) = m - \#\text{ independent loops}$$

  - $N(D^T)$ is the set of edge currents that do not result in any current building up at any nodes. We can find a basis by using the fact that the current around a loop will not build up at any node.

$$\dim(N(D^T)) = \#\text{ independent loops}$$

- $R(D^T)$ is the set of possible current accumulations at each node. Since the current building up at one node must have come from other nodes, in a connected graph, we only have one constraining equation.

$$\dim(R(D^T)) = n - \#\text{ connected components}$$

- The Laplacian: $L = D^T R^{-1} D$
  - L is symmetrical
  - If all resistors are 1Ω, then

$$L_{ii} = \text{ of neighbors of node } i$$

$$L_{ij} = \begin{cases} -1 & \text{if vertex } i,j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$

  - More generally,

$$L_{ii} = \sum_j \frac{1}{R_j} \quad \text{sum over all edges connecting to node } i$$

$$L_{ij} = \begin{cases} -\frac{1}{R_k} & \text{where } R_k \text{ is the resistance between nodes } i,j \\ 0 & \text{otherwise} \end{cases}$$

  - $N(L) = N(D)$
  - Kirchoff's law: $L\mathbf{v} = 0$ when there are no batteries involved.

- When we connect a battery, we have the system

$$L\mathbf{v} = \mathbf{z}$$

where $\mathbf{z}$ represents the current accumulating at each node.

$$\begin{bmatrix} A & B^T \\ B & C \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{v} \end{bmatrix} \begin{bmatrix} \mathbf{J} \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{J} = (A - B^T C^{-1} B)\mathbf{b}$$

Since $\mathbf{J}$ gives the current vector and we know the voltage of the battery, we can solve

$$R_{eq} = \frac{\text{current}}{\text{voltage}}$$

Chapter 3: Orthogonality

3.1 Projections

- Projection of $\mathbf{y}$ onto L is the vector in L that is closest to $\mathbf{y}$

$$P_{\mathbf{x}}(\mathbf{y}) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|^2} \mathbf{x} = \langle \hat{\mathbf{x}}, \mathbf{y} \rangle \hat{\mathbf{x}}$$

$$R(P) = L \quad N(P) = L^{\perp}$$

- $Q = I - P$ projects onto $L^{\perp}$

- P is an orthogonal projection matrix if $P^2 = P$ and $P^T = P$

$$P_{\mathbf{x}} = \frac{\mathbf{x}\mathbf{x}^T}{\|\mathbf{x}\|^2}$$

$$P_{R(A)} = A(A^TA)^{-1}A^T = \tilde{A}(\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T$$

- Least squares equation to solve $A\mathbf{x} = \mathbf{b}$ when there is no solution

$$A^TA\mathbf{x}^* = A^T\mathbf{b}$$

  - A^TA is invertible iff A is invertible

$$\mathbf{x}_{LS} = (A^TA)^{-1}A^T\mathbf{b}$$

- Pseudoinverse of A: $(A^TA)^{-1}A^T$
- Polynomial regression $m < n$

$$p(x) = a_1x^{n-1} + ... + a_{n-1}x + a_n$$

$$V\mathbf{a} = \mathbf{y}$$

$$V = \begin{bmatrix} x_1^{m-1} & ... & x_1 & 1 \\ \vdots & & \vdots & \vdots \\ x_n^{m-1} & ... & x_n & 1 \end{bmatrix} \quad \mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\mathbf{a}_{LS} = (V^TV)^{-1}V^T\mathbf{y}$$

3.2 Complex vector spaces and inner product

- Complex number $z = a + bi$
  - Modulus: $|z| = \sqrt{a^2 + b^2}$
  - Complex conjugate: $\overline{z} = a - bi$
  - Argument: $\theta = \arctan\left(\frac{b}{a}\right)$
- Some formulas
  - $z = |z|(\cos\theta + i\sin\theta) = Re^{i\theta}$
  - $e^{i\theta} = \cos\theta + i\sin\theta$
- Complex inner product

$$\langle z, z\rangle = \overline{z_1} \cdot z_1 + ... + \overline{z_n} \cdot z_n$$

- Properties
  - $\langle s\mathbf{v}, \mathbf{w}\rangle = \overline{s}\,\langle \mathbf{v}, \mathbf{w}\rangle$
  - $\langle \mathbf{v}, s\mathbf{w}\rangle = s\,\langle \mathbf{v}, \mathbf{w}\rangle$
  - $\langle \mathbf{v}, \mathbf{w}\rangle = \overline{\langle \mathbf{w}, \mathbf{v}\rangle}$
  - $\langle \mathbf{v}, \mathbf{w}\rangle = \overline{\mathbf{v}}^T\mathbf{w}$
  - $\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, A^*\mathbf{w}\rangle$
- Adjoint of A: $A^* = \overline{A}^T$

3.3 Orthonormal bases, orthogonal & unitary matrices

- A basis $\{\mathbf{q}_1, ..., \mathbf{q}_n\}$ is an orthonormal basis for V if

$$\left\langle \mathbf{q}_i, \mathbf{q}_j \right\rangle = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

- For any $\mathbf{v} \in V$, there exists a unique set of scalars $c_1, ..., c_n$ such that

$$\mathbf{v} = c_1\mathbf{q}_1 + ... + c_n\mathbf{q}_n = \sum_{j=1}^{n} \left\langle \mathbf{q}_j, \mathbf{v} \right\rangle \mathbf{q}_j$$

- A square matrix Q whose columns form an ONB is called
  - Orthogonal if all entries are real
  - Unitary if entries are complex
- Properties
  - $QQ^* = Q^*Q = I$
  - Rows of Q is also an orthogonal/unitary matrix
  - $\|Q\mathbf{v}\|_2 = \|\mathbf{v}\|_2$

3.4 Fourier series

- 1-period signals are modeled by

$$L^2[0,1] = \left\{ f : [0,1] \to C : \int_0^1 |f(t)|^2 dt < \infty \right\}$$

- The inner product

$$\langle f, g \rangle = \int_0^1 \overline{f(t)} g(t) dx$$

$$\langle f, f \rangle = \int_0^1 |f(x)|^2 dx = \|f\|_2^2$$

- An orthonormal basis for $L^2[0,1]$ is given by

$$\{e^{2\pi int} : n \in Z\}$$

- So any $f \in L^2[0,1]$ can be decomposed as

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi int}$$

$$c_n = \left\langle e^{2\pi int}, f \right\rangle = \int_0^1 f(t) e^{-2\pi int} dt$$

- The average value of $f(x)$ on $[0,1]$ is given by

$$\int_0^1 f(t)dt = \int_0^1 f(t) e^{-2\pi i \cdot 0 \cdot t} = c_0$$

- c_k and c_{-k} correspond to the frequency content of $f(x)$ at frequency $2\pi|k|$

- If $f(t)$ is periodic with period T, then we have $L^2[0,T]$ and an ONB is given by

$$\left\{\frac{1}{\sqrt{T}}e^{2\pi int/T} : n \in Z\right\}$$

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi int/T} dt$$

$$c_n = \frac{1}{T}\left\langle e^{2\pi int/T}, f\right\rangle = \frac{1}{T}\int_0^T f(t)e^{-2\pi int/T}dt$$

- Parseval's formula

$$\sum_{n=-\infty}^{\infty} |c_n|^2 = \frac{1}{T}\int_0^T |f(t)|^2 dt$$

$$\sum_{n=-\infty}^{\infty} |c_n|^2 = \int_0^1 |f(t)|^2 dt$$

Chapter 4: Eigenvalues and Eigenvectors

4.1 Eigenvalues and eigenvectors

- Eigenvalue/eigenvector pair of $n \times m$ matrix: $A\mathbf{v} = \lambda\mathbf{v}$
- Eigenvalues are the roots of the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$. For each λ_j, solve $(A - \lambda_j I)\mathbf{v} = \mathbf{0}$. This subspace is the eigenspace E_{λ_j}.
- The algebraic multiplicity m_j of an eigenvalue λ_j is the power to which $(\lambda - \lambda_j)$ divides the polynomial. The geometric multiplicity d_j is $\dim(E_{\lambda_j})$. $1 \leq d_j \leq m_j$ for all j.
- An eigenbasis exists iff $d_j = m_j$ for each j. In particular, if there are n distinct eigenvalues, then $d_j = m_j = 1$ for each j.
- A is diagonalizable if it has an eigenbasis

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

- From diagonalization, we easily see that

$$\det(A) = \lambda_1 \cdots \lambda_n$$

$$\mathrm{tr}(A) = \lambda_1 + ... + \lambda_n$$

$$A^k = SD^kS^{-1} \text{ for all } k \in Z^+ \text{ and for } k \in Z^- \text{ if } D^{-1} \text{ exists}$$

4.2 Hermitian matrices

- A square matrix is Hermitian if

$$A^* = A$$

  A Hermitian matrix that is real is also symmetric.
- A is Hermitian iff $\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, A\mathbf{w}\rangle$
- A Hermitian matrix has
  - Real eigenvalues
  - Eigenvectors that correspond to distinct eigenvalues are orthogonal
- Every Hermitian matrix is unitarily diagonalizable

$$A = UDU^*$$

4.3 Power method

- Assumptions
  - $|\lambda_1| > |\lambda_2| \geq ... \geq |\lambda_n|$
  - $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is an eigenbasis
  - $\|\mathbf{v}_j\| = 1$ for all j
- Algorithm

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n \quad \text{random}$$
$$A^k\mathbf{x}_0 \approx \lambda_1^k c_1 \mathbf{v}_1$$
$$\mathbf{v}_1 = \frac{A^k\mathbf{x}_0}{\|A^k\mathbf{x}_0\|}$$
$$\lambda_1 = \langle \mathbf{v}_1, A\mathbf{v}_1 \rangle$$

- The random vector $\mathbf{x}_0$ should have a component in the $\mathbf{v}_1$ direction, otherwise power method fails
- To find the eigenvalue closest to s, use the power method on $(A - sI)^{-1}$. This returns $\mathbf{v}_{j^*}$, and we find $\lambda_{j^*} = \langle \mathbf{v}_{j^*}, A\mathbf{v}_{j^*} \rangle$.

4.4 Recursion relations

- Consider the Fibonacci sequence: $F_{n+1} = F_n + F_{n-1}$

$$\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- Then diagonalize A to easily compute A^k
- If there exists $\lambda_i = e^{i\frac{2\pi}{T}}$, then choosing $\mathbf{x}_0 = \mathbf{v}_i$ causes periodic oscillations of period T

4.6 Markov chains

- p_{ij} is the transition probability from j to i
- $x_{n,i}$ is the probability of being at state i at time n. Define the state vector $\mathbf{x}_n = \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ x_{n,3} \end{bmatrix}$
- After n iterations,

$$\mathbf{x}_n = P^n\mathbf{x}_0$$

- A matrix P is stochastic if

$$P = \begin{bmatrix} p_{11} & ... & p_{1k} \\ \vdots & & \vdots \\ p_{k1} & ... & p_{kk} \end{bmatrix}$$

such that

$$0 \leq p_{ij} \leq 1 \quad \text{(non-zero probabilities)}$$
$$\sum_{i=1}^{k} p_{ij} = 1 \text{ for each } j \quad \text{(column sum is 1)}$$

- When n is large and $\lim_{n\to\infty} P^n\mathbf{x}$ exists,

$$P^n\mathbf{x}_0 = P\mathbf{x} = \mathbf{x}$$

So $P\mathbf{x} = \mathbf{x}$ and $\mathbf{x}$ is a stationary point and corresponds to $\lambda = 1$. Eigenvector $\mathbf{v}_1$ corresponding to $\lambda_1 = 1$ has non-negative entries. All other eigenvalues of P satisfy $|\lambda_j| \leq 1$. Eigenvectors corresponding to $|\lambda_j| < 1$ have entries that add up to 0.

- If P or P^k for some $k \in N$ has all positive entries, then λ_1 is the dominant eigenvalue, and normalizing $\mathbf{v}_1$ so that its 1-norm is 1 gives the stationary point. This occurs when there is a nonzero probability of ending up anywhere in the graph.

- For Google pagerank, we define a stochastic matrix that represents the probability of each page linking to every other page. If there are no links out of a node, then we assign equal probabilities from that node to any node (including staying at itself).

- Because P often has so many 0 entries, it almost always has many eigenvalues $|\lambda| = 1$, so we can't use power method to compute $\mathbf{v}_1$. To remedy this, we choose the damping factor $0 \leq \alpha \leq 1$ and use the Google matrix.

$$G = \alpha S + (1-\alpha)Q \quad \text{where} \quad Q = \begin{bmatrix} \frac{1}{N} & \cdots & \frac{1}{N} \\ \vdots & & \vdots \\ \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix}$$

This guarantees that $|\lambda_1| = 1$ is the dominant eigenvalue and the power method can be used.

4.7 Singular value decomposition

- Singular value decomposition generalizes diagonalization to non-square matrices.

$$A = U \cdot \Sigma \cdot V^*$$

- Observations that help us derive the SVD
  - All eigenvalues of A^*A are non-negative
  - A^*A and AA^* have identical non-zero eigenvalues (but may have different numbers of zero eigenvalues)
  - Both A^*A and AA^* are Hermitian (and thus unitarily diagonalizable)

- We get $A = U\Sigma V^*$, where
  - V is formed by eigenvectors of A^*A

$$A^*A = V\Sigma_1^2 V^*$$

  - U is formed by eigenvectors of AA^*

$$AA^* = U\Sigma_2^2 U^*$$

  - Σ is formed by the $\min\{m, n\}$ eigenvalues of A^*A and AA^*

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & 0 & \cdots & 0 \\ & \sigma_2 & & & 0 & \cdots & 0 \\ & & \ddots & & \vdots & & \vdots \\ & & & \sigma_m & 0 & \cdots & 0 \end{bmatrix} \qquad \text{for } m < n$$

- Let r be the number of non-zero singular values of A
  - $N(A) = \{\mathbf{v}_{r+1}, ..., \mathbf{v}_n\}$

- $R(A) = \{\mathbf{u}_1, ..., \mathbf{u}_r\}$
- $N(A^*) = \{\mathbf{u}_1, ..., \mathbf{u}_r\}^\perp = \{\mathbf{u}_{r+1}, ..., \mathbf{u}_m\}$

- Matrix norms
  - $\|A\| = \sigma_1$
  - $\|A\|_{HS}^2 = \sigma_1^2 + ... + \sigma_r^2$

- The left-inverse of A is given by

$$A^\dagger = V\Sigma^\dagger U^*$$

where

$$\Sigma^\dagger = \begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n} \\ 0 & ... & 0 \\ \vdots & & \vdots \\ 0 & ... & 0 \end{bmatrix}$$