Chapter 7 Notes - LA

John Yang

April 29, 2022

Contents

7 Distance and Approximation

7.1 Inner Product Spaces

- An inner product on a vector space V is an operation that assigns to every pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ such that the following properties hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and all scalars c:
  - $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
  - $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
  - $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
  - $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ IFF $\mathbf{u} = \mathbf{0}$
- A vector space with an inner product is called an inner product space.
- Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in an inner product space V and let c be a scalar.
  - $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v} + \mathbf{w} \rangle$
  - $\langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
  - $\langle \mathbf{u}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle = 0$
- Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V.
  - The length (or norm) of $\mathbf{v}$ is $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.
  - The distance between $\mathbf{u}$ and $\mathbf{v}$ is $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$
  - $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.
- Pythagoras' Theorem: Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. Then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal IFF
$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

- The Cauchy-Schwarz Inequality: Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. Then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

with equality holding IFF $\mathbf{u}$ and $\mathbf{v}$ are scalar multiples of each other.

- The triangle inequality: Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. Then

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

7.2 Norms and Distance Functions

- A norm on a vector space V is a mapping that associates with each vector $\mathbf{v}$ a real number $\|\mathbf{v}\|$, called the norm of $\mathbf{v}$, such that the following properties are satisfied for all vectors $\mathbf{u}$ and $\mathbf{v}$ and all scalars c:
  - $\|\mathbf{v}\| \geq 0$, and $\|\mathbf{v}\| = 0$ IFF $\mathbf{v} = \mathbf{0}$
  - $\|c\mathbf{v}\| = |c| \|\mathbf{v}\|$
  - $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
- A vector space with a norm is called a normed vector space.
- We define a distance function for any norm as:

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

- Let d be a distance function defined on a normed linear space V. The following properties hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V:
  - $d(\mathbf{u}, \mathbf{v}) \geq 0$, and $d(\mathbf{u}, \mathbf{v}) = 0$ IFF $\mathbf{u} = \mathbf{v}$
  - $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$
  - $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$
- A matrix norm on M_{nn} is a mapping that associates with each $n \times n$ matrix A a real number $\|A\|$, called the norm of A, such that the following properties are satisfied for all $n \times n$ matrices A and B and all scalars c.
  - $\|A\| \geq 0$ and $\|A\| = 0$ IFF $A = O$.
  - $\|cA\| = |c| \|A\|$
  - $\|A + B\| \leq \|A\| + \|B\|$
  - $\|AB\| \leq \|A\| \|B\|$
- A matrix norm on M_{nn} is said to be compatible with a vector norm on $\|\mathbf{x}\|$ on $\mathbb{R}^n$ if, for all $n \times n$ matrices A and all vectors $\mathbf{x}$ in $\mathbb{R}^n$, we have

$$\|A\mathbf{x}\| \leq \|A\| \|\mathbf{x}\|$$

- The Frobenius norm is given by

$$\|A\|_F = \sqrt{\sum_{i,j=1}^{n} a_{ij}^2}$$

- If $\|\mathbf{x}\|$ is a vector norm on $\mathbb{R}^n$, then $\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$ defines a matrix norm on M_{nn} that is compatible with the vector norm that induces it.
- The matrix norm $\|A\|$ in the previous is called the operator norm induced by the vector norm $\|\mathbf{x}\|$

- Let A be an $n \times n$ matrix with column vectors $\mathbf{a}_i$ and row vectors $\mathbf{A}_i$ for $i = 1, \ldots, n$.

$$\text{a. } \|A\|_1 = \max_{j=1,\ldots,n} \left\{ \|\mathbf{a}_j\|_s \right\} = \max_{j=1,\ldots,n} \left\{ \sum_{i=1}^{n} |a_{ij}| \right\}$$

$$\text{b. } \|A\|_\infty = \max_{i=1,\ldots,n} \left\{ \|\mathbf{A}_i\|_s \right\} = \max_{i=1,\ldots,n} \left\{ \sum_{j=1}^{n} |a_{ij}| \right\}$$

- A matrix A is ill-conditioned if small changes in its entries can produce large changes in the solutions to $A\mathbf{x} = \mathbf{b}$. If small changes in the entries of A produce only small changes in the solutions to $A\mathbf{x} = \mathbf{b}$, then A is called well-conditioned.

7.3 Least Squares Approximation

- If A is an $m \times n$ matrix and $\mathbf{b}$ is in $\mathbb{R}^m$, a least squares solution of $\overline{A\mathbf{x}} = \mathbf{b}$ is a vector $\overline{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\overline{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

  for all $\mathbf{x}$ in $\mathbb{R}^n$.

- The Least Squares Theorem: Let A be an $m \times n$ matrix and let $\mathbf{b}$ be in $\mathbb{R}^m$. Then $A\mathbf{x} = \mathbf{b}$ always has at least one least squares solution $\overline{\mathbf{x}}$. Moreover:
  - $\overline{\mathbf{x}}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\overline{\mathbf{x}}$ is a solution of the normal equations $A^T A\overline{\mathbf{x}} = A^T\mathbf{b}$.
  - A has linearly independent columns if and only if $A^T A$ is invertible. In this case, the least squares solution of $A\mathbf{x} = \mathbf{b}$ is unique and is given by

$$\overline{\mathbf{x}} = \left(A^T A\right)^{-1} A^T\mathbf{b}$$

- Let A be an $m \times n$ matrix with linearly independent columns and let $\mathbf{b}$ be in $\mathbb{R}^m$. If $A = QR$ is a QR factorization of A, then the unique least squares solution $\overline{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ is

$$\overline{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$$

- Let W be a subspace of $\mathbb{R}^m$ and let A be an $m \times n$ matrix whose columns form a basis for W. If $\mathbf{v}$ is any vector in $\mathbb{R}^m$, then the orthogonal projection of $\mathbf{v}$ onto W is the vector

$$\text{proj}_W(\mathbf{v}) = A\left(A^T A\right)^{-1} A^T\mathbf{v}$$

  The linear transformation $P : \mathbb{R}^m \to \mathbb{R}^m$ that projects $\mathbb{R}^m$ onto W has $A\left(A^T A\right)^{-1} A^T$ as its standard matrix.

- If A is a matrix with linearly independent columns, then the pseudoinverse of A is the matrix A^+defined by

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

- Let A be a matrix with linearly independent columns. Then the pseudoinverse $A^+$of A satisfies the following properties, called the Penrose conditions for A:
  - $AA^+A = A$
  - $A^+AA^+ = A^+$
  - $AA^+$and A^+A are symmetric.

7.4 The Singular Value Decomposition

- If A is an $m \times n$ matrix, the singular values of A are the square roots of the eigenvalues of $A^T A$ and are denoted by $\sigma_1, \ldots, \sigma_n$. It is conventional to arrange the singular values so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

- The Singular Value Decomposition: Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Then there exist an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ matrix Σ of the form shown in Equation (1) such that

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

- The Outer Product Form of the SVD: Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Let $\mathbf{u}_1, \ldots, \mathbf{u}_r$ be left singular vectors and let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be right singular vectors of A corresponding to these singular values. Then

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

- Let $A = U \Sigma V^T$ be a singular value decomposition of an $m \times n$ matrix A. Let $\sigma_1, \ldots, \sigma_r$ be all the nonzero singular values of A. Then:
  - The rank of A is r.
  - $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is an orthonormal basis for $\operatorname{col}(A)$.
  - $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ is an orthonormal basis for $\operatorname{null}\left(A^T\right)$.
  - $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for $\operatorname{row}(A)$.
  - $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for $\operatorname{null}(A)$.

- Let A be an $m \times n$ matrix with rank r. Then the image of the unit sphere in $\mathbb{R}^n$ under the matrix transformation that maps $\mathbf{x}$ to $A\mathbf{x}$ is
  - the surface of an ellipsoid in $\mathbb{R}^m$ if $r = n$.
  - a solid ellipsoid in $\mathbb{R}^m$ if $r < n$.

- Let A be an $m \times n$ matrix and let $\sigma_1, \ldots, \sigma_r$ be all the nonzero singular values of A. Then

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$

- If A is an $m \times n$ matrix and Q is an $m \times m$ orthogonal matrix, then

$$\|QA\|_F = \|A\|_F$$

- Let $A = U \Sigma V^T$ be an SVD for an $m \times n$ matrix A, where $\Sigma = \begin{bmatrix} D & O \\ O & O \end{bmatrix}$ and D is an $r \times r$ diagonal matrix containing the nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ of A. The pseudoinverse (or Moore-Penrose inverse) of A is the $n \times m$ matrix A^+defined by

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

  where $\Sigma^+$is the $n \times m$ matrix

$$\Sigma^+ = \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}$$

- The least squares problem $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution $\overline{\mathbf{x}}$ of minimal length that is given by

$$\overline{\mathbf{x}} = A^+ \mathbf{b}$$

- The Fundamental Theorem of invertible matrices: Final Version.

- A is invertible
- $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$
- $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- The reduced row echelon form of A is I_n.
- A is the product of elementary matrices.
- $\text{rank}(A) = n$
- $\text{nullity}(A) = 0$
- The column vectors of A are linearly independent
- The column vectors of A span $\mathbb{R}^n$
- The column vectors of A form a basis for $\mathbb{R}^n$
- The row vectors of A are linearly independent
- The row vectors of A span $\mathbb{R}^n$
- The row vectors of A form a basis for $\mathbb{R}^n$
- $\det A \neq 0$
- 0 is not an eigenvalue of A
- T is invertible.
- T is one-to-one.
- T is onto.
- $\ker(T) = \{\mathbf{0}\}$
- $\text{range}(T) = W$
- 0 is not a singular value of A.

7.5 Applications

- General problem of approximating functions can be stated as: Given a continuous function f on an interval $[a, b]$ and a subspace W of $\mathscr{C}[a, b]$, find the function "closest" to f in W.
- The n'th order Fourier approximation to f on $[-\pi, \pi]$:

$$
\begin{aligned}
a_0 &= \frac{\langle 1, f\rangle}{\langle 1, 1\rangle} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx \\
a_k &= \frac{\langle \cos kx, f\rangle}{\langle \cos kx, \cos kx\rangle} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx dx \\
b_k &= \frac{\langle \sin kx, f\rangle}{\langle \sin kx, \sin kx\rangle} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx dx
\end{aligned}
$$