Math Foundations: Review for CSE 203B

Chester Holtz

chholtz@eng.ucsd.edu
OH: Fri 6:30 - 7:30 pm

Overview

- Notations & general assumptions
- Vector norms, inner products
- Linear spaces, subspaces, linear transformations
- Eigenvalues / eigenvectors, rank, SVD, inverse
- Matrix norms
- Matrix and vector differential

Notations & basic assumptions

- Greek alphabet α, β, γ denote real numbers
- Small letters x, y, z denote vectors
- Capital letters denote matrices A, B, C

Notations & basic assumptions

- Greek alphabet α, β, γ denote real numbers
- Small letters x, y, z denote vectors
- Capital letters denote matrices A, B, C
- $\mathbb{R}$ is the one dimensional Euclidean space
- $\mathbb{R}^n$ is the *n*-dimensional *vector* Euclidean space
- $\mathbb{R}^{m \times n}$ is the $m \times n$ dimensional *matrix* Euclidean space
- $\mathbb{R}_+$ denotes the range $[0, +\infty)$, $\mathbb{R}_{++}$ denotes the range $(0,_\infty)$
- $1_n \in \mathbb{R}^n$ denotes a vector with 1 in all entries
- For any vector $x \in \mathbb{R}^n$, $|x|_i = |x|_i \quad \forall i = 1, \dots, n$

Vector norms, Inner product

A function $f : x \in \mathbb{R}^n \rightarrow y \in \mathbb{R}_+$ is called a norm if the following conditions are satisfied:

1. (zero element) $f(x) \geq 0$ and $f(x) = 0$ iff $x = 0$

Vector norms, Inner product

A function $f : x \in \mathbb{R}^n \to y \in \mathbb{R}_+$ is called a norm if the following conditions are satisfied:

1. (zero element) $f(x) \geq 0$ and $f(x) = 0$ iff $x = 0$
2. (Homogeneity) For any $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$, $f(\alpha x) = |\alpha| f(x)$

Vector norms, Inner product

A function $f : x \in \mathbb{R}^n \to y \in \mathbb{R}_+$ is called a norm if the following conditions are satisfied:

1. (zero element) $f(x) \geq 0$ and $f(x) = 0$ iff $x = 0$
2. (Homogeneity) For any $\alpha \in \mathbb{R}$and $x \in \mathbb{R}^n$, $f(\alpha x) = |\alpha| f(x)$
3. (Triangle inequality) $x, y \in \mathbb{R}^n$ satisfy $f(x) + f(y) \geq f(x + y)$

Vector norms, Inner product

1. (zero element) $f(x) \geq 0$ and $f(x) = 0$ iff $x = 0$
2. (Homogeneity) For any $\alpha \in \mathbb{R}$and $x \in \mathbb{R}^n$, $f(\alpha x) = |\alpha| f(x)$
3. (Triangle inequality) $x, y \in \mathbb{R}^n$ satisfy $f(x) + f(y) \geq f(x + y)$

Example

- ℓ_1 norm "$|| \cdot ||_1$" is defined as $||x||_1 = (|x_1| + |x_2| + \ldots |x_n|)$
- ℓ_2 norm "$|| \cdot ||_2$" is defined as $||x||_2 = (|x_1|^2 + |x_2|^2 + \ldots |x_n|^2)^{\frac{1}{2}}$

Vector norms, Inner product

In general, an ℓ_p norm $(p \geq 1)$ is defined as

$$||x||_p = (|x_1|^p + |x_2|^p + \dots |x_n|^p)^{\frac{1}{p}}$$

Note:

- for $p < 1$, triangle inequality is violated.
- $||x||_\infty = \lim_{p \to +\infty} ||x||_p \max\{|x_1|, |x_2|, \dots, |x_n|\}$

Vector norms, **Inner product**

The inner product. $\langle \cdot, \cdot \rangle$ in $\mathbb{R}^n$ is defined as

$$\langle x, y \rangle = \sum_i x_i y_i$$

Vector norms, Inner product

The inner product. $\langle \cdot, \cdot \rangle$ in $\mathbb{R}^n$ is defined as

$$\langle x, y \rangle = \sum_i x_i y_i$$

Note that $\langle x, x \rangle = ||x||^2$. Two vectors x and y are orthogonal if $\langle x, y \rangle = 0$

Vector norms, Inner product

The inner product. $\langle \cdot, \cdot \rangle$ in $\mathbb{R}^n$ is defined as

$$\langle x, y \rangle = \sum_i x_i y_i$$

Note that $\langle x, x \rangle = ||x||^2$. Two vectors x and y are orthogonal if $\langle x, y \rangle = 0$

Furthermore, if $p \geq q$, then for any $x \in \mathbb{R}^n$, $||x||_p \leq ||x||_q$. In particular,

$$||x||_1 \geq ||x||_2 \geq ||x||_\infty$$

Vector norms, Inner product

$$||x||_1 \leq \sqrt{n}||x||_2 \quad ||x||_2 \leq \sqrt{n}||x||_\infty$$

Proof

Vector norms, Inner product

$$||x||_1 \leq \sqrt{n}||x||_2 \quad ||x||_2 \leq \sqrt{n}||x||_\infty$$

Proof

$$||x||_1 = \langle 1_n, |x| \rangle \leq ||1_n||_2 \, |||x|||_2 = \sqrt{n}||x||_2$$

Cauchy Schwarz inequality:

$$|\langle u, v \rangle|^2 \leq \langle u, u \rangle \langle v, v \rangle$$

or

$$|\langle u, v \rangle| \leq ||u|| \, ||v||$$

Vector norms, Inner product

Given a norm $||x||_A$, its dual norm is defined as

$$||x||_{A^*} = \max_{||y||_A \leq 1} \langle x, y \rangle = \max_{||y||_A = 1} \langle x, y \rangle = \max_z \frac{\langle x, z \rangle}{||z||_A}$$

Vector norms, Inner product

Given a norm $||x||_A$, its dual norm is defined as

$$||x||_{A^*} = \max_{||y||_A \leq 1} \langle x, y \rangle = \max_{||y||_A = 1} \langle x, y \rangle = \max_z \frac{\langle x, z \rangle}{||z||_A}$$

- The dual norm's dual norm is itself. $||x||_{(A^*)^*} = ||x||_A$

Vector norms, Inner product

Given a norm $||x||_A$, its dual norm is defined as

$$||x||_{A^*} = \max_{||y||_A \leq 1} \langle x, y \rangle = \max_{||y||_A = 1} \langle x, y \rangle = \max_z \frac{\langle x, z \rangle}{||z||_A}$$

- The dual norm's dual norm is itself. $||x||_{(A^*)^*} = ||x||_A$
- The ℓ_2 norm is self-dual

Vector norms, Inner product

Given a norm $||x||_A$, its dual norm is defined as

$$||x||_{A^*} = \max_{||y||_A \leq 1} \langle x, y \rangle = \max_{||y||_A = 1} \langle x, y \rangle = \max_{z} \frac{\langle x, z \rangle}{||z||_A}$$

- The dual norm's dual norm is itself. $||x||_{(A^*)^*} = ||x||_A$
- The ℓ_2 norm is self-dual
- In general, the dual norm of an ℓ_p norm is an ℓ_q norm where p, q satisfy $1/p + 1/q = 1$

Vector norms, Inner product

Given a norm $||x||_A$, its dual norm is defined as

$$||x||_{A^*} = \max_{||y||_A \leq 1} \langle x, y \rangle = \max_{||y||_A = 1} \langle x, y \rangle = \max_z \frac{\langle x, z \rangle}{||z||_A}$$

- The dual norm's dual norm is itself. $||x||_{(A^*)^*} = ||x||_A$
- The ℓ_2 norm is self-dual
- In general, the dual norm of an ℓ_p norm is an ℓ_q norm where p, q satisfy $1/p + 1/q = 1$
- (Holder inequality): $\langle x, y \rangle \leq ||x||_A ||y||_{A^*}$

Linear space, subspace, linear transformation

A set S is a linear space if

- $0 \in S$
- given any two points $x, y \in S$ and scalars $\alpha, \beta \in \mathbb{R}$.

$$\alpha x + \beta y \in S$$

Linear space, subspace, linear transformation

- $0 \in S$
- given any two points $x, y \in S$ and scalars $\alpha, \beta \in \mathbb{R}$, $\alpha x + \beta y \in S$

examples

- $\emptyset$?
- 0 ?
- $\{0\}$?
- $\{x | Ax = b\}$?

Linear space, **subspace**, linear transformation

Let S be a linear space. A set S' is a subspace if S' is a linear space and also a subset of S.

Linear space, subspace, **linear transformation**

Let S be a linear space. A function $L(\cdot)$ is a linear transformation if given $x, y \in S$ and scalars $\alpha, \beta \in \mathbb{R}$,

$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y)$$

Note 1-1 correspondence between linear transformations and matrices.

Linear space, subspace, linear transformation

Expressing a subspace

A bunch of vectors. The range space of a matrix X:

$$\mathrm{span}\{x_1, x_2, \ldots, x_n\} = \left\{ \sum_{i=1}^{n} \alpha_i x_i | \alpha_i \in \mathbb{R} \right\} = \{X\alpha | \alpha\}$$

Linear space, subspace, linear transformation

Expressing a subspace

A bunch of vectors. The range space of a matrix X:

$$\mathrm{span} x_1, x_2, \ldots, x_n = \left\{ \sum_{i=1}^{n} \alpha_i x_i | \alpha_i \in \mathbb{R} \right\} = \{X\alpha | \alpha\}$$

The null space of X:

$$\{\alpha | X\alpha = 0\}$$

Eigenvalues / eigenvectors, rank, SVD, inverse

The transpose of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as $A^T \in \mathbb{R}^{n \times n}$:

$$(A^T)_{ij} = A_{ji}$$

Eigenvalues / eigenvectors, rank, SVD, inverse

The transpose of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as $A^T \in \mathbb{R}^{n \times n}$:

$$(A^T)_{ij} = A_{ji}$$

Can verify that

$$(AB)^T = B^T A^T$$

Eigenvalues / eigenvectors, rank, SVD, inverse

A matrix $B \in \mathbb{R}^{n \times n}$ is the inverse of an invertible matrix $A \in \mathbb{R}^{n \times n}$ if:

$$AB = I \quad \text{and} \quad BA = I$$

Note the following properties:

- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^T)^{-1} = (A^{-1})^T$

Eigenvalues / eigenvectors, rank, SVD, inverse

Given a square matrix $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$, $(x \neq 0)$ is called its eigenvector and $\lambda \in \mathbb{R}$ is its associated eigenvalue if:

$$Ax = \lambda x$$

Eigenvalues / eigenvectors, rank, SVD, inverse

Given a square matrix $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$, $(x \neq 0)$ is called its eigenvector and $\lambda \in \mathbb{R}$ is its associated eigenvalue if:

$$Ax = \lambda x$$

Properties

- If the matrix A is symmetric, any two eigenvectors (corresponding to different eigenvalues) are orthogonal.
- $\det A = \prod i \lambda_i$
- The rank of A is equal to the number of non-zero eigenvalues.
- If A is invertible, $1/\lambda_i$ is an eigenvalue of A^{-1}
- $\lambda_{\max} = \sup_{x \neq 0} \frac{x^T A x}{x^T x}$

Eigenvalues / eigenvectors, rank, SVD, inverse

If $A^T = A$, $Ax_1 = \lambda_1 x_1$, $Ax_2 = \lambda_2 x_2$, and $\lambda_1 \neq \lambda_2$, then $x_1^T x_2 = 0$

proof

Consider $X_1^T A x_2$. We have that

$$x_1^T A x_2 = x_1^T (A x_2) = x_1^T (\lambda_2 x_2) = \lambda_2 x_1^T x_2$$

and

$$x_1^T A x_2 = (x_1^T A) x_2 = (A^T x_1)^T x_2 = (A x_1)^T x_2 = \lambda_1 x_1^T x_2$$

So

$$\lambda_2 x_1^T x_2 = \lambda_1 x_1^T x_2$$

and since $\lambda_1 \neq \lambda_2, x_1^T x_2 = 0$.

Eigenvalues / eigenvectors, rank, SVD, inverse

The rank of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as

$$\mathrm{rank}(A) = \min \left\{ r | A = \sum_{i=1}^{r} x_i {y_i}^T, x_i, y_i \in \mathbb{R}^n \right\}$$

Eigenvalues / eigenvectors, rank, SVD, inverse

The rank of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as

$$\mathrm{rank}(A) = \min \left\{ r | A = \sum_{i=1}^{r} x_i {y_i}^T, x_i, y_i \in \mathbb{R}^n \right\}$$

Properties

- $\mathrm{rank}(A) \leq \min\{m, n\}$ (equality = "full-rank")
- $\mathrm{rank}(A) = \mathrm{rank}(A^T)$
- $\mathrm{rank}(AB) \leq \min\{\mathrm{rank}(AB)\}$
- $\mathrm{rank}(A + B) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$
- $\mathrm{rank}(A) + \mathrm{Nullity}(A) = \mathrm{Dim}(V)$ (rank-nullity theorem)

Eigenvalues / eigenvectors, rank, SVD, inverse

The determinant of a square matrix $A \in \mathbb{R}^{n \times n}$ is a scalar-valued function $\det(A) : \mathbb{R}^{n \times n} \to \mathbb{R}$.
Consider the set of all linear combinations of the rows of A:

$$S = \{v \in \mathbb{R}^n | v = \sum_{i=1}^{n} \alpha_i a_i, 0 \leq \alpha_i \leq 1, i = 1, \ldots, n\}$$

$|\det(A)|$ is the area of the n-dimensional parallelotope.

Eigenvalues / eigenvectors, rank, SVD, inverse

The determinant of a square matrix $A \in \mathbb{R}^{n \times n}$ is a scalar-valued function $\det(A) : \mathbb{R}^{n \times n} \to \mathbb{R}$.
Consider the set of all linear combinations of the rows of A:

$$S = \{v \in \mathbb{R}^n | v = \sum_{i=1}^{n} \alpha_i a_i, 0 \leq \alpha_i \leq 1, i = 1, \ldots, n\}$$

$|\det(A)|$ is the area of the n-dimensional parallelotope.

- If $\operatorname{rank}(A) < n$, $\det(A) = 0$
- If $\operatorname{rank}(A) = n$, $\det(A) \neq 0$
- see wiki_link for more properties.

Eigenvalues / eigenvectors, rank, SVD, inverse

Given any matrix $A \in \mathbb{R}^{m \times n}$,

$$A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i U_{i\cdot} V_{i\cdot}^T$$

where $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$ have orthogonal columns and $\Sigma = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ is a diagonal matrix with positive diagonal elements "singular values".

Eigenvalues / eigenvectors, rank, SVD, inverse

Given any matrix $A \in \mathbb{R}^{m \times n}$,

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i U_{i.} V_{i.}^T$$

where $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$ have orthogonal columns and $\Sigma = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ is a diagonal matrix with positive diagonal elements "singular values".

- $\mathrm{rank}(A) = r$
- $||Ax|| \leq \sigma_1 ||x||$. why?

Eigenvalues / eigenvectors, rank, SVD, inverse

A matrix $B \in \mathbb{R}^{n \times n}$ is called positive semi-definite (PSD), if the following are satisfied:

- B is symmetric
- $\forall x \in \mathbb{R}^n$, $x^T B x \geq 0$

Note B is PSD if B can be written: $B = U \Sigma U^T$, where $U^T U = I$

Matrix norms

- Frobeneus norm: $A_F = \left(\sum_{i,m} |A_{ij}|^2|\right)^{\frac{1}{2}} = \left(\sum_{i=1} \sigma_i^2\right)^{\frac{1}{2}}$
- spectral (trace) norm:
  $||A||_{\mathrm{spec}} \max_{||x||=1} ||Ax|| = \max_{||x||=1,||y||=1} y^T Ax = \sigma_1(A)$
- nuclear norm: $||A||_* = \sum_i \sigma_i(A) = \mathrm{trace}(\Sigma)$

Note if A is a vector, $||A||_F = ||A||_2$.

Matrix norms

The inner product $\langle \cdot, \cdot \rangle$ in $\mathbb{R}^{m \times n}$ is defined as:

$$\langle X, Y \rangle = \sum_{ij} X_{ij} Y_{ij} = \mathrm{trace}(X^T Y)$$

In general,
$\mathrm{trace}(AB) = \mathrm{trace}(BA) = \mathrm{trace}(A^T B^T) = \mathrm{trace}(B^T A^T)$

Matrix and vector differential

Let $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ be a (scalar-valued) continuous & differentiable function. It's differential (gradient) is defined as:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

Matrix and vector differential

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

example

Let $f(x) = 1^T x = \sum_i x_i$.

$$\nabla f(x) = \mathbf{1}$$

Matrix and vector differential

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

example

Let $f(x) = x^T x = \sum_i x_i^2$.

$$\nabla f(x) = 2x$$

Matrix and vector differential

The product and chain rules hold when dealing with gradients of vector functions:

- Product rule: $\nabla(f(x)g(x)) = f(x)\nabla g(x) + \nabla f(x)g(x)$
- Chain rule: $\frac{\partial}{\partial t}f(g(t)) = \nabla f(g(t))^T \frac{\partial g}{\partial t}$

Matrix and vector differential

The Hessian $\nabla^2 f = H$ is a matrix with entries $= f(x)$'s second-order derivatives:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x}{(\partial x_1)^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f(x)}{(\partial x_n)^2} \end{bmatrix}$$

Matrix and vector differential

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x}{(\partial x_1)^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f(x)}{(\partial x_n)^2} \end{bmatrix}$$

example

$f(x) = \frac{1}{2}x^T Ax = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$, $A = I$.
$\nabla^2 f(x) = A$

Matrix and vector differential

Let $f(X) : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ be a (scalar-valued) function. It's differential (gradient) is defined as:

$$\frac{\partial f(X)}{\partial X} = \begin{bmatrix} \frac{\partial f(X)}{\partial X_{11}} & \cdots & \frac{\partial f(X)}{\partial X_{in}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(X)}{\partial X_{m1}} & \cdots & \frac{\partial f(X)}{\partial X_{mn}} \end{bmatrix}$$

Linear equalities and inequalities

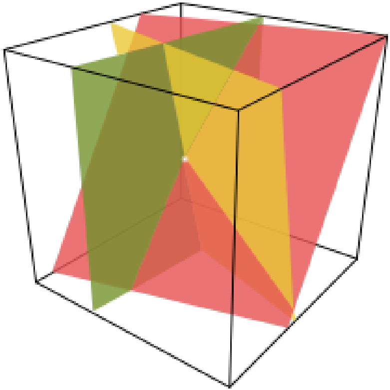
A linear system can be described as the matrix equality $Ax = b$. A solution exists if there is an assignment to the entries of x such that

$$a_{i1}x_1 + a_{i1}x_2 + \ldots + a_{in}x_n = b_i$$

The LHS defines a linear combination of A's column vectors - i.e. the system as a solution if b is in the space spanned by the columns of A.

Linear equalities and inequalities

The solution to system of linear equalities corresponds to the point of intersection of m hyperplanes.



Linear equalities and inequalities

Alternatively, the solution set to system of linear inequalities: $\{x|Ax \leq b\}$ defines the intersection of m half-planes.