EECS 545 Linear Algebra Review

Sudeep Katakol

Jan 7, 2022

Review class

- Note to self: Turn on zoom recording
- Goal is to provide a *quick review* of the mathematical concepts from linear algebra that we will be using throughout the course.
- Interactive format, participate!
- Exercises: Think 30s - 1 min on your own and then discuss with neighbors.

Table of Contents

Vectors and norms

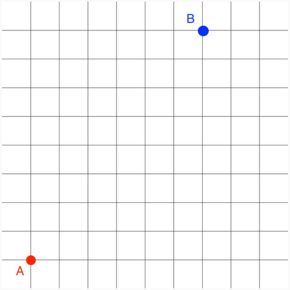
Vectors

- A vector $\mathbf{x} \in \mathbb{R}^n$ is a stack of n real values.
- $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x_1 \; x_2 \; \dots \; x_n]^T$
- For instance, in ML context, $\mathbf{x}$ could denote the **features** or input data. In the housing price example from Lecture 1 (Slide 46), x_1 could denote the number of rooms, x_2 could denote the area code, etc.

Norms

- Norms are a measure of **magnitude** of the vector.
- Formally, a norm $\|\cdot\| : \mathbb{R}^n \to [0, \infty)$ is a non-negative valued function which satisfies the four properties below:
  - Non-negative: $\|\mathbf{x}\| \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$
  - Positive: $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$
  - Homogeneous: $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\| \quad \forall \alpha \in \mathbb{R}$
  - Triangle inequality: $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- Examples:
  - Euclidean (l_2) norm (Default choice): $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$
  - Manhattan distance (l_1 norm): $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \ldots + |x_n|$
  - In general l_p norm ($p \geq 1$): $\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p}$
  - l_∞ norm: $\|\mathbf{x}\|_\infty := \lim_{p \to \infty} \|\mathbf{x}\|_p = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$

- **Exercise**: Your goal is to move from point A $(0, 0)$ to point B $(6, 8)$ on the grid. Draw the paths of your walk if its required that the number of footsteps you take is proportional to the (i) Euclidean distance and (ii) Manhattan distance.



- Non-examples:
  - l_p for $0 < p < 1$.
  - l_0 norm $\|\mathbf{x}\|_0 := \lim_{p \to 0} \|\mathbf{x}\|_p^p = \sum_{i=1}^{n} \mathbb{I}\{x_i \neq 0\}$

Orthogonality

- Inner product of two vectors: $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T\mathbf{x} = \sum_{i=1}^{n} x_i y_i$ (Scalar)
- **Exercise**: Relate the Euclidean norm $\|\mathbf{x}\|_2$ and inner product $\langle \mathbf{x}, \mathbf{x} \rangle$.
- **Orthogonality**: Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal* (denoted by $\mathbf{x} \perp \mathbf{y}$) iff their inner product is zero i.e., $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.
- **Orthogonal set of vectors** is a set of vectors in which any two vectors are orthogonal to each other. Examples:
  - $\{[1\ 0]^T, [0\ 1]^T\}$
  - $\{[1\ -1\ 0]^T, [1\ 1\ 0]^T, [0\ 0\ 1]^T\}$
- **Orthonormal set of vectors**: Orthogonal set of vectors with each vector having unit norm.

Linear Independence

- A set of vectors are said to be **linearly dependent** if one of the vectors can be written as a linear combination of the rest.
- If there is no such vector, the set of vectors are deemed to be **linearly independent**.
- Formally, a set of vector $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K\}$ are said to be **linearly independent** if $\sum_{i=1}^{K} \alpha_i \mathbf{x}_i = 0$ for some $\alpha_i \in \mathbb{R}$, then $\alpha_i = 0 \quad \forall i \in \{1, 2, \ldots, K\}$.
- How is this formal definition consistent with our previous definition?

Linear Independence

- **T/F**: Zero-vector can be present in a linearly independent set of vectors.
- **T/F**: $\mathbf{x}, \mathbf{y}$ are linearly dependent **iff** they are scalar multiples of each other?
- **T/F**: Orthonormal set of non-zero vectors are linearly independent.

Matrices

- A matrix $\mathsf{A} \in \mathbb{R}^{m \times n}$ of size $m \times n$ is grid of real values i.e.,

$$\mathsf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- In terms of columns: $\mathsf{A} = \begin{bmatrix} \mathbf{a}_{:,1} & \mathbf{a}_{:,2} & \cdots & \mathbf{a}_{:,n} \end{bmatrix}$
- In terms of rows: $\mathsf{A} = \begin{bmatrix} \mathbf{a}_{1,:}^T \\ \mathbf{a}_{2,:}^T \\ \vdots \\ \mathbf{a}_{m,:}^T \end{bmatrix}$
- **Exercise**: Write A^T in terms of rows and columns of A.

Matrices

- Matrices can be used to represent data. For instance, a collection of features for ML: $\mathbf{X} = \begin{bmatrix} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \cdots & \mathbf{x}^{(n)} \end{bmatrix} \in \mathbb{R}^{m \times n}$ where $\mathbf{x}^{(i)} \in \mathbb{R}^m$ is the i^{th} feature.
- Matrices also represent a map. Any linear function $f : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ i.e., $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Note that $\mathbf{y}_i = \mathbf{a}_{i,:}^T \mathbf{x} = \sum_{j=1}^{n} a_{ij} x_j$ where $\mathbf{y} = f(\mathbf{x})$.

- A square matrix $\mathsf{A} \in \mathbb{R}^{n \times n}$ is invertible iff there is a matrix $\mathsf{B} \in \mathbb{R}^{n \times n}$ such that $\mathsf{AB} = \mathsf{I}_n = \mathsf{BA}$. Such a B is denoted as A^{-1}.
- Note that it is enough to show only the first equality.
- The linear function represented by A i.e., $f(\mathsf{x}) = \mathsf{Ax}$ is invertible iff A is invertible.
- Property: A is invertible iff $\det \mathsf{A} \neq 0$.

Rank

- Rank of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined as the number of the linearly independent columns (or rows) in the matrix.
- **Exercise**: Determine the rank of the matrix $\mathbf{A} = \mathbf{x}\mathbf{y}^T$ where $\mathbf{x}$ and $\mathbf{y}$ are n-dimensional vectors.
- Invertible matrices have full rank i.e., rank($\mathbf{A}$) = n.

Orthogonal matrices

- A matrix $\mathbf{A}$ is said to be orthogonal if it is square and its columns are **orthonormal**.
- Formally, $\mathbf{A}$ is orthogonal if $\mathbf{A}^T\mathbf{A} = \mathbf{I}_n = \mathbf{A}\mathbf{A}^T$
- $\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \mathbf{a}_{:,1}^T \\ \mathbf{a}_{:,2}^T \\ \vdots \\ \mathbf{a}_{:,n}^T \end{bmatrix} \begin{bmatrix} \mathbf{a}_{:,1} & \mathbf{a}_{:,2} & \cdots & \mathbf{a}_{:,m} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{:,1}^T\mathbf{a}_{:,1} & \mathbf{a}_{:,2}^T\mathbf{a}_{:,2} & \cdots & \mathbf{a}_{:,n}^T\mathbf{a}_{:,n} \\ \mathbf{a}_{:,2}^T\mathbf{a}_{:,1} & \mathbf{a}_{:,2}^T\mathbf{a}_{:,2} & \cdots & \mathbf{a}_{:,2}^T\mathbf{a}_{:,n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{:,n}^T\mathbf{a}_{:,1} & \mathbf{a}_{:,n}^T\mathbf{a}_{:,2} & \cdots & \mathbf{a}_{:,n}^T\mathbf{a}_{:,n} \end{bmatrix}$
- Similarly one can observe that rows of orthogonal matrices are also orthonormal.
- **T/F**: An orthogonal matrix has full rank.

Orthogonal matrices

- **T/F**: If $\mathbf{U}$ is orthogonal, then $\|\mathbf{Ux}\| = \|\mathbf{x}\|$?
- Orthogonal matrices are “rotation” matrices.
- Consider a 2×2 orthogonal matrix with $\det \mathbf{U} = 1$: $\mathbf{U} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$
- Consider another orthogonal matrix with determinant -1:
  $\mathbf{V} = \begin{bmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Eigenvalues and Eigenvectors

- A square matrix $\mathsf{A} \in \mathbb{R}^{n \times n}$ has *eigenvalue* $\lambda \in \mathbb{C}$ if there exists a **non-zero** vector $\mathsf{x} \in \mathbb{C}^n$ called as the *eigenvector* such that $\mathsf{Ax} = \lambda \mathsf{x}$. Eigenvalues and eigenvectors can be complex in general.
- The eigenvalues of A are solutions to $\det(\mathsf{A} - \lambda \mathsf{I}) = 0$. Thus, there are n eigenvalues of matrix A.
- WLOG, we consider only eigenvectors which have unit-norm.
- T/F: Let A be a 3×3 (real) matrix. Can it have 2 real eigenvalues and 1 complex eigenvalue?
- $\det(\mathsf{A}) = \lambda_1 \lambda_2 \cdots \lambda_n$.

Eigendecomposition

- Some matrices (diagonalizable) can be decomposed as $\mathsf{A} = \mathsf{P}\Lambda\mathsf{P}^{-1}$ where P is (typically) a matrix whose columns form the eigenvectors and $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix which contains the eigenvalues.
- A matrix A is called symmetric if $\mathsf{A} = \mathsf{A}^T$.
- Spectral theorem says that a symmetric matrix A has n real eigenvalues and all the eigenvectors corresponding to these eigenvalues are orthogonal.
- Thus, symmetric matrices have an orthogonal eigendecomposition i.e., $\mathsf{A} = \mathsf{U}\Lambda\mathsf{U}^T$ where U is an orthogonal matrix.
- Note that rank(A) is equal to non-zero eigenvalues when A is symmetric, but not necessarily otherwise.

Positive (Semi-) Definite Matrices

- A **symmetric** matrix $\mathbf{A}(=\mathbf{A}^T) \in \mathbb{R}^{n\times n}$ is said to be *positive semi-definite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Notation: $\mathbf{A} \succeq 0$
- A **symmetric** matrix $\mathbf{A}(=\mathbf{A}^T) \in \mathbb{R}^{n\times n}$ is said to be *positive definite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all **non-zero** $\mathbf{x}$ i.e., $\mathbf{x} \neq 0, \mathbf{x} \in \mathbb{R}^n$. Notation: $\mathbf{A} \succ 0$
- A matrix $\mathbf{A}$ is called negative (semi-) definite if $-\mathbf{A}$ is positive (semi-) definite.
- Notation:
  - Negative definite $\mathbf{A} \prec 0$
  - Negative semi-definite $\mathbf{A} \preceq 0$.
- Positive semi-definite matrices have non-negative eigenvalues:
  $\mathbf{A} \succeq 0 \iff \lambda_i \geq 0 \quad \forall\, 1 \leq i \leq n$
- Positive definite matrices have positive eigenvalues:
  $\mathbf{A} \succ 0 \iff \lambda_i > 0 \quad \forall\, 1 \leq i \leq n$

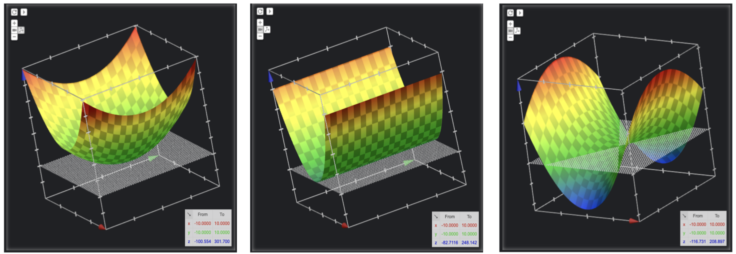
Matrix Calculus

- Let $f(\mathsf{x}) = \mathsf{x}^T\mathsf{A}\mathsf{x}$ where $\mathsf{A} \in \mathbb{R}^{n\times n}$be a scalar valued function. Verify that $f(\mathsf{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$. Consider the following choices for $\mathsf{A} \in \mathbb{R}^{2\times 2}$:
  1. $\mathsf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$
  2. $\mathsf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$
  3. $\mathsf{A} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$

  Depict the function pictorially by sketching the surface $\{(x_1, x_2, f([x_1\ x_2]^T)| -2 \le x_1, x_2 \le 2\}$.

- Let $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ where $\mathbf{A} \in \mathbb{R}^{n\times n}$be a scalar valued function. Verify that $f(\mathbf{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$. Consider the following choices for $\mathbf{A} \in \mathbb{R}^{2\times 2}$:
  1. $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$
  2. $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$
  3. $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$

  Depict the function pictorially by sketching the surface $\{(x_1, x_2, f([x_1\ x_2]^T)| -2 \le x_1, x_2 \le 2\}$.



How to sketch when A is not diagonal?

- Let $f(\mathsf{x}) = \mathsf{x}^T \mathsf{A} \mathsf{x}$ where $\mathsf{A} \succeq 0$.
- $f(\mathsf{x}) = \mathsf{x}^T \, \mathsf{U} \Lambda \mathsf{U}^T \, \mathsf{x} = (\mathsf{U}^T \mathsf{x})^T \, \Lambda \, \mathsf{U}^T \mathsf{x}$.
- Let $\mathsf{y} = \mathsf{U}^T \mathsf{x}$ so that $f(\mathsf{x}) = \mathsf{y}^T \Lambda \mathsf{y}$
- Sketch f in the coordinates of y (previous page).
- Note that $\mathsf{x} = \mathsf{U}\mathsf{y}$ i.e., x is obtained by ”rotating” y.
- Rotate the coordinates to get the sketch in terms of coordinates of x.

Gradient

- Generalization of first derivative.
- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a scalar valued function. Then the gradient of f is a function $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ such that $\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$
- Then the gradient of f at $\mathbf{x}_0$ is given by $\nabla f(\mathbf{x}_0) = \begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_{\mathbf{x}=\mathbf{x}_0} \\ \vdots \\ \left.\frac{\partial f}{\partial x_n}\right|_{\mathbf{x}=\mathbf{x}_0} \end{bmatrix}$
- **Exercise**: $f(\mathbf{z}) = \mathbf{b}^T\mathbf{z}$. What is the gradient $\nabla_{\mathbf{z}} f$ evaluated at $[2\ 1]^T$ for $\mathbf{b} = [-1\ 2]^T$?

Gradient

- **Exercise**: $\nabla f(\mathbf{x})$ when $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is symmetric?

Hessian

- Generalization of the second derivative
- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a scalar valued function. Then the hessian of f is a function $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ such that

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- **Exercise**: $\nabla^2 f(\mathbf{x})$ when $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is symmetric?

Hessian

- **Exercise**: $\nabla^2 f(\mathbf{x})$ when $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$?

- Helpful resource: http://www.matrixcalculus.org/

Blank Pages

Note to self: Stop recording.

Blank Pages