EE445 Mod3-Lec1: Spectral Properties of Matrices

References:

- [CE-OptMod]: Chapter 3.3, 4, 5

Outline

1. Review Eigenvalues & Eigenvectors
2. Symmetric Matrices
3. **Mod3-L2**: Introduction to Singular values and SVD

Why are Spectral Properties Important in ML+OPT?

- Computational efficiency
- Analysis
- Dimensionality reduction
- Numerical stability

How will we see it used?

1. Kernel methods
2. Principle component analysis (unsupervised ML)
3. Principle component regression
4. (time permitting) spectral clustering

Reminder: Eigenvalues & Eigenvectors

Some basics:

- **Definition**: The polynomial $p(\lambda) = \det(A - \lambda I)$ is called the characteristic polynomial of A. The roots of $p(\lambda) = 0$ are the eigenvalues of A.
- **Definition**: A nonzero vector x satisfying $Ax = \lambda x$ is a (right) eigenvector for the eigenvalue λ. A nonzero vector y such that $y^* A = \lambda y^*$ is a (left) eigenvector for the eigenvalue λ.
- Recall that $y^* = (\bar{y})^\top$.
- let $\{x_1, \ldots, x_n\}$ be the eigenvectors of A
  - Orthogonal eigenvectors: $\langle x_i, x_j \rangle = x_i^\top x_j = 0$, $i \neq j$
  - Orthonormal eigenvectors: $\langle x_i, x_j \rangle = x_i^\top x_j = 0$, $i \neq j$ and $\|x_i\| = 1$ for all i

Reminder: Eigenvalues & Eigenvectors

Why important?

- Many ML algorithms involve transforming the matrix A into simpler, or *canonical forms*, from which it is easy to compute its eigenvalues and eigenvectors.
- These transformations are called **similarity transformations**

Similarity transforms

- **Definition**: Let S be any nonsingular matrix. Then A and $B = S^{-1}AS$ are called **similar matrices**, and S is a **similarity transformation**.
- **Proposition.** Similar matrices A and B has the same eigenvalues. Moreover,

$$x \text{ is a right eigenvector of } A \iff S^{-1}x \text{ is a right eigenvector of } B$$

$$y \text{ is a left eigenvector of } A \iff S^{*}y \text{ is a left eigenvector of } B$$

- Some special matrices are similar to diagonal matrices—i.e., for some matrices A, there is a similarity transform S such that $\Lambda = S^{-1}AS$ is diagonal, and Λ contains the eigenvalues of A.
- These matrices are called **diagonalizable**.

Part 2. Special Matrices [Symmetric and Positive (semi) definite]

Symmetric Matrices

Symmetric Matrix: The matrix $A \in \mathbb{R}^{n\times n}$ is symmetric if $A^\top = A$

- Symmetric matrices are one of the most important matrices in linear algebra and ML
- Mod2-L4: we often use kernel matrices $K = [K(x^{(i)}, x^{(j)})]$ and these are symmetric—i.e., $K = K^\top$—since $K(x^{(i)}, x^{(j)}) = K(x^{(j)}, x^{(i)})$
- Mod2-L2: Gram matrices $A^\top A$ and $AA^\top$ are symmetric;
  - in fact we can study all kinds of properties of a matrix A such as the range and null spaces using these gram matrices (cf. Finite Rank Operator Lemma)

Symmetric Matrices: Examples

1. The graph Laplacian is a symmetric matrix

$$L_{ij} = \begin{cases} \#\text{ of edges incident to node } i, & \text{if } i = j \\ -1, & \text{if there is an edge } (i,j) \\ 0, & \text{otherwise} \end{cases}$$

2. Sample covariance matrix

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - \bar{x})(x^{(i)} - \bar{x})^\top, \quad \text{where } \; \bar{x} = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$$

3. Hessian of a function: $H = \nabla^2 f(x)$, where $H_{ij} = \frac{\partial^2}{\partial x_i \partial x_j} f(x)$

Quadratic Functions

- Symmetric matrices play an important role not just in ML but also OPT
- We have seen how to formulate least squares regression as a optimization problem with a quadratic objective:

$$\|Ax - b\|_2^2 = (Ax - b)^\top (Ax - b)$$

- A quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ is a second-order multivariate polynomial in x, that is a function containing a linear combination of all possible monomials of degree at most two—i.e.,

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j + \sum_{i=1}^{n} c_i x_i + d \iff f(x) = x^\top A x + c^\top x + d$$

- Least squares:

$$x^\top A x - 2b^\top A x + b^\top b$$

Quadratic Functions

- using properties of symmetric matrices, we can express any quadratic function as a quadratic form.
- Fact: $x^\top Ax$ is a scalar so that

$$x^\top Ax = x^\top A^\top x \implies x^\top Ax = \frac{1}{2}x^\top \underbrace{(A + A^\top)}_{=:H} x$$

- Aside:

$$A = \underbrace{\frac{1}{2}(A + A^\top)}_{\text{symmetric part}} + \underbrace{\frac{1}{2}(A - A^\top)}_{\text{antisymmetric part}}$$

- Hence, we have

$$f(x) = \frac{1}{2}x^\top Hx + c^\top x + d = \frac{1}{2}\begin{bmatrix} x \\ 1 \end{bmatrix}^\top \begin{bmatrix} H & c \\ c^\top & 2d \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}$$

Symmetric Matrices: Eigendecomposition (Spectral Theorem)

- Every symmetric matrix A can be **diagonalized** as $A = V\Lambda V^\top$ with V formed by the orthonormal eigenvectors of A and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ a diagonal matrix of the eigenvalues of A

$$A = \underbrace{\begin{bmatrix} | & \cdots & | \\ v_1 & \cdots & v_n \\ | & \cdots & | \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \lambda_{n-1} & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} - & v_1^\top & - \\ \vdots & \vdots & \vdots \\ - & v_n^\top & - \end{bmatrix}}_{V^\top}$$

  or equivalently, $A = \lambda_1 v_1 v_1^\top + \cdots + \lambda_n v_n v_n^\top$ (i.e., weighted sum of dyads)
- Additionally, $VV^\top = V^\top V = I$
- This factorization property and the fact that S has n orthogonal eigenvectors are two **important properties** for a symmetric matrix.

Example Problem: Eigenvalues are Real

Problem: Consider a symmetric matrix A. Show that the eigenvalues of A are real.
Solution. Consider $Ax = \lambda x$ with $x \neq 0$. Recall that

$$\langle\, x, x \,\rangle = x^* x = (\bar{x})^\top x$$

Then,

$$\lambda \langle\, x, x \,\rangle = (\bar{x})^\top (\lambda x) = (\bar{x})^\top A x = (A^\top \bar{x})^\top x = (A\bar{x})^\top x = (\bar{A}\bar{x})^\top x = (\bar{\lambda}\bar{x})^\top x = \bar{\lambda} \langle\, x, x \,\rangle$$

so that $\lambda = \bar{\lambda}$.

Example Problem: Orthogonality of Eigenvectors

Problem: Consider a symmetric matrix A. Show that eigenvectors corresponding to distinct eigenvalues are orthogonal.
Solution. Indeed consider eigenvalue-eigenvector pair (λ, x) and eigenvalue-eigenvector pair (μ, z) with $\mu \neq \lambda$. Then,

$$\lambda \langle x, z \rangle = \langle Ax, z \rangle = (Ax)^\top z = x^\top A^\top z = x^\top Az = \mu \langle x, z \rangle \implies \langle x, z \rangle = x^\top z = 0$$

Matrix powers with eigendecomposition

- Recall from **Mod1** we saw many applications with matrix powers such as computing the number of paths of length k in a graph
- For symmetric matrices, computing matrix powers is easy
- Indeed, $A = A^\top$ has orthonormal eigendecomposition $A = V\Lambda V^\top$ so that

$$A^k = \underbrace{(V\Lambda V^\top)\cdots(V\Lambda V^\top)}_{k \text{ times}} = V\Lambda^k V^\top \quad \text{since } V^\top V = I$$

Positive Definite Matrices

- Another important class of matrices are positive definite matrices
- The matrix A is **positive definite** if $\langle Ax, x \rangle > 0$; sometimes we write $A \succ 0$
- And, A is **positive semidefinite (PSD)** if $\langle Ax, x \rangle \geq 0$; sometimes we write $A \succeq 0$
- Positive definite matrices need not be symmetric, but often we are interested in positive definite symmetric matrices
- Eigenvalues: let $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ be the order set of eigenvalues of $A = A^\top$

$$A \succeq 0 \iff \lambda_i(A) \geq 0, \ \forall i \in \{1, \ldots, n\}$$

$$A \succ 0 \iff \lambda_i(A) > 0, \ \forall i \in \{1, \ldots, n\}$$

Example Problem

- Show that $A \succeq 0 \iff \lambda_i(A) \geq 0,\ \forall i \in \{1, \ldots, n\}$
- solution: A is symmetric and hence it can be diagonalized. Let $A = V\Lambda V^\top$ so that

$$x^\top A x = \underbrace{x^\top V}_{=:z^\top} \Lambda V^\top = z^\top \Lambda z = \sum_{i=1}^{n} \lambda_i(A) z_i^2$$

Hence

$$x^\top A x \geq 0\ \forall x \in \mathbb{R}^n \iff z^\top \Lambda z \geq 0\ \forall z \in \mathbb{R}^n$$

And, the right hand side of the equivalence above is itself equivalent to $\lambda_i(A) \geq 0$ for all $i \in \{1, \ldots, n\}$

Examples of PSD Matrices from ML and OPT

- **Mod2-L4**: we often use kernel matrices $K = [K(x^{(i)}, x^{(j)})]$ and these are symmetric and in general PSD
- **Mod2-L2**: Gram matrices $A^\top A$ and $AA^\top$ are PSD

Problem: Show that Gram and Kernel matrices are PSD.
solution. Consider the general case of a kernel matrix:
$G_{ij} = K(x^{(i)}, x^{(j)}) = \langle\, \phi(x^{(i)}), \phi(x^{(j)})\,\rangle$ for i, j. THen for any vector v we have

$$v^\top G v = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i v_j G_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i v_j \,\langle\, \phi(x^{(i)}), \phi(x^{(j)})\,\rangle$$
$$= \left\langle \sum_{i=1}^{n} v_i \phi(x^{(i)}), \sum_{j=1}^{n} v_j \phi(x^{(j)}) \right\rangle = \left\| \sum_{i=1}^{n} v_i \phi(x^{(i)}) \right\|^2 \geq 0$$

Example Problem

Problem. Show that a matrix A is PSD if and only if $A = B^\top B$ for some real matrix B.
Solution.
$(\Longrightarrow)$: Now suppose A is PSD. Let $AV = V\Lambda$ be the eigendecomposition of A and set $B = \sqrt{\Lambda}V^\top$ where $\sqrt{\Lambda} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$. The matrix exists since the eigenvalues are non-negative. Hence

$$B^\top B = V\sqrt{\Lambda}\sqrt{\Lambda}V^\top = V\Lambda V^\top = AVV^\top = A.$$

$(\Longleftarrow)$: Suppose $A = B^\top B$ so that for any vector v we have

$$v^\top A v = v^\top B^\top B v = \|Bv\|^2 \geq 0 \implies A \text{ is PSD}$$