CSE 840: Computational Foundations of Artificial Intelligence February 12, 2025

Symmetric Matrices, Spectral Theorem for Symmetric Matrices,
Positive Definite Matrices, Variational Characterization of Eigenvalues

Instructor: Vishnu Boddeti *Scribe: Auden Garrard and Roshan Atluri, Yuyuan Tian*

1 Introduction

1.1 Symmetric Matrices

Definition: A matrix $A \in \mathbb{R}^{n \times n}$ is called *symmetric* if $A = A^\top$. A matrix $A \in \mathbb{C}^{n \times n}$ is called *Hermitian* if $A = \overline{A}^\top$.

Addition Info: **Examples**

$$A = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 7 \\ 0 & 7 & 9 \end{bmatrix}$$

Proposition: Let $A \in \mathbb{C}^{n \times n}$ be Hermitian. Then all eigenvalues of A are real-valued. Eigenvectors that correspond to distinct eigenvalues are orthogonal.

Proof: Let λ be an eigenvalue of A with eigenvector x. Then

$$Ax = \lambda x$$

$$\lambda \langle x, x \rangle = \langle \lambda x, x \rangle = \langle Ax, x \rangle$$

Since A is Hermitian,

$$\langle Ax, x \rangle = \langle x, Ax \rangle = \langle x, \lambda x \rangle = \overline{\lambda} \langle x, x \rangle$$

$$\Rightarrow \lambda \langle x, x \rangle = \overline{\lambda} \langle x, x \rangle$$

$$\Rightarrow \lambda = \overline{\lambda} \in \mathbb{R} \quad \text{(unless } x = 0 \text{ vector)}$$

$$\Rightarrow \lambda \text{ is real.}$$

Suppose (λ_1, x_1) and (λ_2, x_2) are eigenvalue-eigenvector pairs of A. Then

$$\lambda_1 \langle x_1, x_2 \rangle = \langle \lambda_1 x_1, x_2 \rangle = \langle A x_1, x_2 \rangle = \langle x_1, A x_2 \rangle$$

$$= \langle x_1, \lambda_2 x_2 \rangle = \overline{\lambda_2} \langle x_1, x_2 \rangle$$

Since $\lambda_2 = \overline{\lambda_2}$ (from Hermitian property),

$$\Rightarrow \lambda_1 \langle x_1, x_2 \rangle = \lambda_2 \langle x_1, x_2 \rangle$$

$$
\begin{gathered}
0 = \lambda_1 \langle x_1, x_2 \rangle - \lambda_2 \langle x_1, x_2 \rangle \\
0 = (\lambda_1 - \lambda_2) \langle x_1, x_2 \rangle \\
\Rightarrow \text{either } \lambda_1 = \lambda_2 \text{ or if } \lambda_1 \neq \lambda_2 \text{ then } \langle x_1, x_2 \rangle = 0 \\
\Rightarrow x_1 \perp x_2
\end{gathered}
$$

Definition: An operator $T \in \mathcal{L}(V)$ on a pre-Hilbert space V is called *self-adjoint* if

$$\langle Tu, w \rangle = \langle u, Tw \rangle$$

for all $u, w \in V$.

Sometimes it is called a *Hermitian operator* (on $\mathbb{C}^n$) or a *Symmetric operator* (on $\mathbb{R}^n$).

Remark: Over $\mathbb{C}^n$, self-adjoint operators are represented by Hermitian matrices. On $\mathbb{R}^n$, a self-adjoint operator is represented by a symmetric matrix.

Proposition: Let $T \in \mathcal{L}(V)$ be self-adjoint. Then T has at least one eigenvalue, and it is real-valued. (This holds on both $\mathbb{C}^n$ and $\mathbb{R}^n$.)

Proof (sketch): Let $n := \dim V$. Choose $v \neq 0$, and consider the set of vectors

$$v,\ Tv,\ T^2 v,\ \ldots,\ T^n v$$

These vectors must be linearly dependent (since we have $n+1$ vectors in an n-dimensional space).

So there exist scalars $a_0, a_1, \ldots, a_n$ such that

$$a_0 v + a_1 Tv + \cdots + a_n T^n v = 0$$

Now consider the polynomial with these coefficients:

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = 0$$

This polynomial can be factored as:

$$\underbrace{C(x^2 + b_1 x + c_1) \cdots (x^2 + b_n x + c_n)}_{\text{Quadratic terms}} \underbrace{(x - \lambda_1) \cdots (x - \lambda_m)}_{\text{linear terms}}$$

where the quadratic terms represent irreducible factors over $\mathbb{R}$ (if any), and the linear terms correspond to real eigenvalues $\lambda_1, \ldots, \lambda_m$.

Replace x by T in the polynomial expression:

$$0 = a_0 v + a_1 Tv + \cdots + a_n T^n v = \left(C \underbrace{(\cdots)}_{\text{quadratic}} \underbrace{(\cdots)}_{\text{linear}} \right) (T) v$$

Now we can show: the quadratic terms are invertible, and we are left with (at least one) linear factor:

$$0 = (T - \lambda_1 I) \cdots (T - \lambda_m I) v$$

There must exist at least one index i such that $(T - \lambda_i I)$ is not invertible.

So,

$$(T - \lambda_i I)v = 0 \quad \Rightarrow \quad Tv = \lambda_i v$$

$$\Rightarrow \lambda_i \text{ is an eigenvalue of } T.$$

Addition Info: **Proof that Symmetric Matrices Have Orthogonal Eigenvectors**

Consider a symmetric matrix $A \in \mathbb{R}^{n \times n}$, and let λ_1 and λ_2 be distinct eigenvalues of A with corresponding eigenvectors $\vec{v}_1$ and $\vec{v}_2$, respectively. We aim to show that $\vec{v}_1$ and $\vec{v}_2$ are orthogonal.

From the definition of eigenvectors and eigenvalues, we have:

$$A\vec{v}_1 = \lambda_1 \vec{v}_1, \quad A\vec{v}_2 = \lambda_2 \vec{v}_2.$$

Multiplying both sides of the first equation on the left by $\vec{v}_2^T$ and both sides of the second equation on the left by $\vec{v}_1^T$, we get:

$$\vec{v}_2^T A\vec{v}_1 = \lambda_1 \vec{v}_2^T \vec{v}_1, \quad \vec{v}_1^T A\vec{v}_2 = \lambda_2 \vec{v}_1^T \vec{v}_2.$$

Notice that each of these expressions is a scalar. Therefore,

$$\vec{v}_1^T A\vec{v}_2 = (\vec{v}_1^T A\vec{v}_2)^T = \vec{v}_2^T A^T \vec{v}_1 = \vec{v}_2^T A\vec{v}_1,$$

where the last equality follows from the fact that A is symmetric, i.e., $A = A^T$.

Equating the right-hand sides of the two expressions:

$$\lambda_1 \vec{v}_2^T \vec{v}_1 = \lambda_2 \vec{v}_1^T \vec{v}_2.$$

Since $\lambda_1 \neq \lambda_2$, it follows that

$$\vec{v}_2^T \vec{v}_1 = \vec{v}_1^T \vec{v}_2 = 0,$$

demonstrating that $\vec{v}_1$ and $\vec{v}_2$ are orthogonal.

1.2 Spectral Theorem for Symmetric/Hermitian Matrices

Theorem: A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is orthogonally diagonalizable: there exists an orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ and a diagonal matrix $D \in \mathbb{R}^{n \times n}$ such that

$$A = QDQ^\top$$

where

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

and

$$A = \sum_{i=1}^{n} \lambda_i q_i q_i^\top$$

where each $q_i q_i^\top$ is a rank-1 matrix.

Theorem: A Hermitian matrix $A \in \mathbb{C}^{n\times n}$ is unitarily diagonalizable: there exists a unitary matrix U and a diagonal matrix D such that

$$A = UD\overline{U}^\top$$

and the entries of D are real-valued.

Addition Info:

Proof that Hermitian Matrices are Unitarily Diagonalizable

Let $u_1, u_2, \ldots, u_n$ be an orthonormal basis of eigenvectors, and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the corresponding eigenvalues. Define U to be the matrix with u_k as the k^{th} column, and let Λ be the diagonal matrix with λ_k as the k^{th} diagonal entry.

To show that U is unitary, consider the (i,j)-entry of UU^*. This entry is given by the inner product $\langle u_i, u_j \rangle$, which equals 1 when $i = j$ and 0 otherwise, since the eigenvectors are orthonormal. Thus,

$$UU^* = I.$$

Taking the conjugate transpose of both sides gives

$$U^*U = (UU^*)^* = I,$$

so we also have

$$U^{-1} = U^*,$$

and hence U is unitary.

Now, we prove that $A = U\Lambda U^*$. Consider the effect of $U\Lambda U^*$ on an eigenvector $v_k = u_k$. We compute:

$$U\Lambda U^* v_k = U\Lambda e_k = U\lambda_k e_k = \lambda_k U e_k = \lambda_k v_k = Av_k.$$

Since $\{v_1, v_2, \ldots, v_n\}$ forms a basis for $\mathbb{C}^n$, every vector $x \in \mathbb{C}^n$ can be written as a linear combination of the v_k. Therefore,

$$U\Lambda U^* x = Ax \quad \text{for all } x \in \mathbb{C}^n.$$

It follows that

$$A = U\Lambda U^*.$$

1.3 Positive Definite Matrices

Definition: A matrix $A \in \mathbb{R}^{n\times n}$ is called *positive definite* (PD) if for all $x \in \mathbb{R}^n$, $x \neq 0$,

$$x^\top Ax > 0$$

For *positive semi-definite* (PSD) matrices, $\forall x \in \mathbb{R}^n$, $x \neq 0$,

$$x^\top Ax \geq 0$$

Definition: A matrix $A \in \mathbb{C}^{n\times n}$ is called a *Gram matrix* if there exists a set of vectors $v_1, \dots, v_n \in \mathbb{C}^n$ such that

$$a_{ij} = \langle v_i, v_j \rangle$$

Note: Gram matrices are Hermitian (and similarly, on $\mathbb{R}^{n\times n}$, Gram matrices are symmetric).

Let $V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$, then

$$G = V^\top V, \quad C = V\overline{V}^\top$$

CAUTION: Over $\mathbb{C}$, we have that positive definite (PD) $\Rightarrow$ self-adjoint.

However, over $\mathbb{R}$, this is **not** true!

$\Rightarrow$ There are matrices which are PD but not symmetric.

Example:

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

$$x^\top A x = x_1^2 + x_2^2 > 0 \quad \text{for all } x \neq 0$$

$$\Rightarrow A \text{ is PD but not symmetric.}$$

However, over $\mathbb{C}$, the same matrix is *not* PD, since $x_1^2 + x_2^2$ can be negative (not necessarily positive definite).

Theorem: Let $A \in \mathbb{C}^{n\times n}$ be Hermitian. Then the following are equivalent:

(i) A is positive semi-definite (PD), i.e., $x^* A x \geq 0$ for all $x \in \mathbb{C}^n$.

(ii) All eigenvalues of A are ≥ 0 (> 0).

(iii) The mapping $\langle \cdot, \cdot \rangle_A : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ defined by

$$\langle x, y \rangle_A := \overline{y}^\top A x$$

satisfies all properties of an inner product *except* one: if $\langle x, x \rangle_A = 0$, this does not imply $x = 0$. (This mapping is an inner product only on a subspace.)

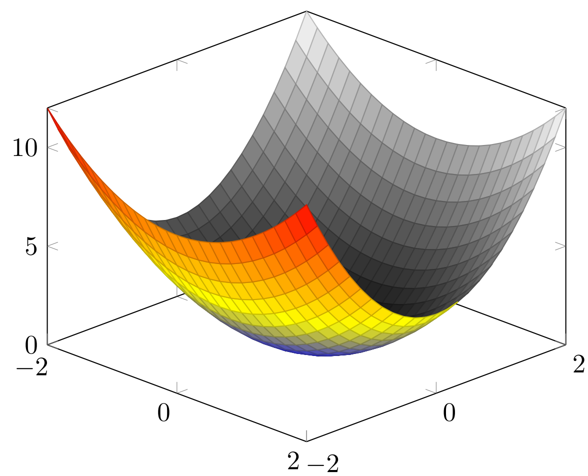
(iv) A is a Gram matrix of n vectors which are not necessarily linearly independent, i.e., (which are linearly independent)

$$a_{ij} = \langle x_i, x_j \rangle$$

where $x_1, \dots, x_n \in \mathbb{C}^n$.

Addition Info:

Quadratic Form Visualization



The function above plotted is

$$f(x, y) = x^2 + 2y^2$$

which comes from the quadratic form:

$$x^T Ax = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Expanding this:

$$x^2 + 2y^2.$$

Since the function is always non-negative and only equals zero at (0,0), this confirms that A is positive definite because all its eigenvalues are strictly positive. Geometrically, this corresponds to a paraboloid that always opens upwards.

Additionally, the above statement indicates that if one of the eigenvalues were negative, this would create a saddle point, breaking one of the passive variables. Finally, since all of the eigenvalues are strictly positive, this guarantees that A is positive definite, never producing negative values.

1.4 Roots of Positive Semi-Definite Matrices

Theorem: Let $A \in \mathbb{R}^{n\times n}$ be symmetric and positive semi-definite (PSD). Then there exists a matrix $B \in \mathbb{R}^{n\times n}$, also PSD, such that

$$A = B^2$$

The matrix B is called the *square root* of A, denoted as

$$B = A^{1/2}$$

Proof: By the spectral theorem,

$$A = UDU^\top$$

where U is orthogonal and D is a diagonal matrix with non-negative eigenvalues:

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}, \quad \lambda_i \geq 0$$

Define

$$\sqrt{D} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix}$$

Then set

$$B := U\sqrt{D}U^\top$$

Addition Info:

Example

Consider the positive semi-definite matrix:

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$$

The eigenvalues of A are 4 and 9, both non-negative. The square root of A is given by:

$$B = \sqrt{A} = \begin{bmatrix} \sqrt{4} & 0 \\ 0 & \sqrt{9} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

which satisfies:

$$B^2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}^2 = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} = A.$$

1.5 Variational Characterization of Eigenvalues

Definition: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The Rayleigh quotient R_A by

$$R_A : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}, \quad x \mapsto \frac{x^\top Ax}{x^\top x}$$

This is called the *Rayleigh coefficient* of A.

Addition Info:

Example Let

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Then

$$R_A(x) = \frac{x^\top Ax}{x^\top x} = \frac{[1\ 0]\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}}{[1\ 0]\begin{bmatrix} 1 \\ 0 \end{bmatrix}} = \frac{[1\ 0]\begin{bmatrix} 2 \\ 1 \end{bmatrix}}{1} = \frac{2}{1} = 2$$

Proposition: Let A be symmetric, and let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of A with corresponding eigenvectors $v_1, \ldots, v_n$.

Then:

$$\min_{x \in \mathbb{R}^n,\ \|x\|=1} R_A(x) = \min_{\|x\|=1} x^\top A x = \lambda_1, \quad \text{attained at } x = v_1$$

$$\max_{x \in \mathbb{R}^n,\ \|x\|=1} R_A(x) = \max_{\|x\|=1} x^\top A x = \lambda_n, \quad \text{attained at } x = v_n$$

Intuition: Assume A is expressed in terms of the orthonormal basis $v_1, \ldots, v_n$, so that

$$A = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

Let y be a vector, also represented in the same basis:

$$y = y_1 v_1 + y_2 v_2 + \cdots + y_n v_n$$

Then,

$$y^\top A y = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

Among the standard basis vectors:

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

the smallest result of $y^\top A y$ is given by choosing

$$y = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

This corresponds to v_1, and the value of $y^\top A y$ would be λ_1.

Proof (sketch): Assume we start with the standard basis. Let

$$Q = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}$$

be the basis transformation matrix. Since Q is orthogonal, we have

$$A = Q^\top \Lambda Q$$

where Λ is diagonal.

For a vector $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ be a vector in the original basis, and define $y := Q^\top x$.

We consider the Rayleigh quotient:

$$R_A(y) = \frac{y^\top A y}{y^\top y} = \frac{(Q^\top x)^\top A (Q^\top x)}{(Q^\top x)^\top (Q^\top x)}$$

Since $(Q^\top x)^\top = x^\top Q$ and Q is orthogonal (so $Q^\top Q = I$), this becomes:

$$= \frac{x^\top Q\, Q^\top \Lambda Q\, Q^\top x}{x^\top Q Q^\top x} = \frac{x^\top \Lambda x}{x^\top x}$$

$$= \frac{\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2}{\|x\|^2}$$

Hence,

$$\min_{\|y\|=1} R_A(y) = \min_{\|x\|=1} \left(\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2\right)$$

Note: Q is orthogonal, so it preserves norms.

The minimum of $R_A(y)$ is attained when

$$x = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow y = Q^\top x = v_1$$

with value

$$\min_{\|y\|=1} R_A(y) = \lambda_1$$

Proposition: Consider the constrained minimization problem

$$\min_{\substack{\|x\|=1 \\ x \perp v_1}} R(x)$$

The solution to this problem is $x = v_2$, and R(x) $= \lambda_2$

Intuition: Consider the restriction of operator A to the subspace

$$V_1^\perp := (\text{span}\{v_1\})^\perp$$

On this subspace, A is invariant and symmetric, so we can apply the Rayleigh quotient again on this smaller space.

Let

$$V_1^\perp = \text{span}\{v_2, v_3, \ldots, v_n\}$$

If we apply the Rayleigh to $V_1^\perp$, we get the next solution:

$$\lambda_2, \quad v_2$$

Theorem: (Min–Max Theorem)

Let $A \in \mathbb{R}^{n\times n}$ be symmetric with eigenvalues

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$$

Then the k-th eigenvalue satisfies:

$$\begin{aligned}\lambda_k &= \min_{\substack{U\subset\mathbb{R}^n \\ \dim U = k}} \max_{x\in U\setminus\{0\}} R_A(x) \\ &= \max_{\substack{U\subset\mathbb{R}^n \\ \dim U = n-k+1}} \min_{x\in U\setminus\{0\}} R_A(x)\end{aligned}$$

Intuition: For $k = 3$:

- Consider the subspace U spanned by v_1, v_2, v_3. As we saw before,

$$\max_{x\in U} R_A(x) = \lambda_3, \quad \text{attained by } v_3$$

- Consider another subspace U spanned by v_9, v_{10}, v_{11},

$$\max_{x\in U} R_A(x) = \lambda_{11}$$

bibliography

https://rubenvannieuwpoort.nl/posts/the-spectral-theorem-for-hermitian-matrices

https://www.youtube.com/watch?v=OXLalScAMl0