

Chapter 2

Linear algebra

2.1 Eigenvalues and eigenvectors

In this section, we shall concern with square matrices only, unless stated otherwise. As we have seen, the adjacency matrix and incidence matrix of a graph are algebraic structures which contain graphical information. This is the motivation for us to dig deeper into linear algebra. Eigenvalues and eigenvectors say a lot about a matrix, hence we would hope that they tell us something about the graph.

2.1.1 Preliminaries

The *eigenvalues* of a matrix A are the numbers λ such that the equation $(\lambda I - A)x = 0$ has a non-zero solution vector, in which case the solution vector x is the corresponding *eigenvector*. Since $\vec{0}$ is always a solution, it would be unique if $\det(\lambda I - A) \neq 0$. Hence, the eigenvalues are solutions to the equation $\det(\lambda I - A) = 0$, i.e. $p_A(\lambda) = 0$ where $p_A(\lambda)$ is the *characteristic polynomial* of A .

Theorem 2.1. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of an $n \times n$ complex matrix A , then

(i) $\lambda_1 + \dots + \lambda_n = \text{tr } A$ ($:= \sum_i a_{ii}$).

(ii) $\lambda_1 \dots \lambda_n = \det A$.

Proof. Suppose

$$p_A(\lambda) = \prod_i (\lambda - \lambda_i) = \lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_1\lambda + c_0$$

It is easy to see that $c_{n-1} = -(\sum_i a_{ii})$, proving (i). On the other hand, $c_0 = (-1)^n \det A$, proving (ii). $\square$

2.1.2 The diagonal form

Proposition 2.2. Suppose the $n \times n$ matrix A has n linearly independent eigenvectors x_i , $1 \leq i \leq n$. Then if these vectors are columns of matrix S , it follows that $S^{-1}AS = \Lambda$ where Λ is a diagonal matrix with $(\Lambda)_{ii} = \lambda_i$.

Proof. It's easy to see that $AS = S\Lambda$. $\square$

Proposition 2.3. If $x_1, \dots, x_k$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_k$, then $x_1, \dots, x_k$ are linearly independent.

Proof. When $k = 2$, suppose $c_1 x_1 + c_2 x_2 = 0$. Multiplying by A gives $c_1 \lambda_1 x_1 + c_2 \lambda_2 x_2 = 0$. Subtracting λ_2 times the previous equation we have

$$c_1(\lambda_1 - \lambda_2)x_1 = 0$$

Hence, $c_1 = 0$ since $\lambda_1 \neq \lambda_2$ and $x_1 \neq 0$. The general case can be done by induction. $\square$

Proposition 2.4. If $\lambda_1, \dots, \lambda_n$ are eigenvalues of A , then $\lambda_1^k, \dots, \lambda_n^k$ are eigenvalues of A^k . If S diagonalizes A , i.e. $S^{-1}AS = \Lambda$, then $S^{-1}A^kS = \Lambda^k$

Proof. Easy. $\square$

2.1.3 Symmetric and Hermitian matrices

We'll be working on $\mathbb{C}$, so let's fix some terminologies. For $x \in \mathbb{C}^n$, let $\bar{x}$ denotes its complex conjugate. The inner product of x and y is defined to be

$$\bar{x}^T y = \bar{x}_1 y_1 + \dots + \bar{x}_n y_n$$

If A is any complex matrix, the *Hermitian transpose* A^* of A is defined to be $\bar{A}^T$. Note that just as in the real matrix case, $(AB)^* = B^* A^*$.

A is said to be *Hermitian* if $A = A^*$. If A contains only real entries, being Hermitian implies being symmetric. Also notice that if A is Hermitian then the diagonal entries are reals.

Lemma 2.5. We have

1. If A is Hermitian, then for all $x \in \mathbb{C}^n$, $x^* Ax$ is real.
2. Every eigenvalue of a Hermitian matrix is real.
3. The eigenvectors of a Hermitian matrix, if come from distinct eigenvalues are orthogonal to one another.

Proof. It is straightforward that

1. $(x^* Ax)^* = x^* A^* x^{**} = x^* Ax$.
2. $Ax = \lambda x$ implies $\lambda = \frac{x^* Ax}{x^* x}$.
3. Suppose $Ax = \lambda_1 x$, $Ay = \lambda_2 y$, $\lambda_1 \neq \lambda_2$ and $A = A^*$.

$$(\lambda_1 x)^* y = (Ax)^* y = x^* Ay = x^* (\lambda_2 y)$$

Hence, $(\lambda_1 - \lambda_2)x^* y = 0$.

$\square$

2.1.4 Orthonormal and unitary matrices

A real matrix Q is said to be *orthogonal* if $Q^T Q = I$. A complex matrix U is *unitary* if $U^* U = I$. In other words, the columns of U (and Q) are orthonormal. Clearly orthogonal is a special case of unitary. We state without proof a simple proposition.

Proposition 2.6. *Let U be a unitary matrix, then*

- (i) $(Ux)^*(Uy) = x^*y$, and $\|Ux\|^2 = \|x\|^2$.
- (ii) Every eigenvalue λ of U has modulus 1 (i.e. $|\lambda| = 1$).
- (iii) Eigenvectors corresponding to distinct eigenvalues of U are orthonormal.
- (iv) If U' is another unitary matrix, then UU' is unitary.

2.1.5 The Spectral theorem and the Jordan form

First, a few definitions. Two matrices A and B are *similar* iff there is an invertible matrix M such that $M^{-1}AM = B$.

Proposition 2.7. *If $B = M^{-1}AM$, then A and B have the same eigenvalues. An eigenvector x of A corresponds to an eigenvector $M^{-1}x$ of B .*

Proof. $Ax = \lambda x$ implies $(M^{-1}A)x = \lambda M^{-1}x$, or $(BM^{-1})x = \lambda(M^{-1}x)$. Another proof would go

$$\det(xI - B) = \det(M^{-1}(xI)M - M^{-1}AM) = \det(M^{-1}(xI - A)M) = \det(xI - A)$$

□

The vector space spanned by all eigenvectors corresponding to a particular eigenvalue λ is called the *eigenspace* associated with λ , or the λ -eigenspace. We shall often use V_λ to denote this space if the underlying matrix is clear from context.

Corollary 2.8. *If A and B are similar, then the corresponding eigenspaces of A and B have the same dimension.*

Proof. Suppose $B = M^{-1}AM$, then the mapping $\phi : x \rightarrow M^{-1}x$ is an invertible linear transformation from one eigenspace of A to the corresponding eigenspace of B . □

Lemma 2.9 (Schur's lemma). *For any $n \times n$ matrix A , there is a unitary matrix U such that $U^{-1}AU$ is upper triangular.*

Proof. Note that the eigenvalues of A are then lying on the diagonal of $U^{-1}AU$.

We show this by induction on n . The lemma holds when $n = 1$. When $n > 1$, over $\mathbb{C}$ A must have at least one eigenvalue λ_1 . Let x_1 be a corresponding eigenvector. Use *Gram-Schmidt* process to get a set of orthonormal vectors $\{x_1, x_2, \dots, x_n\}$. Let U_1 be the matrix whose columns are this set of vectors in order. Clearly,

$$AU_1 = U_1 \begin{bmatrix} \lambda_1 & * & * & \dots & * \\ 0 & * & * & \dots & * \\ 0 & * & * & \dots & * \\ \dots & \dots & \dots & \dots & \dots \\ 0 & * & * & \dots & * \end{bmatrix}, \text{ so } U_1^{-1}AU_1 = \begin{bmatrix} \lambda_1 & * & * & \dots & * \\ 0 & * & * & \dots & * \\ 0 & * & * & \dots & * \\ \dots & \dots & \dots & \dots & \dots \\ 0 & * & * & \dots & * \end{bmatrix}$$

Now, let $A' = (U_1^{-1}AU_1)_{11}$ (crossing off row 1 and column 1 of $U_1^{-1}AU_1$). Then by induction there exists an $(n-1) \times (n-1)$ unitary matrix M such that $M^{-1}A'M$ is upper triangular. Let U_2 be the $n \times n$ matrix obtained by adding a new row and new column to M with all new entries equal 0 except $(U_2)_{11} = 1$. Clearly U_2 is unitary and $U_2^{-1}(U_1^{-1}AU_1)U_2$ is upper triangular. Letting $U = U_1U_2$ completes the proof. $\square$

The following theorem is one of the most important theorems in linear algebra, besides the Jordan form given next.

Theorem 2.10 (Spectral theorem). *Every real symmetric matrix can be diagonalized by an orthogonal matrix, and every Hermitian matrix can be diagonalized by a unitary matrix:*

$$(\text{real case}) \quad Q^{-1}AQ = \Lambda \quad (\text{complex case}) \quad U^{-1}AU = \Lambda$$

Moreover, in both cases all eigenvalues are real.

Proof. The real case follows from the complex case. Firstly, by Schur's lemma there is a unitary matrix U such that $U^{-1}AU$ is upper triangular. Moreover,

$$(U^{-1}AU)^* = U^*A^*(U^{-1})^* = U^{-1}AU$$

i.e. $U^{-1}AU$ is also Hermitian. But an upper triangular Hermitian matrix must be diagonal. The realness of eigenvalues follow from Lemma 2.5. $\square$

Theorem 2.11 (The Jordan form). *If a matrix A has s linearly independent eigenvectors, then it is similar to a matrix which is in **Jordan form** with s square blocks on the diagonal:*

$$M^{-1}AM = \begin{bmatrix} B_1 & 0 & 0 & \dots & 0 \\ 0 & B_2 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & B_s \end{bmatrix}$$

Each block has exactly one 1-dimensional eigenspace, one eigenvalue, and 1's just above the diagonal:

$$B_j = \begin{bmatrix} \lambda_j & 1 & 0 & \dots & 0 \\ 0 & \lambda_j & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda_j \end{bmatrix}$$

Proof. You could read Appendix B of [132]. The proof is not difficult but tedious. The only thing we need to remember is that this theorem leads to the following nice corollary. I might come back to prove this if needed later on. $\square$

Corollary 2.12. *The following hold*

1. $\text{rank}(A) = \sum_{\lambda_i \neq 0} m(\lambda_i) + m(0) - \dim(V_0)$.
2. If A is Hermitian, then the λ -eigenspace has dimension equal the multiplicity of λ as a solution to equation $p_A(x) = 0$.
3. In fact, in Hermitian case $\mathbb{C}^n = \bigoplus_i V_{\lambda_i}$ where V_{λ_i} denotes the λ_i -eigenspace.

Proof. This follows directly from the *Jordan form* and our observation in Corollary 2.8. We are mostly concerned with the dimensions of eigenspaces, so we can think about Λ instead of A . Similar matrices have the same rank, so A and its Jordan form have the same rank. The Jordan form of A has rank equal the total number of non-zero eigenvalues on the diagonal plus the number of 1's in the Jordan blocks corresponding to the eigenvalue 0, which is exactly $m(0) - \dim(V_0)$.

When A is Hermitian, it is diagonalizable. Every eigenvector corresponding to an *occurrence* of an eigenvalue λ is linearly independent from all others (including the eigenvector corresponding to another instance of the same λ). $\square$

2.1.6 The minimum polynomial

I found the following very nice theorem stated without proof in a book called “Matrix Methods” by Richard Bronson. I’m sure we could find a proof in either [71] or [46], but I wasn’t able to get them from the library. Here I present my little proof.

Theorem 2.13. *Suppose B_k is a Jordan block of size $(l + 1) \times (l + 1)$ corresponding to the eigenvalue λ_k of A , i.e.*

$$B_k = \begin{bmatrix} \lambda_k & 1 & 0 & \dots & 0 \\ 0 & \lambda_k & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \dots & \dots & \dots & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda_k \end{bmatrix}$$

Then, for any polynomial $q(\lambda) \in \mathbb{C}[\lambda]$

$$q(B_k) = \begin{bmatrix} q(\lambda_k) & \frac{q'(\lambda_k)}{1!} & \frac{q''(\lambda_k)}{2!} & \dots & \frac{q^{(l)}(\lambda_k)}{l!} \\ 0 & q(\lambda_k) & \frac{q'(\lambda_k)}{1!} & \dots & \frac{q^{(l-1)}(\lambda_k)}{(l-1)!} \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \dots & \dots & \dots & \dots & \frac{q'(\lambda_k)}{1!} \\ 0 & 0 & 0 & \dots & q(\lambda_k) \end{bmatrix} \quad (2.1)$$

Proof. We only need to consider the case $q(x) = x^j, j \geq 0$, and then extend linearly into all polynomials. The case $j = 0$ is clear. Suppose equation (2.1) holds for $q(x) = x^{j-1}, j \geq 1$. Then, when $q(x) = x^j$ we have

$$\begin{aligned}
q(B_k) &= B_k^{j-1} B_k \\
&= \begin{bmatrix} \lambda_k^{j-1} & \binom{j-1}{1} \lambda_k^{j-2} & \binom{j-1}{2} \lambda_k^{j-3} & \cdots & \binom{j-1}{l} \lambda_k^{j-l-1} \\ 0 & \lambda_k^{j-1} & \binom{j-1}{1} \lambda_k^{j-2} & \cdots & \binom{j-1}{l-1} \lambda_k^{j-l} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \binom{j-1}{1} \lambda_k^{j-2} \\ 0 & 0 & 0 & 0 & \lambda_k^{j-1} \end{bmatrix} \begin{bmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \lambda_k & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_k \end{bmatrix} \\
&= \begin{bmatrix} \lambda_k^j & \binom{j}{1} \lambda_k^{j-1} & \binom{j}{2} \lambda_k^{j-2} & \cdots & \binom{j}{l} \lambda_k^{j-l} \\ 0 & \lambda_k^j & \binom{j}{1} \lambda_k^{j-1} & \cdots & \binom{j}{l-1} \lambda_k^{j-l+1} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \binom{j}{1} \lambda_k^{j-1} \\ 0 & 0 & 0 & 0 & \lambda_k^j \end{bmatrix}
\end{aligned}$$

□

The *minimum polynomial* $m(\lambda)$ of an $n \times n$ matrix A over a complex vector space V is the monic polynomial of lowest degree such that $m(A) = 0$.

Lemma 2.14. *With the terminologies just stated, we have*

1. $m(\lambda)$ divides $p_A(\lambda)$.
2. Every root of $p_A(\lambda)$ is also a root of $m(\lambda)$.
3. A is diagonalizable iff $m_A(\lambda)$ has no multiple roots.
4. If $\{\lambda_i\}_{i=1}^s$ are distinct eigenvalues of a Hermitian matrix A , then $m(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)$.

Proof. 1. $m(\lambda)$ must divide every polynomial $q(\lambda)$ with $q(A) = 0$, since otherwise $q(\lambda) = h(\lambda)m(\lambda) + r(\lambda)$ implies $r(A) = 0$ while $r(\lambda)$ has smaller degree than $m(\lambda)$. On the other hand, by Cayley-Hamilton Theorem, $p_A(A) = 0$.

2. Notice that $Ax = \lambda x$ implies $A^i x = \lambda^i x$. Thus for any eigenvalue λ_k of A with corresponding eigenvector x , $0 = m(A)x = \sum_i c_i A^i x = \sum_i c_i \lambda_k^i x = m(\lambda_k)x$. This implies λ_k is a root of $m(\lambda)$.
3. ($\Rightarrow$). Suppose $M^{-1}AM = \Lambda$ for some invertible matrix M , and $\lambda_1, \dots, \lambda_s$ are distinct eigenvalues of A . By 1 and 2, we only need to show A is a root of $m_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)$. It is easy to see that for any polynomial $q(\lambda)$, $q(A) = Mq(\Lambda)M^{-1}$. In particular, since $p_A(\Lambda) = 0$ we get the desired result.

($\Leftarrow$). Now we assume $m_A(\lambda)$ has no multiple root, which implies $m_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)$. By Proposition 2.2, we shall show that A has n linearly independent eigenvectors. Firstly, notice that if the Jordan form of A is

$$M^{-1}AM = \begin{bmatrix} B_1 & 0 & 0 & \cdots & 0 \\ 0 & B_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & B_s \end{bmatrix}$$

then, for any $q(\lambda) \in \mathbb{C}[\lambda]$ we have

$$\begin{aligned} M^{-1}q(A)M &= q \left(\begin{bmatrix} B_1 & 0 & 0 & \dots & 0 \\ 0 & B_2 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & B_s \end{bmatrix} \right) \\ &= \begin{bmatrix} q(B_1) & 0 & 0 & \dots & 0 \\ 0 & q(B_2) & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & q(B_s) \end{bmatrix} \end{aligned}$$

So, $\prod_{i=1}^s (A - \lambda_i I) = 0$ implies $\prod_{i=1}^s (B_k - \lambda_i I) = 0$ for all $k = 1, \dots, s$. If A does not have n linearly independent eigenvectors, one of the blocks B_k must have size > 1 . Applying Theorem 2.13 with $q(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)$, we see that $q(B_k)$ does not vanish since $q'(\lambda_i) \neq 0, \forall i \in [s]$. Contradiction!

4. Follows from 3 since a Hermitian matrix is diagonalizable. □

2.2 Positive definite matrices

The purpose of this section is to develop several important background facts. The first is the necessary and sufficient conditions for a real symmetric matrix A (or Hermitian in general) to be *positive definite*. This is essentially the conditions for a *quadratic form* on $\mathbb{R}^n$ to have a minimum at some point. The second is the *Sylvester law of inertia*. Another is the *Rayleigh's principle*. Lastly, some *minimax and maximin principles* related to the *Rayleigh's quotient* shall be developed. These principles eventually lead to the theorem about *interlacing of eigenvalues*, which is very useful when our matrix is the adjacency matrix of a graph.

2.2.1 Some analysis

Let us first recall two key theorems from real analysis, stated without proofs. For $f : \mathbb{R}^n \rightarrow \mathbb{R}$, if for some $a \in \mathbb{R}^n$ $\frac{\partial f}{\partial x_i}(a) = 0$, then a is called a *stationary point* of f .

Theorem 2.15 (The second derivative test). Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and its partial derivatives up to and including order 2 are continuous in a ball $B(a, r)$ (centered at $a \in \mathbb{R}^n$, radius r). Suppose that f has a stationary point at a . For $h = (h_1, \dots, h_n)$, define $\Delta f(a, h) = f(a + h) - f(a)$; also define

$$Q(h) = \frac{1}{2!} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(a) h_i h_j$$

then,

1. If $Q(h) > 0$ for $h \neq 0$, then f has a strict local minimum at a .
2. If $Q(h) < 0$ for $h \neq 0$, then f has a strict local maximum at a .

3. If $Q(h)$ has a positive maximum and a negative minimum, then $\Delta f(a, h)$ changes sign in any ball $B(a, \rho)$ such that $\rho < r$.

Note. (3.) says that at any close neighborhood of a , there are some points b and c such that $f(b) > f(a)$ and $f(c) < f(a)$.

Example 2.16. Let us look at a quadratic form $F(x_1, \dots, x_n)$ with all real coefficients, i.e. every term of F has degree at most 2. Let A be the matrix defined by $a_{ij} = \partial^2 F / \partial x_i \partial x_j$. Clearly, A is a real symmetric matrix. For any vector $h \in \mathbb{R}^n$,

$$\begin{aligned} h^T A h &= \begin{bmatrix} h_1 & h_2 & \dots & h_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix} \\ &= \sum_{i,j=1}^n a_{ij} h_i h_j \\ &= 2Q(h) \end{aligned}$$

So, $F(x_1, \dots, x_n)$ has a minimum at $(0, \dots, 0)$ (which is a stationary point of F) iff $h^T A h > 0$ for all $h \neq 0$.

Definition 2.17. A non-singular $n \times n$ Hermitian matrix A is said to be *positive definite* if $x^* A x > 0$ for all non zero vector $x \in \mathbb{C}^n$. A is *positive semidefinite* if we only require $x^* A x \geq 0$. The terms *negative definite* and *negative semidefinite* can be defined similarly.

Note. Continuing with our example, clearly $F(x_1, \dots, x_n)$ has a minimum at $(0, \dots, 0)$ iff A is positive definite. Also, since we already showed that if A is Hermitian, then $x^* A x$ is real, the definitions given above make sense.

A function f is in C^1 on some domain $D \subseteq \mathbb{R}^n$ if f and all its first order derivatives are continuous on D . For $a \in D$, $\nabla f(a) := (\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a))$.

Theorem 2.18 (Lagrange's multiplier rule). Suppose that $f, \varphi_1, \dots, \varphi_k$ are C^1 functions on an open set D in $\mathbb{R}^n$ containing a point a , that the vectors $\nabla \varphi_1(a), \dots, \nabla \varphi_k(a)$ are linearly independent, and that f takes on its minimum among all points of D_0 at x_0 , where D_0 is the subset of D so that for all $x \in D_0$,

$$\varphi_i(x_1, \dots, x_n) = 0, \quad i = 1, \dots, k$$

Then, if $F : \mathbb{R}^{n+k} \rightarrow \mathbb{R}$ is defined to be

$$F(x, \lambda) = f(x) - \sum_{i=1}^k \lambda_i \varphi_i(x)$$

then there exists $\lambda^0 \in \mathbb{R}^k$ such that

$$\begin{aligned} \frac{\partial F}{\partial x_i}(a, \lambda^0) &= 0 \quad i = 1, \dots, n \\ \frac{\partial F}{\partial \lambda_j^0}(a, \lambda^0) &= 0 \quad i = 1, \dots, k \end{aligned}$$

Note. This theorem essentially says that the maxima (or minima) of f subject to the *side conditions* $\varphi_1 = \cdots = \varphi_k = 0$ are among the maxima (or minima) of the function F without any constraints.

Example 2.19. To find the maximum of $f(x) = x_1 + 3x_2 - 2x_3$ on the sphere $14 - (x_1^2 + x_2^2 + x_3^2) = 0$, we let

$$F(x, \lambda) = x_1 + 3x_2 - 2x_3 + \lambda(x_1^2 + x_2^2 + x_3^2 - 14)$$

Then, $\frac{\partial F}{\partial x_1} = 1 + 2\lambda x_1$, $\frac{\partial F}{\partial x_2} = 3 + 2\lambda x_2$, $\frac{\partial F}{\partial x_3} = -2 + 2\lambda x_3$, and $\frac{\partial F}{\partial \lambda} = x_1^2 + x_2^2 + x_3^2 - 14$. Solving $\frac{\partial F}{\partial x_i} = 0$ we obtain two solutions $(x, \lambda) = (1, 3, -2, -1/2)$ and $(-1, -3, 2, 1/2)$. Which of these solutions give a maximum or a minimum? We apply the second derivative test. All second derivatives of F are 0 except $\partial^2 F / \partial x_1^2 = \partial^2 F / \partial x_2^2 = \partial^2 F / \partial x_3^2 = 2\lambda$. $Q(h) = 2\lambda(h_1^2 + h_2^2 + h_3^2)$ has the same sign as λ . Hence, the first solution gives the maximum value of 14, the second solution gives the minimum value of -13 .

2.2.2 Conditions for positive-definiteness

Now we are ready to specify the necessary and sufficient conditions for a Hermitian matrix to be positive definite, or positive semidefinite for that matter.

Theorem 2.20. *Each of the following tests is a necessary and sufficient condition for the real symmetric matrix A to be positive definite.*

- (a) $x^T A x > 0$ for all non-zero vector x .
- (b) All the eigenvalues of A satisfy $\lambda_i > 0$.
- (c) All the upper left submatrices A_k have positive determinants.
- (d) If we apply Gaussian elimination on A without row exchanges, all the pivots satisfy $p > 0$.

Note. (a) and (b) hold for Hermitian matrices also.

Proof. ($a \Rightarrow b$). Suppose x_i is a unit λ_i -eigenvector, then $0 < x_i^T A x_i = x_i^T \lambda_i x_i = \lambda_i$.

($b \Rightarrow a$). Since A is real symmetric, it has a full set of orthonormal eigenvectors $\{x_1, \dots, x_n\}$ by the Spectral theorem. For each $x \in \mathbb{R}^n$, suppose $x = c_1 x_1 + \cdots + c_n x_n$, then

$$Ax = A(c_1 x_1 + \cdots + c_n x_n) = c_1 \lambda_1 x_1 + \cdots + c_n \lambda_n x_n$$

Because the x_i are orthonormal, we get

$$\begin{aligned} x^T A x &= (c_1 x_1^T + \cdots + c_n x_n^T)(c_1 \lambda_1 x_1 + \cdots + c_n \lambda_n x_n) \\ &= \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2 \end{aligned}$$

Thus, every $\lambda_i > 0$ implies $x^T A x > 0$.

($a \Rightarrow c$). We know $\det A = \lambda_1 \cdots \lambda_n > 0$. To prove the same result for all A_k , we look at a non-zero vector x whose last $n - k$ components are 0, then

$$x^T A x = \begin{bmatrix} x_k^T & 0 \end{bmatrix} \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = x_k^T A_k x_k$$

$\det A_k > 0$ follows by induction.

($c \Rightarrow d$). Without row exchanges, the pivot d_k in Gaussian elimination is $\det A_k / \det A_{k-1}$. This can also be proved easily by induction.

($d \Rightarrow a$). Gaussian elimination gives us a LDU factorization of A where all diagonal entries of L and U are 1's. Also, the diagonal entries d_i of D is exactly the i^{th} pivot p_i . The fact that A is symmetric implies $L = U^T$, hence $A = LDL^T$, which gives

$$x^T A x = (x^T L)(D)(L^T x) = d_1(L^T x)_1^2 + d_2(L^T x)_2^2 + \cdots + d_n(L^T x)_n^2$$

Since L is fully ranked, $L^T x \neq 0$ whenever $x \neq 0$. So the pivots $d_i > 0$ implies $x^T A x > 0$ for all non-zero vectors x . $\square$

2.2.3 The Rayleigh's quotient and the variational characterizations

For a Hermitian matrix A , the following is known as the *Rayleigh's quotient* :

$$R(x) = \frac{x^* A x}{x^* x}$$

Theorem 2.21 (Rayleigh-Ritz). Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of a real symmetric matrix A . Then, the quotient $R(x)$ is maximized at any λ_1 -eigenvector $x = x_1$ with maximum value λ_1 . $R(x)$ is minimized at any λ_n -eigenvector $x = x_n$ with minimum value λ_n ,

Proof. Let Q be a matrix whose columns are a set of orthonormal eigenvectors $\{x_1, \dots, x_n\}$ of A corresponding to $\lambda_1, \dots, \lambda_n$, respectively. Writing x as a linear combination of columns of Q : $x = Qy$, then since $Q^T A Q = \Lambda$ we have

$$R(x) = \frac{x^T A x}{x^T x} = \frac{(Q^T y)^T A (Q^T y)}{(Q^T y)^T (Q^T y)} = \frac{y^T \Lambda y}{y^T y} = \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2}$$

Hence,

$$\lambda_1 \geq R(x) = \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2} \geq \lambda_n$$

Moreover, $R(x) = \lambda_1$ when $y_1 \neq 0$ and $y_i = 0, \forall i > 1$. This means $x = Qy$ is a λ_1 -eigenvector. The case $R(x) = \lambda_n$ case is proved similarly. $\square$

An equivalent statement of the principle is as follows.

Corollary 2.22. Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of a real symmetric matrix A . Over all non-zero unit vectors $x \in \mathbb{R}^n$, $x^T A x$ is maximized at a unit λ_1 -eigenvector, with maximum value λ_1 , and minimized at a unit λ_n -eigenvector, with minimum value λ_n .

Rayleigh's principle essentially states that

$$\lambda_1 = \max_{x \in \mathbb{R}^n} R(x) \quad \text{and} \quad \lambda_n = \min_{x \in \mathbb{R}^n} R(x)$$

What about the rest of the eigenvalues ? Here is a simple answer, stated without proof. The proof is simple enough.

Theorem 2.23. Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of a Hermitian matrix A , and $u_1, \dots, u_n$ are the corresponding set of orthonormal eigenvectors. Then,

$$\begin{aligned} \lambda_k &= \max_{\substack{0 \neq x \in \mathbb{C}^n \\ x \perp u_1, \dots, u_{k-1}}} R_A(x) \\ \lambda_k &= \min_{\substack{0 \neq x \in \mathbb{C}^n \\ x \perp u_{k+1}, \dots, u_n}} R_A(x) \end{aligned}$$

The theorem has a pitfall that sometime we don't know the eigenvectors. The following generalization of Rayleigh's principle, sometime referred to as the *minimax and maximin principles for eigenvalues*, fill the hole by not requiring us to know that eigenvectors.

Theorem 2.24 (Courant-Fisher). Let V_k be the set of all k -dimensional subspaces of $\mathbb{C}^n$. Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ be the eigenvalues of a Hermitian matrix A . Then,

$$\begin{aligned}\lambda_k &= \max_{S \in V_k} \left[\min_{\substack{x \in S \\ x \neq 0}} R(x) \right] \\ &= \min_{S \in V_{n-k+1}} \left[\max_{\substack{x \in S \\ x \neq 0}} R(x) \right]\end{aligned}$$

Note. It should be noted that the previous two theorems are often referred to as the *variational characterization* of the eigenvalues.

Proof. Let $U = [u_1, u_2, \dots, u_n]$ be the unitary matrix with unit eigenvectors $u_1, \dots, u_n$ corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$. Let us first fix $S \in V_k$ and let S' be the image of S under the invertible linear transformation represented by U^* . Obviously, $\dim(S') = k$. We have already known that $R(x)$ is bounded, so it is safe to say the following, with $x \neq 0, y \neq 0$ being implicit.

$$\begin{aligned}\inf_{x \in S} R(x) &= \inf_{x \in S} \frac{x^* A x}{x^* x} \\ &= \inf_{x \in S} \frac{(U^* x)^* \Lambda (U^* x)}{(U^* x)^* (U^* x)} \\ &= \inf_{y \in S'} \frac{y^* \Lambda y}{y^* y} \\ &\leq \inf_{\substack{y \in S' \\ y_1 = \dots = y_{k-1} = 0}} \frac{y^* \Lambda y}{y^* y} \\ &= \inf_{\substack{y \in S' \\ y_1 = \dots = y_{k-1} = 0}} \frac{\lambda_1 y_1^2 + \dots + \lambda_n y_n^2}{y_1^2 + \dots + y_n^2} \\ &= \inf_{\substack{y \in S' \\ y_1 = \dots = y_{k-1} = 0}} \frac{\lambda_k y_k^2 + \dots + \lambda_n y_n^2}{y_k^2 + \dots + y_n^2} \\ &\leq \lambda_k\end{aligned}$$

The inequality in line 4 is justified by the fact that there is a non zero vector $y \in S'$ such that $y_1 = \dots = y_{k-1} = 0$. To get this vector, put k basis vectors of S' into the rows of a $k \times n$ matrix and do Gaussian elimination.

Now, S was chosen arbitrarily, so it is also true that

$$\sup_{S \in V_k} \inf_{x \in S} R(x) \leq \lambda_k$$

Moreover, $R(u_k) = (U^* u_k)^* \Lambda (U^* u_k) = e_k^* \Lambda e_k = \lambda_k$. Thus, the infimum and supremum can be changed to minimum and maximum, and the inequality can be changed to equality. The other equality can be proven similarly. $\square$

This theorem has a very important and beautiful corollary, called the *Interlacing of eigenvalues* to be presented in the next section. Let us introduce a simple corollary.

Corollary 2.25. *Let $A \in M_n$ be Hermitian, let k be a given integer with $1 \leq k \leq n$, let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ be the eigenvalues of A , and let S_k be a given k -dimensional subspace of $\mathbb{C}^n$. The following hold*

- (a) *If there exists c_1 such that $R(x) \leq c_1$ for all $x \in S_k$, then $c_1 \geq \lambda_{n-k+1} \geq \dots \geq \lambda_n$.*
- (b) *If there exists c_2 such that $R(x) \geq c_2$ for all $x \in S_k$, then $\lambda_1 \geq \dots \geq \lambda_k \geq c_2$.*

Proof. It is almost straightforward from the Courant-Fisher theorem that

$$(a) \quad c_1 \geq \max_{0 \neq x \in S_k} R(x) \geq \min_{\dim(S)=n-(n-k+1)+1} \max_{0 \neq x \in S} R(x) = \lambda_{n-k+1}$$

$$(b) \quad c_2 \leq \min_{0 \neq x \in S_k} R(x) \leq \max_{\dim(S)=k} \min_{0 \neq x \in S} R(x) = \lambda_k$$

□

2.2.4 Applications of the variational characterizations

Throughout the rest of this section, we use M_n to denote the set of all $n \times n$ matrices over $\mathbb{C}$ (i.e. $M_n = \mathbb{C}^{n^2}$). The first application of the variational characterization is a theorem by Weyl.

Theorem 2.26 (Weyl, year?). *Let $A, B \in M_n$ be Hermitian and the eigenvalues $\lambda_i(A)$, $\lambda_i(B)$, and $\lambda_i(A+B)$ be arranged in decreasing order, i.e. $\lambda_1(X) \geq \dots \geq \lambda_n(X)$, for $X = A, B, A+B$. For each $k = 1, 2, \dots, n$ we have*

$$\lambda_k(A) + \lambda_1(B) \geq \lambda_k(A+B) \geq \lambda_k(A) + \lambda_n(B)$$

Proof. By noticing that $\lambda_1(B) \geq R_B(x) \geq \lambda_n(B)$, for all $x \in \mathbb{C}^n$, we have

$$\begin{aligned} \lambda_k(A+B) &= \max_{S \in V_k} \min_{0 \neq x \in S} \frac{x * (A+B)x}{x * x} \\ &= \max_{S \in V_k} \min_{0 \neq x \in S} (R_A(x) + R_B(x)) \\ &\geq \max_{S \in V_k} \min_{0 \neq x \in S} (R_A(x) + \lambda_n(B)) \\ &= \lambda_k(A) + \lambda_n(B) \end{aligned}$$

The other direction is proven similarly. □

Notice that if B is positive semidefinite, then $\lambda_1(B) \geq 0$. We have the following trivial corollary, called the *monotonicity theorem* since it says that the eigenvalues keep increasing as positive semidefinite matrices are added.

Corollary 2.27 (Monotonicity Theorem). Suppose A and B are Hermitian with B being positive semidefinite, and the eigenvalues of A and $A + B$ sorted in decreasing order, then

$$\lambda_k(A) \leq \lambda_k(A + B)$$

Theorem 2.28 (Interlacing of eigenvalues). Let A be a Hermitian matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Let B be the matrix obtained from A by removing row i and column i , for any $i \in [n]$. Suppose B has eigenvalues $\theta_1 \geq \dots \geq \theta_{n-1}$, then

$$\lambda_1 \geq \theta_1 \geq \lambda_2 \geq \theta_2 \geq \dots \geq \theta_{n-1} \geq \lambda_n$$

Note. A proof of this theorem using Spectral Decomposition Theorem can also be given, but it is not very instructive.

Proof. We can safely assume that $i = n$ for the ease of presentation. We would like to show that as $1 \leq k \leq n - 1$, $\lambda_k \geq \theta_k \geq \lambda_{k+1}$. Let $x = [y^T x_n]^T \in \mathbb{C}^n$ where $y \in \mathbb{C}^{n-1}$. Note that if $x_n = 0$ then $x^* A x = y^* B y$. We first use the maximin form of Courant-Fisher theorem to write

$$\begin{aligned} \lambda_k &= \max_{\substack{S \subseteq \mathbb{C}^n \\ \dim(S)=k}} \min_{0 \neq x \in S} \frac{x^* A x}{x^* x} \\ &\geq \max_{\substack{S \subseteq \{e_n\}^\perp \\ \dim(S)=k}} \min_{0 \neq x \in S} \frac{x^* A x}{x^* x} \\ &= \max_{\substack{S \subseteq \{e_n\}^\perp \\ \dim(S)=k}} \min_{\substack{0 \neq x \in S \\ x_n=0}} \frac{x^* A x}{x^* x} \\ &= \max_{\substack{S \subseteq \mathbb{C}^{n-1} \\ \dim(S)=k}} \min_{0 \neq y \in S} \frac{y^* B y}{y^* y} \\ &= \theta_k \end{aligned}$$

Now, we use the minimax form of the theorem to obtain $\lambda_{k+1} \leq \theta_k$.

$$\begin{aligned} \lambda_{k+1} &= \min_{\substack{S \subseteq \mathbb{C}^n \\ \dim(S)=n-(k+1)+1}} \max_{0 \neq x \in S} \frac{x^* A x}{x^* x} \\ &\leq \min_{\substack{S \subseteq \{e_n\}^\perp \\ \dim(S)=n-k}} \max_{0 \neq x \in S} \frac{x^* A x}{x^* x} \\ &= \min_{\substack{S \subseteq \{e_n\}^\perp \\ \dim(S)=n-k}} \max_{\substack{0 \neq x \in S \\ x_n=0}} \frac{x^* A x}{x^* x} \\ &= \min_{\substack{S \subseteq \mathbb{C}^{n-1} \\ \dim(S)=(n-1)-k+1}} \max_{0 \neq y \in S} \frac{y^* B y}{y^* y} \\ &= \theta_k \end{aligned}$$

□

2.2.5 Sylvester's law of inertia

Material in this section follows closely that in the book *Matrix Analysis* by Horn and Johnson [57].

Definition 2.29. Let $A, B \in M_n$ be given. If there exists a non-singular matrix S such that

- $B = SAS^*$, then B is said to be $*$ -congruent to A .
- $B = SAS^T$, then B is said to be T -congruent to A .

Note. These two notion of congruence must be closely related; they are the same if S is a real matrix. When it is not important to distinguish between the two, we use the term *congruence* without a prefix. Since S was required to be non-singular, congruent matrices have the same rank. Also note that, if A is Hermitian then so is SAS^* ; if A is symmetric, then SAS^T is also symmetric.

Proposition 2.30. Both $*$ -congruence and T -congruence are equivalent relations.

Proof. It is easy to verify that the relations are reflexive, symmetric and transitive. Only need to notice that S is non-singular. $\square$

The set M_n , therefore, is partitioned into equivalence classes by congruence. As an abstract problem, we may seek a canonical representative of each equivalence class under each type of congruence. *Sylvester's law of inertia* gives us the affirmative answer for the $*$ -congruence case, and thus also gives the answer for the set of real symmetric matrices.

Definition 2.31. Let $A \in M_n$ be a Hermitian matrix. The *inertia* of A is the ordered triple

$$i(A) = (i_+(A), i_-(A), i_0(A))$$

where $i_+(A)$ is the number of positive eigenvalues of A , $i_-(A)$ is the number of negative eigenvalues of A , and $i_0(A)$ is the number of zero eigenvalues of A , all counting multiplicity. The *signature* of A is the quantity $i_+(A) - i_-(A)$.

Note. Since $\text{rank}(A) = i_+(A) + i_-(A)$, the signature and the rank of A uniquely identify the inertia of A .

Theorem 2.32 (Sylvester's law of inertia). Let $A, B \in M_n$ be Hermitian matrices. A and B are $*$ -congruent if and only if A and B have the same inertia.

Proof. ($\Rightarrow$). Firstly, for any Hermitian matrix $C \in M_n$, C can be diagonalized by a unitary matrix U , i.e. $C = U\Lambda U^*$, with Λ being diagonal containing all eigenvalues of C . By multiplying U with a permutation matrix, we can safely assume that down the main diagonal of Λ , all positive eigenvalues go first: $\lambda_1, \dots, \lambda_{i_+}$, then the negatives: $\lambda_{i_++1}, \dots, \lambda_{i_++i_-}$, and the rest are 0's. Thus, if we set $D = \text{diag}(\sqrt{|\lambda_1|}, \dots, \sqrt{|\lambda_{i_++i_-}|}, 0, \dots, 0)$, then

$$\Lambda = D \begin{bmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & -1 & & & & \\ & & & & \ddots & & & \\ & & & & & -1 & & \\ & & & & & & 0 & \\ & & & & & & & \ddots \\ & & & & & & & & 0 \end{bmatrix} \quad D = DI_C D$$

with the entries not shown being 0. I_C is called the *inertia matrix* of C . Consequently, letting $S = UD$ (S is clearly non-singular) we get

$$C = U\Lambda U^* = UDI_CDU^* = SI_CS^* \quad (2.2)$$

Hence, if A and B have the same inertia, then they could be written in the form (2.2) with possibly a different S for each, but $I_A = I_B$. Since $*$ -congruence is transitive, A is $*$ -congruent to B as they are both congruent to I_A .

($\Leftarrow$). Now, assume $A = SBS^*$ for some non-singular matrix S . A and B have the same rank, so $i_0(A) = i_0(B)$. We are left to show that $i_+(A) = i_+(B)$. For convenience, let $a = i_+(A)$ and $b = i_+(B)$. Let $u_1, \dots, u_a$ be the orthonormal eigenvectors for the positive eigenvalues of A , so that $\dim(\text{Span}\{u_1, \dots, u_a\}) = a$. If $x = c_1u_1 + \dots + c_a u_a$, then $x^*Ax = \lambda_1|c_1|^2 + \dots + \lambda_a|c_a|^2 > 0$. But then

$$x^*Ax = x^*SBS^*x = (S^*x)^*B(S^*x) > 0$$

so $y^*By > 0$ for all vector y in $\text{Span}\{S^*u_1, \dots, S^*u_a\}$, which also has dimension a . By Corollary 2.25, $b \geq a$. A similar argument shows that $a \geq b$, which completes the proof. $\square$

Corollary 2.33. *Given $x \in \mathbb{R}^n$, if $x^T Ax$ can be written as the sum of m products involving two linear factors, that is*

$$x^T Ax = \sum_{k=1}^m \left(\sum_{i \in S_k} b_{i,k} x_i \right) \left(\sum_{j \in T_k} c_{j,k} x_j \right)$$

Further assume that A has p positive eigenvalues and q positive eigenvalues (counting multiplicities), then $m \geq \max(p, q)$.

Proof. I have not been able to see why this corollary follows from Sylvester's law yet. A proof of the corollary can be given, but that's not the point. $\square$