Chapter 7. Symmetric Matrices and Quadratic Forms

7.1 Diagonalization of Symmetric Matrices
Schur Triangular Form.

Definition. A matrix \( P \in \mathbb{C}^{n \times n} \) is unitary if \( P^*P = I_n \).

Schur Factorization. Any \( A \in \mathbb{C}^{n \times n} \) can be written in the form \( A = PUP^* \) where \( P \in \mathbb{C}^{n \times n} \) is unitary and \( U \in \mathbb{C}^{n \times n} \) is upper triangular.

This can be proven by induction. The case \( n = 1 \) is clear.

Now suppose the result holds for \( (n-1) \times (n-1) \) matrices and let \( A \in \mathbb{C}^{n \times n} \).

Let \( \{\lambda_1, v_1\} \) be an eigenvalue/eigenvector pair for \( A \) with \( \|v_1\| = 1 \).

Extend \( v_1 \) to an orthonormal basis \( \{v_1, \ldots, v_n\} \) for \( \mathbb{C}^n \) and set \( P_1 = [v_1, \cdots, v_n] \).

Schur Factorization. (cont.) Note \( P_1^* = P_1^{-1} \). we may write

\[
AP_1 = P_1 \begin{bmatrix} \lambda_1 & w \\
0 & M \end{bmatrix}, \quad M \in \mathbb{C}^{(n-1) \times (n-1)}, \quad w \in \mathbb{C}^{1 \times (n-1)}.
\]

By assumption, we can write \( M = QU_0Q^* \), \( Q \) unitary and \( U \) is upper triangular.

Now set \( P_2 = \begin{bmatrix} 1 & 0 \\
0 & Q \end{bmatrix} \), \( P = P_1P_2 \).

Then \( P \) is unitary (check!) and

\[
P^*AP = P_2^* \begin{bmatrix} \lambda_1 & w \\
0 & M \end{bmatrix} P_2 = \begin{bmatrix} \lambda & wQ \\
0 & U_0 \end{bmatrix},
\]

which completes the proof.
Schur Triangular Form.

This result shows that every $A \in \mathbb{C}^{n \times n}$ is similar to an upper triangular matrix $U \in \mathbb{C}^{n \times n}$ via a change of coordinate matrix $P \in \mathbb{C}^{n \times n}$ that is unitary.

That is: every matrix $A$ is unitarily similar to an upper triangular matrix.

Definition. (Normal matrices) A matrix $A \in \mathbb{C}^{n \times n}$ is normal if

$$A^*A = AA^*.$$  

Examples of normal matrices.

- If $A^* = A$ (i.e. $A$ is hermitian), then $A$ is normal.
- If $A \in \mathbb{R}^{n \times n}$ is symmetric ($A = A^T$), then $A$ is normal.
- If $A^* = -A$ (skew-adjoint), then $A$ is normal.
- If $A$ is unitary ($A^*A = I_n$), then $A$ is normal.
Theorem. If \( A \in \mathbb{C}^{n \times n} \) is normal and \((\lambda, v)\) is an eigenvalue/eigenvector pair, then \(\{\bar{\lambda}, v\}\) is an eigenvalue/eigenvector pair for \(A^*\).

Indeed,

\[
\| (A - \lambda I)v \|^2 = \| (A - \lambda I)v \|^* (A - \lambda I)v \\
= v^* (A^* - \bar{\lambda}I) (A - \lambda I)v \\
= v^* (A - \lambda I) (A^* - \bar{\lambda}I) v \\
= \| (A^* - \bar{\lambda}I)v \|^2.
\]
Spectral Theorem. (cont.)

Now suppose $A \in \mathbb{C}^{n \times n}$ is normal, i.e. $AA^* = A^*A$. We begin by writing the Schur factorization of $A$, i.e.

$$A = PUP^*, \quad P = [v_1 \cdots v_n],$$

where $P$ is unitary and $U = [c_{ij}]$ is upper triangular.

First note that $AP = PU$ implies $Av_1 = c_{11}v_1$, and hence (since $A$ is normal) $A^*v_1 = \bar{c}_{11}v_1$.

However, $A^*P = PU^*$, so that

$$\bar{c}_{11}v_1 = A^*v_1 = \bar{c}_{11}v_1 + \cdots + \bar{c}_{1n}v_n$$

By independence of $v_2, \ldots, v_n$, we deduce $c_{1j} = 0$ for $j = 2, \ldots, n$.

We have shown

$$U = \begin{bmatrix} c_{11} & 0 \\ 0 & \tilde{U} \end{bmatrix},$$

where $\tilde{U} \in \mathbb{C}^{(n-1) \times (n-1)}$ is upper triangular.

But now $AP = PU$ gives $Av_2 = c_{22}v_2$, and arguing as above we deduce $c_{2j} = 0$ for $j = 3, \ldots, n$.

Continuing in this way, we deduce that $U$ is diagonal. □
Spectral Theorem. (cont.)

To summarize, $A \in \mathbb{C}^{n \times n}$ is normal ($AA^* = A^*A$) if and only if it can be written as $A = PDP^*$ where $P = [v_1 \cdots v_n]$ is unitary and $D = \text{diag}(\lambda_1, \cdots, \lambda_n)$. Note

- $P$ unitary means $P^{-1} = P^*$
- $A$ is unitarily similar to a diagonal matrix
- $\{\lambda_j, v_j\}$ are eigenvalue-eigenvector pairs for $A$

Theorem. (Spectral Theorem for Self-Adjoint Matrices)

- A matrix $A \in \mathbb{C}^{n \times n}$ is self-adjoint ($A = A^*$) if and only if it is unitarily similar to a real diagonal matrix, i.e. $A = PDP^*$ for some unitary $P \in \mathbb{C}^{n \times n}$ and some diagonal $D \in \mathbb{R}^{n \times n}$.

Indeed, this follows from the spectral theorem for normal matrices. In particular,

$$PDP^* = A = A^* = PD^*P \implies D = D^*,$$

which implies that $D \in \mathbb{R}^{n \times n}$.

Note this implies that self-adjoint matrices have real eigenvalues.
Eigenvectors and eigenvalues for normal matrices. Suppose $A$ is a normal matrix.

- Eigenvectors associated to different eigenvalues are orthogonal:
  
  $$v_1 \cdot Av_2 = \lambda_2 v_1 \cdot v_2,$$
  
  $$v_1 \cdot Av_2 = A^* v_1 \cdot v_2 = \lambda_1 v_1 \cdot v_2.$$

- If the eigenvalues are all real, then $A$ is self-adjoint. (This follows from the spectral theorem.)

Spectral decomposition. If $A \in \mathbb{C}^{n \times n}$ is a normal matrix, then we may write $A = PDP^*$ as above. In particular,

$$A = \lambda_1 v_1 v_1^* + \cdots + \lambda_n v_n v_n^*$$

Recall that

$$\frac{1}{\|v_k\|^2} v_k v_k^* = v_k v_k^*$$

is the projection matrix for the subspace $V_k = \text{span}\{v_k\}$.

Thus, a normal matrix can be written as the sum of scalar multiples of projections on to the eigenspaces.
Chapter 7. Symmetric Matrices and Quadratic Forms

7.2 Quadratic Forms

Definition. Let \( A \in \mathbb{C}^{n \times n} \) be a self-adjoint matrix. The function
\[
Q(x) = x^* Ax, \quad x \in \mathbb{C}^n
\]
is called a quadratic form. Using self-adjointness of \( A \), one finds
\[
Q : \mathbb{C}^n \rightarrow \mathbb{R}.
\]

- If \( Q(x) > 0 \) for all \( x \neq 0 \), we call \( Q \) positive definite.
- If \( Q(x) \geq 0 \) for all \( x \neq 0 \), we call \( Q \) positive semidefinite.
- We define negative definite, negative semidefinite similarly.
- We call \( Q \) indefinite if it attains both positive and negative values.
Characteristic forms. Expanding the inner product, we find that

\[ x^*Ax = \sum_{j=1}^{n} a_{jj}|x_j|^2 + 2 \sum_{i<j} \text{Re}(a_{ij}x_i x_j). \]

For \( A \in \mathbb{R}^{n \times n} \) and \( x \in \mathbb{R}^n \), this reduces to

\[ x^TAx = \sum_{j=1}^{n} a_{jj}x_j^2 + 2 \sum_{i<j} a_{ij}x_i x_j. \]

Example.

\[ x^T \begin{bmatrix} 1 & -2 & 3 \\ -2 & 4 & -5 \\ 3 & -5 & -6 \end{bmatrix} x = x_1^2 + 4x_2^2 - 6x_3^2 - 4x_1x_2 + 6x_1x_3 - 10x_2x_3. \]

Characterization of definiteness. Let \( A \in \mathbb{C}^{n \times n} \), \( Q(x) = x^*Ax \).

- There exists an orthonormal basis \( B = \{v_1, \ldots, v_n\} \) s.t.
  - \( A = PDP^* \), where \( P = [v_1 \cdots v_n] \) and
  - \( D = \text{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n \times n} \).

Then, with \( y = P^{-1}x \)

\[ Q(x) = x^*PDP^*x = (P^{-1}x)^*DP^{-1}x = y^*Dy \]

\[ = \lambda_1|y_1|^2 + \cdots + \lambda_n|y_n|^2. \]

We conclude:

**Theorem.** If \( A \in \mathbb{C}^{n \times n} \) is self-adjoint, then \( Q(x) = x^*Ax \) is positive definite if and only if the eigenvalues of \( A \) are all positive.

(Similarly for negative definite, or semidefinite...)
**Quadratic forms and conic sections.** The equation

\[
ax_1^2 + 2bx_1x_2 + cx_2^2 + dx_1 + ex_2 = f
\]

can be written as

\[
x^T Ax + [d \ e] x = f, \quad A = A^T = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.
\]

By the spectral theorem, there is a basis of eigenvectors \( \{v_1, v_2\} \) that diagonalizes \( A \). That is,

\[
A = PDP^T, \quad P = [v_1 \ v_2], \quad D = \text{diag}(\lambda_1, \lambda_2).
\]

Writing \( y = P^Tx \), the equation becomes

\[
y^T Dy + [d' \ e'] y = f, \quad [d' \ e'] = [d \ e] P,
\]
i.e.

\[
\lambda_1 y_1^2 + \lambda_2 y_2^2 + d'y_1 + e'y_2 = f.
\]

**Principle axis theorem.** The change of variables \( y = P^Tx \) gives

\[
x^T Ax + [d \ e] x = f \iff y^T Dy + [d' \ e'] y = f.
\]

The nature of the conic section can be understood through the quadratic form \( y^T Dy \).

Note that this transforms \( x^*Ax \) into a quadratic form \( y^*Dy \) with no cross-product term.
Example. Consider $x_1^2 - 6x_1x_2 + 9x_2^2$. This corresponds to

$$A = \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}.$$ 

The eigenvalues are $\lambda = 10, 0$ (the quadratic form is positive definite), with eigenspaces

$$E_0 = \text{span}([3, 1]^T), \quad E_{10} = \text{span}([1, -3]^T).$$

Consequently $A = PDPT$, with

$$P = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix}.$$ 

Writing $y = P^Tx$ leads to the quadratic form

$$10y_1^2 + 0y_2^2 = 10y_1^2.$$ 

Example. (cont.) Consider the conic section described by

$$x_1^2 - 6x_1x_2 + 9x_2^2 + 3x_1 + x_2 = 1.$$ 

This can be written $x^TAx + [3 \ 1]^T = 1$. Continuing from above, this is equivalent to

$$10y_1^2 + [3 \ 1]P^Ty = 10y_1^2 + \sqrt{10}y_2 = 1,$$

i.e. $y_2 = \frac{\sqrt{10}y_1}{10} - \sqrt{10}y_1^2$.

In the $y_1y_2$ plane, the conic section is a parabola. To go from $x$ coordinates to $y$ coordinates, we apply $P$, which is a rotation.
Recall: A self-adjoint matrix $A \in \mathbb{C}^{n \times n}$ is unitarily similar to a real diagonal matrix. Consequently, we can write

$$A = \lambda_1 u_1 u_1^* + \cdots + \lambda_n u_n u_n^*,$$

where $\lambda_n \leq \cdots \leq \lambda_1 \in \mathbb{R}$ and $\{u_1, \cdots, u_n\}$ is an orthonormal basis.
**Quadraic forms and boundedness.** Let $A$ be self-adjoint. Continuing from above,

$$x^*Ax = \lambda_1 x^* u_1 (u_1^* x) + \cdots + \lambda_n x^* u_n (u_n^* x)$$

$$= \lambda_1 |u_1^* x|^2 + \cdots + \lambda_n |u_n^* x|^2.$$ 

Since $\{u_1, \ldots, u_n\}$ is an orthonormal basis,

$$x = (u_1^* x) u_1 + \cdots + (u_n^* x) u_n \implies \|x\|^2 = |u_1^* x|^2 + \cdots + |u_n^* x|^2.$$ 

We deduce

$$\lambda_n \|x\|^2 \leq x^* Ax \leq \lambda_1 \|x\|^2.$$ 

**Rayleigh principle.** We continue with $A$ as above and set

$$\Omega_0 = \{0\}, \quad \Omega_k := \text{span}\{u_1, \ldots, u_k\}.$$ 

Then for $x \in \Omega_{k-1}^\perp$ we have

$$\|x\|^2 = |u_k^* x|^2 + \cdots + |u_n^* x|^2,$$

$$x^* Ax = \lambda_k |u_k^* x|^2 + \cdots + \lambda_n |u_n^* x|^2.$$ 

Thus (using $\lambda_n \leq \cdots \leq \lambda_1$)

$$\lambda_n \|x\|^2 \leq x^* Ax \leq \lambda_k \|x\|^2.$$ 

$$\implies \lambda_n \leq x^* Ax \leq \lambda_k \quad \text{for all } x \in \Omega_{k-1}^\perp \quad \text{with } \|x\| = 1.$$ 

But since $u_n^* Au_n = \lambda_n$ and $u_k^* Au_k = \lambda_k$, we deduce the Rayleigh principle: for $k = 1, \ldots, n$,

$$\min_{\|x\|=1} x^* Ax = \min_{\|x\|=1, \ x \in \Omega_{k-1}^\perp} x^* Ax = \lambda_n,$$

$$\max_{\|x\|=1} x^* Ax = \lambda_k.$$
**Example.** Let \( Q(x_1, x_2) = 3x_1^2 + 9x_2^2 + 8x_1x_2, \) which corresponds to
\[
A = \begin{bmatrix} 3 & 4 \\ 4 & 9 \end{bmatrix}.
\]
The eigenvalues are \( \lambda_1 = 11 \) and \( \lambda_1 = 2, \) with
\[
\Omega_1 = \text{nul}(A - 11I) = \text{span}\{[1, 2]^T\},
\]
\[
\Omega_1^\perp = \text{nul}(A - I) = \text{span}\{[-2, 1]^T\}.
\]
Note
\[
\min_{\|x\| = 1} x^*Ax = \lambda_2 = 1, \quad \max_{\|x\| = 1} x^*Ax = \lambda_1 = 11,
\]
By the Rayleigh principle, the minimum is obtained on \( \Omega_1^\perp, \) while
the maximum restricted to this set is also equal to \( \lambda_2 = 1. \)

**Example.** (cont.)
The contour curves \( Q(x_1, x_2) = \text{const} \) are ellipses in the \( x_1x_2 \) plane.

Using the change of variables \( y = P^*x, \) where
\[
P = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}
\]
is a rotation by \( \theta \sim 63.44^\circ, \) one finds \( Q(x) = 11y_1^2 + y_2^2. \)

Thus the contour curves \( Q(x_1, x_2) = \text{const} \) are obtained by
rotating the contour curves of \( 11x_1^2 + x_2^2 = \text{const} \) by \( \theta. \)
Singular values. For a matrix \( A \in \mathbb{C}^{n \times p} \), the matrix \( A^*A \in \mathbb{C}^{p \times p} \) is self-adjoint. By the spectral theorem, there exists an orthonormal basis \( B = \{v_1, \ldots, v_p\} \) for \( \mathbb{C}^p \) consisting of eigenvectors for \( A^*A \) with real eigenvalues \( \lambda_1 \geq \cdots \geq \lambda_p \).

Noting that \( x^*(A^*A)x = (Ax)^*Ax = \|Ax\|^2 \geq 0 \) for all \( x \), we deduce
\[
\lambda_j = \lambda_j \|v_j\|^2 = v_j^*(A^*A)v_j \geq 0 \quad \text{for all} \quad j.
\]

**Definition.** With the notation above, we call \( \sigma_j := \sqrt{\lambda_j} \) the singular values of \( A \).

- If \( \text{rank} A = r \), then \( \sigma_{r+1} = \cdots = \sigma_p = 0 \).
- In this case \( \{v_1, \ldots, v_r\} \) is an orthonormal basis for \( \text{col}(A^*) \), while \( \{v_{r+1}, \ldots, v_p\} \) is an orthonormal basis for \( \text{nul}(A) \).
Singular Value Decomposition. Let $A \in \mathbb{C}^{n \times p}$ with rank $A = r$ as above. The vectors

$$
\mathbf{u}_j = \frac{1}{\sigma_j} A \mathbf{v}_j, \quad j = 1, \ldots, r
$$

form an orthonormal basis for $\text{col}(A)$. Indeed,

$$
\mathbf{u}_i \cdot \mathbf{u}_j = \frac{\mathbf{v}_i^*(A^* A) \mathbf{v}_j}{\sigma_i \sigma_j} = \begin{cases} 0 & i \neq j \\ 1 & i = 1. \end{cases}
$$

Next let $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for $\text{col}(A)^\perp$. Defining the unitary matrices $V = [\mathbf{v}_1 \cdots \mathbf{v}_p]$ and $U = [\mathbf{u}_1 \cdots \mathbf{u}_n],$

$$
AV = U \Sigma, \quad \Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{C}^{n \times p}, \quad D = \text{diag}(\sigma_1, \ldots, \sigma_r).
$$

We call $A = U \Sigma V^*$ the singular value decomposition of $A \in \mathbb{C}^{n \times p}$.

SVD and linear transformations. Let $T(x) = Ax$ be a linear transformation $T : \mathbb{C}^p \rightarrow \mathbb{C}^n$.

Writing $A = U \Sigma V^*$ as above, we have $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ and $C = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ are orthonormal bases for $\mathbb{C}^p$ and $\mathbb{C}^n$. Then

$$
U^*(Ax) = \Sigma(V^*x) \implies [T(x)]_C = \Sigma[x]_B,
$$

i.e. there are orthonormal bases for $\mathbb{C}^p$ and $\mathbb{C}^n$ s.t. $T$ can be represented in terms of the matrix $\Sigma$. 
**Transformations of** $\mathbb{R}^2$. If $T : \mathbb{R}^2 \to \mathbb{R}^2$ is given by $T(x) = Ax$, then there exist unitary matrices $U, V$ so that $A = UDV^T$ for $D = \text{diag}(\sigma_1, \sigma_2)$.

Unitary matrices in $\mathbb{R}^{2 \times 2}$ represent rotations/reflections of the plane.

Every linear transformation of the plane is the composition of three transformations: a rotation/reflection, a scaling transformation, and a rotation/reflection.

---

**Moore–Penrose inverse of** $A \in \mathbb{C}^{n \times p}$. Write $V_r = [v_1 \cdots v_r] \in \mathbb{C}^{p \times r}$ and $U_r = [u_1 \cdots u_r] \in \mathbb{C}^{n \times r}$. Then

$$A = U \Sigma V^* = U_r D V_r^*$$

represents a reduced SVD for $A$.

**Definition.** The **Moore–Penrose pseudo inverse** of $A \in \mathbb{C}^{n \times p}$ is defined by

$$A^+ = V_r D^{-1} U_r^* \in \mathbb{C}^{p \times n}.$$

- $AA^+ = U_r U_r^* = \text{proj}_{\text{col}(A)} \in \mathbb{C}^{n \times n}$
- $A^+ A = V_r V_r^* = \text{proj}_{\text{col}(A^+)} \in \mathbb{C}^{p \times p}$
- $AA^+ A = A$, $A^+ AA^+ = A^+$,
- $A^+ = A^{-1}$ whenever $r = p = n$. 
Least squares solutions for $A \in \mathbb{C}^{n \times p}$. Recall that the least squares solutions of $Ax = b$ are the solutions to the normal system $A^*Ax = A^*b$. Equivalently, they are solutions to $Ax = \text{proj}_{\text{col}(A)}b$.

When $\text{rank}(A^*A) = r < p$, there are infinitely many least squares solutions.

Note that since $AA^* = \text{proj}_{\text{col}(A)}$, we have

$$AA^*b = \text{proj}_{\text{col}(A)}(b) \implies A^*b \text{ is a least squares solution.}$$

On the other hand, using $A^*b \in \text{col}(A^*)$, we have for any other least squares solution $\hat{x}$,

$$A\hat{x} - AA^*b = 0 \implies \hat{x} - A^*b \in \text{nul}(A) = \text{col}(A^*)^\perp,$$

so $A^*b \perp \hat{x} - A^*b$. Consequently,

$$||\hat{x}||^2 = ||A^*b||^2 + ||\hat{x} - A^*b||^2.$$

Thus $A^*b$ is the least squares solution of smallest length.

Four fundamental subspaces. Let $A \in \mathbb{C}^{n \times p}$. Consider

- $\text{col}(A)$, $\text{col}(A)^\perp = \text{nul}(A^*)$
- $\text{col}(A^*) = \text{row}(\bar{A})$, $\text{col}(A^*)^\perp = \text{nul}(A)$

Recall the SVD of $A \in \mathbb{C}^{n \times p}$ with $\text{rank}(A) = r$ yields an orthonormal basis $\{v_1, \ldots, v_p\}$ consisting of eigenvectors of $A^*A$, and an orthonormal basis $\{u_1, \ldots, u_r\}$ obtained by completing

$\{u_1, \ldots, u_r\}$, where $u_j = \frac{1}{\sigma_j}Av_j$.

Since $A^*Av_j = \lambda_j v_j$, $Av_j = \sigma_j u_j$:

- $\{v_1, \ldots, v_r\}$ is an orthonormal basis for $\text{col}(A^*A) = \text{col}(A^*) = \text{row}(\bar{A})$
- $\{v_{r+1}, \ldots, v_p\}$ is an orthonormal basis for $\text{col}(A^*)^\perp = \text{nul}(A)$
- $\{u_1, \ldots, u_r\}$ is an orthonormal basis for $\text{col}(A)$
- $\{u_{r+1}, \ldots, u_n\}$ is an orthonormal basis for $\text{col}(A)^\perp = \text{nul}(A^*)$
Review: Matrix Factorizations

Let $A \in \mathbb{C}^{n \times p}$.

- **Permutated LU factorization:** $PA = LU$, where $P \in \mathbb{C}^{n \times n}$ is an invertible permutation matrix, $L \in \mathbb{C}^{n \times n}$ is invertible and lower triangular, and $U \in \mathbb{C}^{n \times p}$ is upper triangular.

- **QR factorization:** $A = QR$, where the columns of $Q \in \mathbb{C}^{n \times p}$ are generated from the columns of $A$ by Gram-Schmidt and $R \in \mathbb{C}^{p \times p}$ is upper triangular.

- **SVD:** $A = U\Sigma V^*$, where $U \in \mathbb{C}^{n \times n}$, $V \in \mathbb{C}^{p \times p}$ are unitary,
  
  $$
  D = \begin{bmatrix}
  D & 0 \\
  0 & 0 \\
  \end{bmatrix} \in \mathbb{C}^{n \times p}, \quad D = \text{diag}(\sigma_1, \ldots, \sigma_r).
  $$

  For $A \in \mathbb{C}^{n \times n}$:

- **Schur factorization:** $A = PUP^*$ where $P$ is unitary and $U$ is upper triangular.

- **Spectral theorems:** $A = PDP^*$, where $P$ is unitary and $D$ is diagonal. This holds if and only if $A$ is normal. The matrix $D$ is real if and only if $A$ is self-adjoint.