Hermitian Forms and Zeros of a Polynomial

Pranshu Gaba

Indian Institute of Science, Bangalore
gabapranshu@ug.iisc.in

November 30, 2017

Abstract

We looked at the general properties of Hermitian (self-adjoint) matrices, and used the Schur-Cohn theorem to find the number of roots of a polynomial lying within and without the unit circle.

1 Introduction

In this paper we see the properties of Hermitian matrices, which are very useful and interesting. We also see and prove the Schur-Cohn theorem to find the number of roots of a polynomial lying within the unit circle.

There are many ways to locate the roots of a polynomial. The Schur-Cohn theorem shows a surprising connection between linear algebra and roots of a polynomial. It will be used to find out how many roots of the polynomial lie inside and outside the unit circle.

First we will define some basic terms that will be used ahead in the paper.

2 Definitions

2.1 Norm of a matrix

2.1.1 Operator norm

Given $A \in \mathbb{M}_n$ (the set of $n \times n$ square matrices with complex elements), the operator norm of A, denoted by $\|A\|$, is defined as

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\|$$

The triangle inequality $\|A + B\| \leq \|A\| + \|B\|$ is satisfied.

2.1.2 Hilbert-Schmidt norm

The Hilbert-Schmidt norm of matrix A, denoted by $\|A\|_2$, is defined as the square root of sum of squares of all entries in A.

$$\|A\|_2 = \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2}$$

Theorem 1. *The operator norm is always less than or equal to the Hilbert-Schmidt norm.*

Proof.

$$\begin{aligned}
\|Ax\|^2 &= \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} x_j \right|^2 \\
&\leq \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |a_{ij}| \, |x_j| \right)^2 \\
&\leq \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \right) \left(\sum_{j=1}^{n} |x_j|^2 \right) = \left(\sum_{i,j} |a_{ij}|^2 \right) \|x\|^2
\end{aligned}$$

Therefore $\dfrac{\|Ax\|}{\|x\|} \leq \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2}$, which is equivalent to $\|A\| \leq \|A\|_2$. $\square$

2.2 Inner product

The inner product is a binary operator on two vectors $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C}$. It satisfies the following conditions for all $x, y, z \in V$ and $a \in \mathbb{C}$:

- It is linear in the first term.

  $\langle ax, y \rangle = a\langle x, y \rangle$

  $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

- It becomes its complex conjugate when the arguments are reversed.

  $\langle x, y \rangle = \overline{\langle y, x \rangle}$

- Inner product of a vector with itself is non-negative.

  $\langle x, x \rangle \geq 0$. Here equality is achieved if and only if $x = 0$.

For vectors on $\mathbb{C}^n$, the inner product $\langle x, y \rangle$ is defined as $x^T \overline{y}$, the vector multiplication of the transpose of the first term with the complex conjugate of the second term. This definition satisfies all the above mentioned conditions.

2.3 Adjoint

Let $A \in \mathbb{M}_n$. The adjoint of matrix A, denoted by A^*, is the matrix that satisfies $\langle A^*x, y\rangle = \langle x, Ay\rangle$.

Theorem 2. *The adjoint of a matrix is obtained by taking the complex conjugate of every element, followed by transposing the matrix.*

Proof. Using properties of inner product on $\mathbb{C}^n$,

$$\begin{aligned} \langle A^*x, y\rangle &= (A^*x)^T\overline{y} \\ &= x^T(A^*)^T\overline{y} \\ = \langle x, Ay\rangle &= x^T\overline{Ay} \end{aligned}$$

Therefore $(A^*)^T = \overline{A}$, or $A^* = (\overline{A})^T$. ☐

Note that the adjoint of the adjoint of a matrix is the original matrix itself, $(A^*)^* = A$.
The adjoint of a matrix, A^*, like A, represents a linear transformation on $\mathbb{C}^n$.

2.4 Positive Definite Matrix

A matrix $A \in \mathbb{M}_n$ that satisfies $\langle Ax, x\rangle \geq 0$ for all $x \in \mathbb{C}^n$ is called a *positive semidefinite matrix* and is denoted as $A \geq 0$. If the inequality is strict, $\langle Ax, x\rangle > 0$, then A is called a *positive definite matrix*, and it is denoted as $A > 0$.

Theorem 3. *Let $A \in \mathbb{M}_n$ be a positive semidefinite matrix. Then all eigenvalues of A are non-negative.*

Proof. Let $\lambda_1, \lambda_2, \dots, \lambda n$ be the eigenvalues of A, and let $x_1, x_2, \dots, x_n$ be the corresponding eigenvectors. For any eigenvector x_i, $\langle Ax_i, x_i\rangle = \langle \lambda_i x_i, x_i\rangle = \lambda_i\langle x_i, x_i\rangle$.
$\lambda_i = \frac{\langle Ax_i, x_i\rangle}{\langle x_i, x_i\rangle}$
$\langle Ax_i, x_i\rangle$ is non-negative because A is positive semidefinite, and $\langle x_i, x_i\rangle$ is positive by definition of inner product. Hence $\lambda_i \geq 0$. All eigenvalues of A are non-negative. ☐

Lemma. *If $A \geq 0$, then $\|A\| = \sup\limits_{\|x\|=1} \langle Ax, x\rangle$*

Proof. $\langle Ax, x\rangle \leq \|Ax\|\,\|x\| \leq \|A\|\,\|x\|^2$. So $\sup_{\|x\|=1}\langle Ax, x\rangle \leq \|A\|$ ☐

Theorem 4. *A^*A is a positive semidefinite for all $A \in \mathbb{M}_n$.*

Proof. $\langle A^*Ax, x\rangle = \langle Ax, Ax\rangle = \|Ax\|^2 \geq 0$ ☐

2.5 Unitary Matrices

A square matrix U that satisfies $U^*U = I$ is called a unitary matrix.

Theorem 5. *The columns of a unitary matrix U form an orthonormal basis.*

Proof. U^{-1} exists, so each column is linearly independent. ☐

Theorem 6. *It preserves inner product, $\langle Ux, Uy\rangle = \langle x, y\rangle$.*

Proof. $\langle x, y\rangle = \langle Ix, y\rangle = \langle U^*Ux, y\rangle = \langle Ux, Uy\rangle$ ☐

2.6 Trace

The trace of matrix A is the sum of the diagonal elements of the matrix.

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \langle Ae_i, e_i \rangle$$

Remark. Here a_{ij} denotes the element in the i^{th} row and j^{th} column. e_i denotes the i^{th} standard basis vector. Note that $a_{ij} = \langle e_i, e_j \rangle$.

Theorem 7. *The trace of A^*A is equal to the square of the Hilbert-Schmidt norm of A.* $\mathrm{Tr}(A^*A) = \|A\|_2^2$

Proof.

$$\begin{aligned}
\mathrm{Tr}(A^*A) &= \sum_{i=1}^{n} \langle A^*Ae_i, e_i \rangle \\
&= \sum_{i=1}^{n} \langle Ae_i, Ae_i \rangle \\
&= \sum_{i=1}^{n} \|Ae_i\|^2 \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ji}|^2 \\
&= \|A\|_2^2
\end{aligned}$$

□

2.7 Hermitian Matrix

A matrix that satisfies $A = A^*$ is called a Hermitian matrix (also known as self-adjoint matrix).

Theorem 8. *All eigenvalues of a Hermitian matrix are real.*

Proof. Let v be an eigenvector of a Hermitian matrix A, and let λ be the corresponding eigenvalue. Then $Av = \lambda v$.

$\langle Av, v \rangle = \langle \lambda v, v \rangle = \lambda \langle v, v \rangle$. Also $\langle Av, v \rangle = \langle v, A^*v \rangle = \langle v, Av \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \langle v, v \rangle$

This means $\lambda \langle v, v \rangle = \overline{\lambda} \langle v, v \rangle$ for any v. Since λ is the same conjugate as its complex conjugate, it implies λ is real. □

The converse of this is also true.

Theorem 9. *[Converse of Theorem 8] If $A \in \mathbb{M}_n$ and $\langle Ax, x \rangle \in \mathbb{R}$ for every x, then $A = A^*$.*

Proof. Let $\alpha \in \mathbb{C}$ and $h, g \in \mathbb{C}^n$. Then $\langle A(h + \alpha g), h + \alpha g \rangle = \langle Ah, h \rangle + \alpha \langle Ag, h \rangle + \overline{\alpha} \langle Ah, g \rangle + |\alpha|^2 \langle Ag, g \rangle$

So $\alpha \langle Ag, h \rangle + \overline{\alpha} \langle Ah, g \rangle = \overline{\alpha} \langle h, Ag \rangle + \alpha \langle g, Ah \rangle$ When $\alpha = 1$, $\langle Ag, h \rangle + \langle Ah, g \rangle = \langle h, Ag \rangle + \langle g, Ah \rangle$

When $\alpha = i$, $i \langle Ag, h \rangle - i \langle Ah, g \rangle = -i \langle h, Ag \rangle + i \langle g, Ah \rangle$

$2i \langle Ag, h \rangle = 2i \langle g, Ah \rangle$ or $\langle Ag, h \rangle = \langle g, Ah \rangle = \langle A^*g, h \rangle$

$Ag = A^*g$ for all g, therefore $A = A^*$. A is Hermitian. □

Corollary. *Every positive semidefinite matrix $A \in \mathbb{M}_n$ is Hermitian.*

Proof. $\langle Ax, x\rangle \geq 0$ so $\langle Ax, x\rangle \in \mathbb{R}$ for all x. By Theorem 9, A is Hermitian. □

Lemma. *If A is Hermitian, then $\|A\| = \sup_{\|h\|=1} |\langle Ah, h\rangle|$.*

Proof. Let $x \in \mathbb{C}^n$. Then for any $g \in \mathbb{C}^n$, we have $|\langle x, g\rangle| \leq \|x\| \, \|g\|$. (By Cauchy-Schwartz).
So $\left|\langle x, \frac{g}{\|g\|}\rangle\right| \leq \|x\|$
So $\sup_{\|g\|=1} |\langle x, g\rangle| \leq \|x\|$ (Equality when $g = \frac{x}{\|x\|}$).
$\|A\| = \sup_{\|h\|=1} \|Ah\|$
$= \sup_{\|h\|=1} \sup_{\|g\|=1} |\langle Ah, g\rangle|$
If $h, g \in \mathbb{C}^n$ with $\|h\| = \|g\|$, then

$$\begin{aligned}
\langle A(h \pm g), h \pm g\rangle &= \langle Ah, h\rangle \pm \langle Ah, g\rangle \pm \langle Ag, h\rangle + \langle Ag, g\rangle \\
&= \langle Ah, h\rangle \pm \langle Ah, g\rangle \pm \langle g, Ah\rangle + \langle Ag, g\rangle \\
&= \langle Ah, h\rangle \pm 2\mathrm{Re}(\langle Ah, g\rangle) + \langle Ag, g\rangle
\end{aligned}$$

$$4\mathrm{Re}\langle Ah, g\rangle =$$

□

Lemma. *If $\langle Ah, h\rangle = 0$ for all h, then $A = 0$.*

Proof. Left to the reader □

2.8 Diagonalization

Hermitian matrices can be diagonalized?

Theorem 10. *For every Hermitian matrix A, there exists a diagonal matrix Λ such that $A = U^*\Lambda U$. Here U is some unitary matrix.*

Proof. The matrix U is a unitary matrix, which means $U^*U = I$
Start with any eigenvalue λ_i of A. Let x_i be a unit eigenvector corresponding to λ_i.
Let $\mathcal{M} = \{y \in \mathbb{C}^n, \ \langle y, x\rangle = 0\}$
$\langle Ay, x\rangle = \langle y, Ax\rangle = \lambda\langle y, x\rangle = 0$.

□

2.9 Projectors

A matrix P is a projector if $P^2 = P$ and $P^* = P$

Theorem 11. *There exists a subspace M of $\mathbb{C}^n$ such that $Pm = m \ \forall m \in M$, and $Px = 0 \ \forall x \in m^\perp$*

Proof. Left to the reader □

2.10 Shift Matrix

Let S, the *shift matrix*, be the $n \times n$ square matrix $\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$.

Note that S is a nilpotent matrix of order n, i.e. S^n is a zero matrix.

Related useful matrices:

$$S^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

$$S^*S = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

$$I - SS^* = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

The last one $I - S^*S$ is a projector.

3 Schur-Cohn Theorem

Given any polynomial $p(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n$ with complex coefficients, we are interested in finding how many of its roots lie within the unit circle and how many roots lie outside. Without loss of generality, let $a_0 = 1$ as it does not change the roots of the polynomial.

Suppose p has roots α_i. Then $p(z) = (z - \alpha_1)(z - \alpha_2) \cdots (z - \alpha_n)$.

Then $p(S)$ is $\begin{bmatrix} a_n & a_{n-1} & \ddots & \ddots & a_1 \\ 0 & a_n & a_{n-1} & \ddots & \ddots \\ 0 & 0 & a_n & \ddots & \ddots \\ 0 & 0 & 0 & \ddots & a_{n-1} \\ 0 & 0 & 0 & 0 & a_n \end{bmatrix}$.

This can be factorized as $p(S) = (S - \alpha_1 I)(S - \alpha_2 I) \cdots (S - \alpha_n I)$. Let $B_j = S - \alpha_j I$.

Next, define q as the polynomial $\overline{a_n} z^n + \overline{a_{n-1}} z^{n-1} + \cdots + \overline{a_0}$. Note that its roots are $\frac{1}{\overline{\alpha_i}}$. We get $q(z) = (1 - \overline{\alpha_1} z)(1 - \overline{\alpha_2} z) \cdots (1 - \overline{\alpha_n z})$. Also, $q(S) = (I - \overline{\alpha_1} S)(I - \overline{\alpha_2} S) \cdots (I - \overline{\alpha_n} S)$. Let $C_j = I - \overline{\alpha_j} S$.

Let $\underline{H}$ be the Hermitian form $\|q(S)x\|^2 - \|p(S)x\|^2$.

$= \langle q(S)x, q(S)x \rangle - \langle p(S)x, p(S)x \rangle$

$= \langle q^*(S)q(S)x, x \rangle - \langle p^*(S)p(S)x, x \rangle$

$= \langle (q^*(S)q(S) - p^*(S)p(S))x, x \rangle$.

The $n \times n$ matrix corresponding to this Hermitian form is $H = q^*(S)q(S) - p^*(S)p(S)$.

We can now state the Schur-Cohn theorem:

Theorem 12. *The polynomial p, it will have k roots inside the circle, and $n-k$ roots outside the circle iff k eigenvalues of H are positive and $n-k$ are negative.*

4 Proof

We will first prove the Schur-Cohn theorem for $n = 1$, that is for linear polynomials. It will then be extended to polynomials of higher degrees with the help of the Spectral theorem and the Courant-Fischer theorem.

4.1 Linear Polynomial

Let's write $q(S)$ and $p(S)$ as a product of the linear terms. $q(S)^*q(S) - p(S)^*p(S)$
$= (C_1C_2C_3 \dots C_n)^*(C_1C_2C_3 \dots C_n) - (B_1B_2B_3 \dots B_n)^*(B_1B_2B_3 \dots B_n)$

For $n = 1$, this is equal to $C_1^*C_1 - B_1^*B_1$.

$$\begin{aligned}
&C_1^*C_1 - B_1^*B_1 \\
&= (I - \overline{\alpha_1}S)^*(I - \overline{\alpha_1}S) - (S - \alpha_1 I)^*(S - \alpha_1 I) \\
&= (I - \alpha_1 S^*)(I - \overline{\alpha_1}S) - (S^* - \overline{\alpha_1}I)(S - \alpha_1 I) \\
&= (I - \alpha_1 S^* - \overline{\alpha_1}S + |\alpha_1|^2 S^*S) - (S^*S - \alpha_1 S^* - \overline{\alpha_1}S + |\alpha_1|^2 I) \\
&= I - |\alpha_1|^2 I - S^*S + |\alpha_1|^2 S^*S \\
&= (1 - |\alpha_1|^2)(I - S^*S)
\end{aligned}$$

For $n = 1$, $I - S^*S$ is just a 1×1 matrix, so $H = (1 - |\alpha_1|^2)$.

Note that $I - S^*S$ is a posititve definite matrix. If $|\alpha| < 1$, then root of the linear polynomial lies within the unit circle. Also note that H has one negative eigenvalue. Similarly, if $|\alpha| > 1$, then the root of the linear polynomial lies outside the unit circle, and the eigenvalue of H is positive. This shows that the Schur-Cohn theorem is true for $n = 1$. We will now extend the proof for all n.

4.2 Spectral theorem

Theorem 13. *Let $A \in \mathbb{M}_n$ be a Hermitian matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n$. Then A can be written as $U\Lambda U^*$, where U is a unitary matrix, and Λ is a diagonal matrix with real entries.*

Proof. Left to the reader. □

4.3 Courant-Fischer theorem

Theorem 14. *Let $A \in \mathbb{M}_n$ be a Hermitian matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n$. Then*

$$\lambda_k = \min_{\omega_1, \dots, \omega_{n-k} \in \mathbb{C}^n} \max_{\substack{x \neq 0, x \in \mathbb{C}^n \\ x \perp \omega_1, \dots, \omega_{n-k}}} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$$

Proof. If $x \neq 0$, then $\frac{\langle Ax,x\rangle}{\langle x,x\rangle} = \frac{\langle U\Lambda U^* x,x\rangle}{\langle U^*x,U^*x\rangle}$
$= \frac{\langle \Lambda U^*x,U^*x\rangle}{\langle U^*x,U^*x\rangle}$. and $\{U^*x : x \neq 0\} = \{x \in \mathbb{C}^n : x \neq 0\}$

Thus if $\omega_1, \ldots, \omega_{n-k}$ are given, then

$$\sup_{\substack{x\neq 0\\ x\perp\omega_1,\ldots,\ \omega_{n-k}}} \frac{\langle Ax,x\rangle}{\langle x,x\rangle} = \sup_{\substack{y\neq 0\\ y\perp U^*\omega_1,\ldots,\ U^*\omega_{n-k}}} \frac{\langle \Lambda y,y\rangle}{\langle y,y\rangle}$$

$x \perp \omega$ if and only if $y \perp U^*\omega$.

$$= \sup_{\substack{\langle y,y\rangle=1\\ y\perp U^*\omega_1,\ \ldots,\ U^*\omega_{n-k}}} \sum_{i=1}^{n} \lambda_i |y_i|$$

$$\geq \sup_{\substack{\langle y,y\rangle=1\\ y\perp U^*\omega_1,\ \ldots,\ U^*\omega_{n-k}\\ y_1=y_2=\cdots=y_k-1=0}} \sum_{i=1}^{n} \lambda_i |x_i|^2$$

$$= \sup_{\substack{\langle y,y\rangle=1\\ y\perp U^*\omega_1,\ \ldots,\ U^*\omega_{n-k}\\ y_1=y_2=\cdots=y_k-1=0}} \sum_{i=k}^{n} \lambda_i |y_i|^2$$

$$\geq \lambda_k$$

Let $\omega_1 = x_n, \ldots,\ \omega_{n-k} = x_{k+1}$
If $x \perp \omega_i$, as above, then $x = \sum_{i=1}^{k} c_i x_i$.
$\langle Ax, x\rangle = \langle A \sum_{i=1}^{n} c_i X_i, \sum_{i=1}^{n} c_i x_i\rangle$
$= \langle \sum_{i=1}^{n} c_i \lambda_i x_i, \sum_{i=1}^{n} c_i x_i\rangle$
$= \sum_{i=1}^{k} \lambda_i |c_i|^2$
$\leq \lambda_k \sum_{i=1}^{k} |c_i|^2$

$\square$

4.4 Continuing the proof

For general n,

$q^*(S)q(S) - p^*(S)p(S)$
$= (C_1C_2C_3\ldots C_n)^*(C_1C_2C_3\ldots C_n) - (B_1B_2B_3\ldots B_n)^*(B_1B_2B_3\ldots B_n)$
$= C_n^* \cdots C_1^* C_n \cdots C_1 - B_n^* \cdots B_1^* B_n \cdots B_1$
B_i commutes with C_j.
$1 - S^*S = e_1e_1^*$.
$C_j^*C_j - B_j^*B_j = (1-|\alpha_j|^2)(1-S^*S) = (1-|\alpha_j|^2)e_1e_1^*$
We can add and subtract terms to get a telescoping series:
We get $H = \sum_{j=1}^{n}(1-|\alpha_j|^2)V_jV_j^*$, where $V_j = B_1^* \cdots B_{j-1}^* C_{j+1}^* \cdots C_n^* e_1$

5 Extensions

5.1 Arbitrary radius

To find the number of roots of $p(z)$ inside a circle of radius r, repeat the above Schur Cohn theorem but with the polynomial $p(\frac{z}{r})$ instead.

5.2 Limitations

This method fails when the polynomial has one or more roots on the circle.

6 Conclusion

Thank You!

7 Acknowledgements

Prof. Tirthankar Bhattacharya

8 References