Hermitian Forms and Zeros of a Polynomial

Pranshu Gaba

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

December 4, 2017

Abstract

We looked at various concepts in linear algebra such as norms, unitary matrices, and Hermitian matrices, and studied their applications. In particular, we examined the generalized Schur-Cohn theorem, which relates the number of roots of a polynomial lying within and without the unit circle with the parity of eigenvalues of a matrix formed by coefficients of the polynomial.

Acknowledgements

I am grateful to Prof. Tirthankar Bhattacharya for guiding me in this project, introducing me to new concepts and ideas in mathematics, challenging me to improve myself, and for making this an enriching experience altogether.

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

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

Theorem 1. *The operator norm satisfies the triangle inequality* $\|A+B\| \leq \|A\| + \|B\|$.

Proof.

$$\begin{aligned}\|A+B\| = \sup_{x\neq 0} \frac{\|(A+B)x\|}{\|x\|} &\leq \sup_{x\neq 0} \frac{\|Ax\| + \|Bx\|}{\|x\|} \\ &\leq \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|} + \sup_{x\neq 0} \frac{\|Bx\|}{\|x\|} = \|A\| + \|B\|\end{aligned}$$

□

Theorem 2. *The operator norm is submultiplicative,* $\|AB\| \leq \|A\| \|B\|$ *for all square matrices* $A, B \in \mathbb{M}_n$.

Proof.

$$\begin{aligned}\|AB\| &= \sup_{x\neq 0} \frac{\|ABx\|}{\|x\|} \\ &= \sup_{Bx\neq 0} \frac{\|ABx\|}{\|x\|} \\ &= \sup_{Bx\neq 0} \frac{\|ABx\|}{\|Bx\|} \frac{\|Bx\|}{\|x\|} \\ &\leq \sup_{y\neq 0} \frac{\|Ay\|}{\|y\|} \sup_{x\neq 0} \frac{\|Bx\|}{\|x\|} = \|A\| \|B\|\end{aligned}$$

□

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 3. *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 $\|A\| \leq \|A\|_2$. □

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 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 $A > 0$.

For two matrices $A, B \in \mathbb{M}_n$, $A \leq B$ denotes $B - A$ is positive semidefinite. As a result, $\langle (B - A)x, x\rangle \geq 0$, or $\langle Ax, x\rangle \leq \langle Bx, x\rangle$ for all x.

Theorem 4. *Let $A \in \mathbb{M}_n$ be a positive semidefinite matrix. Then all eigenvalues of A are non-negative.*

Proof. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A, and let $x_1, x_2, \ldots, x_n$ be the corresponding eigenvectors, that is: $Ax_i = \lambda x_i$.

For any eigenvector x_i, the inner product $\langle Ax_i, x_i\rangle = \langle \lambda_i x_i, x_i\rangle$. Since λ_i is a scalar, it can come out of the inner product, so $\langle \lambda_i x_i, x_i\rangle = \lambda_i\langle x_i, x_i\rangle$. We get

$$\lambda_i = \frac{\langle Ax_i, x_i\rangle}{\langle x_i, x_i\rangle}$$

Here $\langle Ax_i, x_i\rangle$ 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. □

2.4 Adjoint

Let $A \in \mathbb{M}_n$. The adjoint of matrix A, denoted by A^*, is the matrix that satisfies $\langle A^*x, y\rangle = \langle x, Ay\rangle$. The adjoint of a matrix, A^*, like A, represents a linear transformation on $\mathbb{C}^n$.

Theorem 5. *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 $A^* = (\overline{A})^T$. $\square$

Remark. The adjoint of the adjoint of a matrix is the original matrix itself, $(A^*)^* = A$.

Theorem 6. *The norm of a matrix is equal to the norm of its adjoint.* $\|A\| = \|A^*\|$

Proof.

$$\|A\|^2 = \sup_{\|x\|=1} \|Ax\|^2 = \sup_{\|x\|=1} \langle Ax, Ax\rangle = \sup_{\|x\|=1} \langle A^*Ax, x\rangle = \|A^*A\|$$

$$\|A^*\|^2 = \sup_{\|x\|=1} \|A^*x\|^2 = \sup_{\|x\|=1} \langle A^*x, A^*x\rangle = \sup_{\|x\|=1} \langle AA^*x, x\rangle = \|AA^*\| \leq \|A\| \|A^*\|$$

However, using Theorem 2, we get $\|A^*A\| \leq \|A^*\| \|A\|$ in the first equation. Hence, $\|A\|^2 \leq \|A^*\| \|A\|$, which implies $\|A\| \leq \|A^*\|$. Similarly, in the second equation we get $\|A^*\| \leq \|A\|$.

Since both these inequalities are true, it necessarily means that $\|A\| = \|A^*\|$. $\square$

Theorem 7. *A^*A and AA^* are positive semidefinite for all $A \in \mathbb{M}_n$.*

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

2.5 Unitary Matrix

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

Theorem 8. *The columns of a unitary matrix U form an orthonormal basis.*

Proof. Let u_i be the i^{th} column of U. Then $U = \begin{bmatrix} u_1 & u_2 & u_3 & \cdots & u_n \end{bmatrix}$ and $U^* = \begin{bmatrix} u_1^* \\ u_2^* \\ u_3^* \\ \vdots \\ u_n^* \end{bmatrix}$.

Using the fact that U is a unitary matrix, $U^*U = I$:

$$\begin{bmatrix} u_1^* \\ u_2^* \\ u_3^* \\ \vdots \\ u_n^* \end{bmatrix} \begin{bmatrix} u_1 & u_2 & u_3 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

We see that $u_i^*u_j = \langle u_i, u_j\rangle = \delta_{ij}$. Here δ_{ij} is the Kronecker Delta function which is defined as

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Each column has norm 1 since $\langle u_i, u_i\rangle = \|u_i\|^2 = 1$, and any two columns are orthogonal $\langle u_i, u_j\rangle = 0$. Hence the column vectors $u_1, u_2, \ldots, u_n$ of a unitary matrix form an orthonormal basis. $\square$

Theorem 9. *A unitary matrix preserves inner product:* $\langle Ux, Uy\rangle = \langle x, y\rangle$ *for all* $x, y \in \mathbb{C}_n$.

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 10. *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. The sum of diagonal elements of A^*A is

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

□

2.7 Hermitian Matrix

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

Theorem 11. *Let* $A \in \mathbb{M}_n$ *be a Hermitian matrix. Then all eigenvalues of* A *are real.*

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

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

This implies $\lambda = \overline{\lambda}$ for all v. Hence, λ is real. All eigenvalues of a Hermitian matrix are real. □

Theorem 12. *Eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Proof. Suppose λ and μ are distinct eigenvalues of A. Suppose the corresponding eigenvectors are u and v respectively. We have $Au = \lambda u$ and $Av = \mu v$. Since A is Hermitian,

$$\lambda\langle u, v\rangle = \langle \lambda u, v\rangle = \langle Au, v\rangle = \langle u, Av\rangle = \langle u, \mu v\rangle = \mu\langle u, v\rangle$$

Since $\lambda \neq \mu$, $\langle u, v\rangle$ must be zero. Eigenvectors u and v are orthogonal.

□

A diagonal matrix is a matrix whose non-diagonal entries are all zero. $a_{ij} = 0$ if $i \neq j$

Theorem 13 (Spectral theorem)**.** *For every Hermitian matrix A, there exists a unitary matrix U such that $A = U\Lambda U^*$, where Λ is the diagonal matrix* $\begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ *containing eigenvalues of A.*

Proof. The matrix A is a linear transformation T with respect to the standard basis I. The matrix U is a unitary matrix, so its columns $u_1, u_2, \ldots, u_n$ form an orthonormal basis for Λ. Since U is a unitary vector, $U^* = U^{-1}$. This means Λ is the matrix of the same transformation T with respect to a different basis U, and that A and Λ have the same eigenvalues.

Start with any eigenvalue λ_1 of A. Let x_1 be a unit eigenvector corresponding to λ_1.

The basis transforms from $x_i \mapsto U^* x_i$

$AU = U\Lambda$

□

Let A be Hermitian. Its eigenvalues are $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Using spectral theorem,

$$\frac{\langle Ax, x\rangle}{\langle x, x\rangle} = \frac{\langle U\Lambda U^* x, x\rangle}{\langle x, x\rangle} = \frac{\langle \Lambda y, y\rangle}{\langle y, y\rangle} = \frac{\sum \lambda_i |y_i|^2}{\sum |y_i|^2}$$

$\{x \in \mathbb{C}^n : x \neq 0\} = \{y \in \mathbb{C}^n : y = U^* x \text{ for } x \neq 0\}$

We get the following inequality.

$$\lambda_1 = \lambda_1 \frac{\sum |y_i|^2}{\sum |y_i|^2} \leq \frac{\sum \lambda_i |y_i|^2}{\sum |y_i|^2} \leq \lambda_n \frac{\sum |y_i|^2}{\sum |y_i|^2} = \lambda_n$$

Therefore

$$\lambda_1 \leq \frac{\langle Ax, x\rangle}{\langle x, x\rangle} \leq \lambda_n$$

$$\lambda_1 = \inf_{x \neq 0} \frac{\langle Ax, x\rangle}{\langle x, x\rangle}, \quad \lambda_n = \sup_{x \neq 0} \frac{\langle Ax, x\rangle}{\langle x, x\rangle}$$

Theorem 14. *If $A \in \mathbb{M}_n$ is Hermitian, then $\langle Ax, x\rangle$ is real for all x.*

Proof. $\langle Ax, x\rangle = \langle x, A^* x\rangle = \langle x, Ax\rangle = \overline{\langle Ax, x\rangle}$. □

The converse of this theorem is also true.

Theorem 15 (Converse of Theorem 14)**.** *If $A \in \mathbb{M}_n$ and $\langle Ax, x\rangle \in \mathbb{R}$ for every $x \in \mathbb{C}_n$, then A is a Hermitian matrix.*

Proof. Let $\alpha \in \mathbb{C}$ and $h, g \in \mathbb{C}^n$. Then

$$\langle A(h + \alpha g), h + \alpha g\rangle = \langle Ah, h\rangle + \alpha\langle Ag, h\rangle + \overline{\alpha}\langle Ah, g\rangle + |\alpha|^2 \langle Ag, g\rangle$$

The first and the last terms $\langle Ah, h\rangle$ and $|\alpha|^2 \langle Ag, g\rangle$ are both real. We now look at the sum of second and third terms:

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

We can substitute values of $\alpha = 1$ and $\alpha = i$ in this equation to get a system of linear equations.

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

Solving this system gives us $\langle Ag, h\rangle = \langle g, Ah\rangle$, which is equal to $\langle A^* g, h\rangle$. Hence $Ag = A^* g$ for all g, and $A = A^*$. We see that A is a Hermitian matrix. □

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

Proof. A is a positive semidefinite matrix, so $\langle Ax, x\rangle \geq 0$, so $\langle Ax, x\rangle \in \mathbb{R}$ for all x. By Theorem 15, A is Hermitian. □

Theorem 16. *If A is Hermitian, then $\|A\| = \sup\{|\langle Ah, h\rangle| : \|h\| = 1\}$.*

Proof. By Cauchy Schwarz, $\|Ah\| \, \|h\| \geq |\langle Ah, h\rangle|$. For $\|h\| = 1$, we have $\|A\| \geq |\langle Ah, h\rangle|$. Now we will prove the reverse inequality.

Let $M = \sup\{|\langle Ah, h\rangle| : \|h\| = 1\}$. If $h, g \in \mathbb{C}^n$ with $\|h\| = \|g\| = 1$, then

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

From this, we get a bound on $\mathrm{Re}\langle Ah, g\rangle$:

$$\begin{aligned}
4\mathrm{Re}\langle Ah, g\rangle &= \langle A(h+g), h+g\rangle - \langle A(h-g), h-g\rangle \\
&\leq |\langle A(h+g), h+g\rangle| + |\langle A(h-g), h-g\rangle| \\
&\leq \left|\langle A\frac{h+g}{\|h+g\|}, \frac{h+g}{\|h+g\|}\rangle\right| \|h+g\|^2 + \left|\langle A\frac{h-g}{\|h-g\|}, \frac{h-g}{\|h-g\|}\rangle\right| \|h-g\|^2 \\
&\leq M(\|h+g\|^2 + \|h-g\|^2) \\
&= M(2\|h\|^2 + 2\|g\|^2) = 4M
\end{aligned}$$

We get $\mathrm{Re}\langle Ah, g\rangle \leq M$ for any $h, g \in \mathbb{C}^n$ with $\|h\| = \|g\| = 1$.

If $\langle Ah, g\rangle$ is not real, then we can write it as $\langle Ah, g\rangle = e^{i\theta} |\langle Ah, g\rangle|$. We get $|\langle Ah, g\rangle| = e^{-i\theta}\langle Ah, g\rangle$ is real, and is equal to $\langle Ae^{-i\theta}h, g\rangle$. Since $\left\|e^{-i\theta}h\right\| = 1$, the above inequality still holds:

$$|\langle Ah, g\rangle| = \langle Ae^{-i\theta}h, g\rangle = \mathrm{Re}\langle Ae^{-i\theta}h, g\rangle \leq M$$

We also have the following relation

$$\|A\| = \sup_{\|h\|=1} \|Ah\| = \sup_{\|h=1\|} \sup_{\|g\|=1} |\langle Ah, g\rangle|$$

Therefore $|\langle Ah, g\rangle| \leq M$ implies $\|A\| \leq M$. Since $\|A\| \geq M$ and $\|A\| \leq M$ both hold true, it means that $\|A\| = M = \sup\{|\langle Ah, h\rangle| : \|h\| = 1\}$. □

Corollary. *If $\langle Ax, x\rangle = 0$ for all x, then $A = 0$.*

Proof. Since $\langle Ax, x\rangle = 0$ for all x, we have $\sup|\langle Ax, x\rangle| = 0$. By Theorem 16,$\|A\| = 0$. This is true only when $A = 0$. □

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

Proof. By the Corollary in Theorem 15, every positive semidefinite matrix is Hermitian, and for a positive semidefinite matrix, $|\langle Ax, x\rangle| = \langle Ax, x\rangle$. Hence, $\|A\| = \sup\limits_{\|x\|=1} \langle Ax, x\rangle$ □

Theorem 17 (Courant-Fischer)**.** *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*

$$\begin{aligned}\lambda_k &= \min_{w_1,\ldots,w_{n-k}\in\mathbb{C}^n} \max_{\substack{x\neq 0, x\in\mathbb{C}^n\\ x\perp w_1,\ldots,w_{n-k}}} \frac{\langle Ax, x\rangle}{\langle x, x\rangle} \\ &= \max_{w_1,\ldots,w_{k-1}\in\mathbb{C}^n} \min_{\substack{x\neq 0, x\in\mathbb{C}^n\\ x\perp w_1,\ldots,w_{k-1}}} \frac{\langle Ax, x\rangle}{\langle x, x\rangle}\end{aligned}$$

Proof. If $x \neq 0$, then

$$\frac{\langle Ax, x\rangle}{\langle x, x\rangle} = \frac{\langle U\Lambda U^* x, x\rangle}{\langle U^*x, U^*x\rangle} = \frac{\langle \Lambda U^* x, U^* x\rangle}{\langle U^*x, U^*x\rangle}$$

Since U is a unitary matrix, its columns form an orthonormal basis. U^*x spans and $\{U^*x \mid x \neq 0\} = \{x \in \mathbb{C}^n \mid x \neq 0\}$

Thus if $w_1, \ldots, w_{n-k}$ are given, then

$$\begin{aligned}\sup_{\substack{x\neq 0\\ x\perp w_1,\ldots,\ w_{n-k}}} \frac{\langle Ax, x\rangle}{\langle x, x\rangle} &= \sup_{\substack{y\neq 0\\ y\perp U^*w_1,\ldots,\ U^*w_{n-k}}} \frac{\langle \Lambda y, y\rangle}{\langle y, y\rangle} \\ &= \sup_{\substack{\langle y,y\rangle=1\\ y\perp U^*w_1,\ \ldots,\ U^*w_{n-k}}} \sum_{i=1}^{n} \lambda_i\, |y_i|^2 \\ &\geq \sup_{\substack{\langle y,y\rangle=1\\ y\perp U^*w_1,\ \ldots,\ U^*w_{n-k}\\ y_1=y_2=\cdots=y_{k-1}=0}} \sum_{i=1}^{n} \lambda_i\, |x_i|^2 \\ &= \sup_{\substack{\langle y,y\rangle=1\\ y\perp U^*w_1,\ \ldots,\ U^*w_{n-k}\\ y_1=y_2=\cdots=y_{k-1}=0}} \sum_{i=k}^{n} \lambda_i\, |y_i|^2 \\ &\geq \lambda_k\end{aligned}$$

Let $w_1 = x_n,\ w_2 = x_{n-1},\ \ldots,\ w_{n-k} = x_{k+1}$

If $x \perp w_i$, as above, then $x = \sum_{i=1}^{k} c_i x_i$.

$$
\begin{aligned}
\langle Ax, x\rangle &= \langle A\sum_{i=1}^{n} c_i x_i, \sum_{i=1}^{n} c_i x_i\rangle \\
&= \langle \sum_{i=1}^{n} c_i \lambda_i x_i, \sum_{i=1}^{n} c_i x_i\rangle \\
&= \sum_{i=1}^{k} \lambda_i \left|c_i\right|^2 \\
&\leq \lambda_k \sum_{i=1}^{k} \left|c_i\right|^2
\end{aligned}
$$

□

2.8 Projectors

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

Theorem 18. *There exists a subspace M of $\mathbb{C}^n$ such that*

$$
\begin{cases} Pm = m & \forall m \in M \\ Px = 0 & \forall x \in M^{\perp} \end{cases}
$$

Proof. range$(P) = M$ and range$(I - P) = M^{\perp}$ satisfy the above properties. These subspaces are orthogonal. Let $Pv \in M$, $w - Pw \in M^{\perp}$.

$$\langle w - Pw, Pv\rangle = \langle w, Pv\rangle - \langle Pw, Pv\rangle = \langle w, Pv\rangle - \langle w, P^* Pv\rangle = \langle w, Pv\rangle - \langle w, P^2 v\rangle = \langle w, Pv\rangle - \langle w, Pv\rangle = 0$$

Let $m \in$ range(P), that is, $m = Pv$ for some $v \in \mathbb{C}^n$. Then $Pm = P^2 v = Pv = m$.

Now let $x \in$ range$(I - P)$, that is, $x = (I - P) = v - Pv$ for some $v \in \mathbb{C}^n$. $Px = Pv - P^2 v = Pv - Pv = 0$. □

2.9 Shift Matrix

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

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

The following matrices will be useful in proving the Schur-Cohn theorem. Note that $I - S^* S = e_1 e_1^*$ is a projector matrix.

$$
S^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad S^* S = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \quad I - S^* S = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}
$$

3 Schur-Cohn Theorem

3.1 Theorem

For any polynomial

$$p(z) = a_0 z^n + a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_{n-1} z + a_n, \quad a_i \in \mathbb{C}$$

such that none of the roots lie on the unit circle $|z| = 1$, and $p(0) \neq 0$, we want to find out how many of its roots lie inside the unit circle ($|z| < 1$) and how many roots lie outside ($|z| > 1$).

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 $p(z) = (z - \alpha_1)(z - \alpha_2) \cdots (z - \alpha_n)$.

Note that $p(S)$ is $\begin{bmatrix} a_n & a_{n-1} & a_{n-2} & \cdots & a_1 \\ 0 & a_n & a_{n-1} & \cdots & a_2 \\ 0 & 0 & a_n & \cdots & a_3 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_n \end{bmatrix}$, where S is the shift matrix.

The matrix $p(S)$ can be factorized as

$$p(S) = (S - \alpha_1 I)(S - \alpha_2 I) \cdots (S - \alpha_n I)$$

Denote $S - \alpha_j I$ by B_j. Then $p(S) = \displaystyle\prod_{j=1}^{n} B_j = B_1 B_2 B_3 \cdots B_n$

Next, define q to be the polynomial

$$q(z) = \overline{a_n} z^n + \overline{a_{n-1}} z^{n-1} + \cdots + \overline{a_0}$$

The roots of $q(S)$ are $\frac{1}{\overline{\alpha_i}}$.

$$\begin{aligned} q\left(\frac{1}{\overline{\alpha_i}}\right) &= \frac{\overline{a_n}}{\overline{\alpha_i}^n} + \frac{\overline{a_{n-1}}}{\overline{\alpha_i}^{n-1}} + \cdots + \overline{a_0} \\ &= \frac{1}{\overline{\alpha_i}^n}(\overline{a_n} + \overline{a_{n-1}\alpha_i} + \cdots + \overline{a_0 \alpha_i}^n) \\ &= \frac{\overline{p(\alpha_i)}}{\overline{\alpha_i}^n} = 0 \end{aligned}$$

$q(z)$ can be factorized as $q(z) = (1 - \overline{\alpha_1} z)(1 - \overline{\alpha_2} z) \cdots (1 - \overline{\alpha_n} z)$.

The matrix $q(S)$ can be factorized as

$$q(S) = (I - \overline{\alpha_1} S)(I - \overline{\alpha_2} S) \cdots (I - \overline{\alpha_n} S)$$

Denote $I - \overline{\alpha_j} S$ by C_j. Then $q(S) = \displaystyle\prod_{j=1}^{n} C_j = C_1 C_2 C_3 \cdots C_n$.

We now state the Schur-Cohn theorem:

Theorem 19 (Schur)**.** *Consider the matrix*

$$H = q^*(S) q(S) - p^*(S) p(S)$$

The polynomial p will have all its roots inside the unit circle $|z| = 1$ if and only if H is positive definite. It will have all the roots outside the unit circle if and only if H is negative definite.

Theorem 20 (Cohn Generalization)**.** *The polynomial p, it will have k roots inside the circle $|z| = 1$, and $n - k$ roots outside the circle if and only if k eigenvalues of H are positive and $n - k$ are negative.*

3.2 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.

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 $C_1^*C_1 - B_1^*B_1$.

$$\begin{aligned} C_1^*C_1 - B_1^*B_1 &= (I - \overline{\alpha_1}S)^*(I - \overline{\alpha_1}S) - (S - \alpha_1 I)^*(S - \alpha_1 I) \\ &= (I - \alpha_1 S^*)(I - \overline{\alpha_1}S) - (S^* - \overline{\alpha_1}I)(S - \alpha_1 I) \\ &= (I - \alpha_1 S^* - \overline{\alpha_1}S + |\alpha_1|^2 S^*S) - (S^*S - \alpha_1 S^* - \overline{\alpha_1}S + |\alpha_1|^2 I) \\ &= I - |\alpha_1|^2 I - S^*S + |\alpha_1|^2 S^*S \\ &= (1 - |\alpha_1|^2)(I - S^*S) \\ &= (1 - |\alpha_1|^2)e_1e_1^* \end{aligned}$$

For $n = 1$, $I - S^*S$ is a 1×1 matrix, so $H = \left[1 - |\alpha_1|^2\right]$.
Note that $I - S^*S$ is a posititve definite matrix of rank 1.

- If $|\alpha_1| < 1$, then the root of the linear polynomial lies within the unit circle. Matrix H has one positive eigenvalue.
- Similarly, if $|\alpha_1| > 1$, then the root of the linear polynomial lies outside the unit circle, and the eigenvalue of H is negative.

This shows that the Schur-Cohn theorem is true for $n = 1$. We will now extend the proof for all n.

$$\begin{aligned} H &= q^*(S)q(S) - p^*(S)p(S) \\ &= (C_1C_2C_3 \dots C_n)^*(C_1C_2C_3 \dots C_n) - (B_1B_2B_3 \dots B_n)^*(B_1B_2B_3 \dots B_n) \\ &= (C_n^* \cdots C_1^*)(C_1 \cdots C_n) - (B_n^* \cdots B_1^*)(B_1 \cdots B_n) \end{aligned}$$

B_i commutes with C_j, so we can rearrange their order in each term according to our convenience. We can add and subtract terms to get a telescoping series:

$$\begin{array}{rccc} H = & (C_n^* \cdots C_2^*)C_1^*C_1(C_2 \cdots C_n) & - & (C_n^* \cdots C_2^*)B_1^*B_1(C_2 \cdots C_n) \\ + & B_1^*(C_n^* \cdots C_3^*)C_2^*C_2(C_2 \cdots C_n)B_1 & - & B_1^*(C_n^* \cdots C_3^*)B_2^*B_2(C_3 \cdots C_n)B_1 \\ + & B_1^*B_2^*(C_n^* \cdots C_4^*)C_3^*C_3(C_4 \cdots C_n)B_2B_1 & - & B_1^*B_2^*(C_n^* \cdots C_4^*)B_3^*B_3(C_4 \cdots C_n)B_2B_1 \\ + & \vdots & - & \vdots \\ + & \vdots & - & \vdots \\ + & (B_1^*B_2^* \cdots B_{n-1}^*)C_n^*C_n(B_{n-1} \cdots B_2B_1) & - & (B_1^* \cdots B_{n-1}^*)B_n^*B_n(B_{n-1} \cdots B_1) \end{array}$$

$$= \sum_{j=1}^{n} (B_1^*B_2^* \cdots B_{j-1}^*)(C_n^* \cdots C_{j+1}^*)(C_j^*C_j - B_j^*B_j)(C_{j+1} \cdots C_n)(B_{j-1} \cdots B_2B_1)$$

In the case of p being a linear polynomial, we saw that $C_j^* C_j - B_j^* B_j = (1 - |\alpha_j|^2)(I - S^* S) = (1 - |\alpha_j|^2) e_1 e_1^*$. Hence,

$$\begin{aligned} H &= \sum_{j=1}^{n} (B_1^* B_2^* \cdots B_{j-1}^*)(C_n^* \cdots C_{j+1}^*)((1 - |\alpha_j|^2) e_1 e_1^*)(C_{j+1} \cdots C_n)(B_{j-1} \cdots B_2 B_1) \\ &= \sum_{j=1}^{n} (1 - |\alpha_j|^2) v_j v_j^*, \quad \text{ where } v_j = B_1^* \cdots B_{j-1}^* C_{j+1}^* \cdots C_n^* e_1. \end{aligned}$$

Let V be the matrix with columns v_i, let D be the diagonal matrix with $d_{ii} = 1 - |\alpha_i|^2$,

$$V = \begin{bmatrix} v_1 & v_2 & v_3 & \cdots & v_n \end{bmatrix}$$

$$D = \begin{bmatrix} 1 - |\alpha_1|^2 & 0 & 0 & \cdots & 0 \\ 0 & 1 - |\alpha_2|^2 & 0 & \cdots & 0 \\ 0 & 0 & 1 - |\alpha_3|^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 - |\alpha_n|^2 \end{bmatrix}$$

then

$$H = VDV^*$$

If matrix H is invertible, then matrix V is also invertible.

Lemma. *H is invertible if and only if D is invertible*

Proof. Let H be positive definite ($\Rightarrow$). Then

$$\begin{aligned} \langle Dx, x \rangle &= \langle V^{-1} H (V^*)^{-1} x, x \rangle \\ &= \langle (V^{-1}) H (V^{-1})^*, x \rangle \\ &= \langle H (V^{-1})^*, (V^{-1})^* x \rangle \\ &= \langle Hy, y \rangle \\ &= \sum_{j=1}^{n} (1 - |\alpha_j|^2) |y_j|^2 > 0 \end{aligned}$$

Let D be positive definite ($\Leftarrow$). Then

$$\begin{aligned} \langle Hx, x \rangle &= \langle VDV^* x, x \rangle \\ &= \langle DV^* x, V^* x \rangle \\ &= \langle Dy, y \rangle \\ &= \sum_{j=1}^{n} (1 - |\alpha_j|^2) |y_j|^2 > 0 \end{aligned}$$

$\square$

Hence if all the roots of $p(z)$ lie within the unit circle, then every eigenvalue of D is positive, so every eigenvalue of H is also positive. Similarly if all the roots of $p(z)$ lie outside the unit circle, then every eigenvalue of H is negative.

We will now look at the generalized version by Cohn, where some roots are inside the circle and some roots are outside.

Let the eigenvalues of H be $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k \leq \cdots \leq \lambda_n$. From the Spectral theorem, we get the values of the lowest and the highest eigenvalues.

$$\lambda_1 = \inf_{x \neq 0} \frac{\langle Hx, x \rangle}{\langle x, x \rangle}, \quad \lambda_n = \sup_{x \neq 0} \frac{\langle Hx, x \rangle}{\langle x, x \rangle}$$

To find the intermediate eigenvalues, the Courant-Fischer theorem will be required.

Let λ_k be the smallest positive eigenvalue. Then given any $w_1, w_2, \ldots, w_{n-k} \in \mathbb{C}^n$, we have some $x \neq 0$ such that x is orthogonal to $w_1, w_2, \ldots, w_{n-k}$ and $\langle Dx, x \rangle > 0$.

$$\langle Hx, x \rangle = \langle VDV^* x, x \rangle = \langle DV^* x, V^* x \rangle$$

Let W be the subspace spanned by $w_1, w_2, \ldots, w_{n-k}$. Now let $y \in W^\perp$ so $\langle y, w_j \rangle = 0$ for all j. Let $x = V^* y$. Then $\langle (V^*)^{-1} x, w_i \rangle = 0$, which implies $\langle x, V^{-1} w_i \rangle = 0$. This implies x is orthogonal to all of $V^{-1} w_1, V^{-1} w_2, \ldots, V^{-1} w_{n-k}$.

Hence,

$$\langle Dx, x \rangle > 0 \implies \langle V^{-1} H (V^*)^{-1} x, x \rangle > 0 \implies \langle Hy, y \rangle > 0$$

This means for every postive eigenvalue of D, there is a positive eigenvalue of H. Similarly, for every negative eigenvalue of D, there is a negative eigenvalue of H.

- Thus, the number of roots inside the unit circle is equal to the number of positive eigenvalues of D, which is equal to the number of positive eigenvalue of H.
- Analogously, the number of roots outside the unit circle is equal to the number of negative eigenvalues of D, which is equal to the number of negative eigenvalues of H.

This concludes the proof of the generalized Schur-Cohn theorem. $\square$

3.3 Example

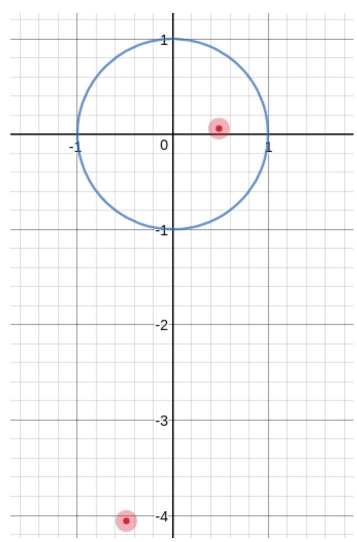
We will apply the Schur-Cohn theorem on the polynomial $p(z) = x^2 + 4ix - 2i$.

$$p(S) = \begin{bmatrix} -2i & 4i \\ 0 & -2i \end{bmatrix}, \quad p^*(S) = \begin{bmatrix} 2i & 0 \\ -4i & 2i \end{bmatrix}, \quad p^*(S)p(S) = \begin{bmatrix} 20 & -8 \\ -8 & 4 \end{bmatrix}$$
$$q(S) = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}, \quad q^*(S) = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}, \quad q^*(S)q(S) = \begin{bmatrix} 17 & 4 \\ 4 & 1 \end{bmatrix}$$

We get matrix H as

$$H = q^*(S)q(S) - p^*(S)p(S) = \begin{bmatrix} -3 & 12 \\ 12 & -3 \end{bmatrix} = 3 \begin{bmatrix} -1 & 4 \\ 4 & -1 \end{bmatrix}$$

The eigenvalues of this matrix are -5 and 3. Since one eigenvalue is positive and one is negative, by Schur-Cohn theorem, it means that the polynomial has one root inside the unit circle and one root outside the unit circle. Indeed when the roots are plotted, this result is verified:



4 Conclusion

The algorithm described above was to determine the location of roots with respect to the unit circle. It can be modified to find the location of roots with respect to a circle of any radius about the origin. 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. This algorithm can be even used to find the number in one half plane of $\mathbb{C}$.

This method fails when the polynomial has one or more roots on the circle, or if there is a conjugate pair of roots. H becomes singular and the algorithm fails. To avoid this, compute the eigenvalues at a circle of a different radius. More workarounds are presented in these papers: Modified versions of this algorithm exist, which are faster but only state if all the roots are within the unit circle or not.

This algorithm has a wide range of applications such as control theory, optimization and digital systems processing, where it is required that all the roots lie in a stability radius of a given point.

References