

Lecture # 12

Inverse Iteration, Gerschgorin Disks, and Sturm Sequences

We begin with an important perturbation theorem for eigenvalues.

Theorem 1 (Gerschgorin Circle Theorem) *Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of $A \in \mathbf{C}^{n \times n}$. Define $\lambda(A) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ as the spectrum of A . Let the row circles of A in the complex plane be given by*

$$R_i = \{z \in \mathbf{C} : |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|\}.$$

Let the column circles of A be

$$C_j = \{z \in \mathbf{C} : |z - a_{jj}| \leq \sum_{i \neq j} |a_{ij}|\}.$$

Then

$$\lambda(A) \subseteq \bigcup_{i=1}^n R_i, \quad \lambda(A) \subseteq \bigcup_{j=1}^n C_j.$$

Also, each connected subset of $\bigcup_{i=1}^n R_i$ and $\bigcup_{j=1}^n C_j$ contains as many eigenvalues as circles.

We prove all but the last statement about connected subsets (which requires a sophisticated continuity argument). Let (λ, \mathbf{x}) be an eigenvalue/eigenvector pair of A . Normalized \mathbf{x} so that $\|\mathbf{x}\|_\infty = 1$. Then

$$(A - \lambda I)\mathbf{x} = 0. \tag{1}$$

Let I be the largest component of \mathbf{x} . Then row I of (1) reads

$$(a_{II} - \lambda)x_I + \sum_{j \neq i} a_{Ij}x_j = 0.$$

Thus

$$|a_{II} - \lambda||x_I| \leq \sum_{j \neq i} |a_{Ij}||x_j|.$$

Since $|x_I| = \|\mathbf{x}\|_\infty = 1$, and $|x_J| \leq \|\mathbf{x}\|_\infty$ we have

$$|a_{II} - \lambda| \leq \sum_{j \neq i} |a_{Ij}|.$$

Thus λ is in the I row circle. Since A^T has the same eigenvalues as A , λ must also be in one of the column circles.

Now we give an example.

Example 1 *Let*

$$A = \begin{pmatrix} 2 & 10^{-10} & 10^{-6} \\ 10^{-3} & 1 & 10^{-4} \\ 10^{-6} & 10^{-8} & 3 \end{pmatrix}.$$

The row circles are

$$\begin{aligned} |z - 2| &\leq 10^{-10} + 10^{-6} \leq 1.0001 \cdot 10^{-6} \\ |z - 1| &\leq 10^{-3} + 10^{-4} \leq 1.1 \cdot 10^{-3} \\ |z - 3| &\leq 10^{-6} + 10^{-8} \leq 1.01 \cdot 10^{-6} \end{aligned}$$

The column circles are

$$\begin{aligned} |z - 2| &\leq 10^{-3} + 10^{-6} \leq 1.001 \cdot 10^{-3} \\ |z - 1| &\leq 10^{-8} + 10^{-10} \leq 1.01 \cdot 10^{-8} \\ |z - 3| &\leq 10^{-4} + 10^{-6} \leq 1.01 \cdot 10^{-4} \end{aligned}$$

The bound from the row circles is best for the eigenvalues near 2 and 3, bound from the column circles is better for the eigenvalue near 1.

According to MATLAB, the actual eigenvalues are a bit closer. They are

$$2 - 8.9972e - 13, 1 - 5.9985e - 13, 3 + 1.5001e - 12.$$

Notice that, since the Gershgorin circles are all disjoint, and A is a real matrix, then A must have real eigenvalues.

Note also that for a symmetric matrix, the row circles and column circles are the same. Moreover, instead of being circles in the complex plane, the Gerschgorin regions will just be intervals on the real line!

We will focus now on the symmetric eigenvalue problem, and we will take a significant departure from the text. Let

$$A = A^T \in \mathbf{R}^{n \times n}$$

be symmetric. We can diagonalize it as follows

$$\begin{aligned} A &= Q\Lambda Q^T, \quad Q^T Q = Q Q^T = I_n, \\ \Lambda &= \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \quad \lambda_n \geq \lambda_{n-1} \geq \dots \geq \lambda_1. \end{aligned}$$

The symmetric eigenvalue problem has a lot of structure; we have elegant algorithms.

The basic inverse iteration routine is

```

w0 initial guess,  $\mu$  shift,  $\lambda^{(0)} = \mu$ ;
Solve  $(A - \mu I)\mathbf{y}_0 = \mathbf{w}_0$ ;
 $\lambda^{(1)} = \mu + 1/\mathbf{y}_0^T \mathbf{w}_0$ ;
 $\mathbf{w}_1 = \mathbf{y}_0/\|\mathbf{y}_0\|_2$ ;  $k = 1$ ;
while  $|\lambda^{(k)} - \lambda^{(k-1)}| > \epsilon$ 
    Solve  $(A - \mu I)\mathbf{y}_k = \mathbf{w}_k$ ;
     $\lambda^{(k+1)} = \mu + 1/\mathbf{y}_k^T \mathbf{w}_k$ ;
     $\mathbf{w}_{k+1} = \mathbf{y}_k/\|\mathbf{y}_k\|_2$ ;
end;

```

We need to improve this in two different ways.

1. Factoring $A - \mu I$ using either LU or QR factorization is expensive. It is especially so if we change μ since that will require a new factorization.
2. We need to choose μ to be close to an eigenvalue, but how do we know enough to get it close to an eigenvalue? Can we get it closer?

As we showed last time, A can be reduced to tridiagonal form. That is, there is an orthogonal Q_0 such that

$$T = Q_0^T A Q_0$$

where T has the

$$T = \begin{pmatrix} \alpha_1 & \beta_1 & 0 & 0 & 0 & 0 \\ \beta_1 & \alpha_2 & \beta_2 & 0 & 0 & 0 \\ 0 & \beta_2 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \beta_{n-1} \\ \cdots & \cdots & \cdots & \cdots & \beta_{n-1} & \alpha_n \end{pmatrix}$$

In fact, we can parameterize this matrix in terms of the $2n - 1$ parameters, $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_{n-1}$.

The Q-R decomposition of a tridiagonal matrix is simple. To get

$$T - \mu I = Q_T R_T$$

we have that Q is the product of $n - 1$ Givens rotations

$$Q = G_1^T \cdots G_{n-1}^T$$

where G_k rotates rows k and $k + 1$, zeroing out β_k . The upper triangular matrix R has the form

$$R = \begin{pmatrix} x & x & x & & \\ & x & x & x & \\ & & x & x & x \\ & & & x & x \\ & & & & x \end{pmatrix}$$

There is also a simple bound for the eigenvalues of T from the Gerschgorin circles of T . The row circles are

$$R_i = \{\lambda \in \mathbf{R} : \alpha_i - |\beta_{i-1}| - |\beta_i| \leq \lambda \leq \alpha_i + |\beta_{i-1}| + |\beta_i|\}.$$

If $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, then

$$\begin{aligned} \lambda_1 &\geq \min_{1 \leq i \leq n} \alpha_i - |\beta_{i-1}| - |\beta_i| = \lambda_{min} \\ \lambda_n &\leq \max_{1 \leq i \leq n} \alpha_i + |\beta_{i-1}| + |\beta_i| = \lambda_{max} \end{aligned}$$

Thus

$$\lambda(A) \in [\lambda_{min}, \lambda_{max}].$$

For a symmetric tridiagonal matrix, it is possible to compute the characteristic polynomial. Let

$$\begin{aligned} T_1 &= (\alpha_1) \\ T_k &= \begin{pmatrix} \alpha_1 & \beta_1 & 0 & 0 & 0 & 0 \\ \beta_1 & \alpha_2 & \beta_2 & 0 & 0 & 0 \\ 0 & \beta_2 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \beta_{k-1} \\ \cdots & \cdots & \cdots & \cdots & \beta_{k-1} & \alpha_k \end{pmatrix} \end{aligned}$$

where $k = 1, 2, \dots, n$. Let $p_0(\lambda) = 1$ and let

$$p_k(\lambda) = \det(T_k - \lambda I).$$

Thus $p_k(\lambda)$ is the characteristic polynomial of T_k . A recursion that uses expansion by minors is

$$\det(T_n - \lambda I) = (\alpha_n - \lambda) \det(T_{n-1} - \lambda I) - \beta_{n-1}^2 \det(T_{n-2} - \lambda I)$$

so that

$$p_n(\lambda) = (\alpha_n - \lambda)p_{n-1}(\lambda) - \beta_{n-1}^2 p_{n-2}(\lambda).$$

Thus we have a recurrence for the characteristic polynomials of T_k .

$$\begin{aligned} p_0(\lambda) &= 1 \\ p_1(\lambda) &= \alpha_1 - \lambda \\ p_{k+1}(\lambda) &= (\alpha_{k+1} - \lambda)p_k(\lambda) - \beta_k^2 p_{k-1}(\lambda). \end{aligned}$$

That produces a sequence of polynomials

$$\{p_0(\lambda), p_1(\lambda), \dots, p_n(\lambda)\}$$

that contain considerable information about the eigenvalues of T . This sequence is called a Sturm sequence.

Next time, we will show how to exploit that information.