STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2019
LECTURE 4

1. Gerschgorin's theorem

- $A = [a_{ij}] \in \mathbb{C}^{n \times n}$, for $i = 1, \dots, n$, we define the *Gerschgorin's discs*
$$G_i \coloneqq \{z \in \mathbb{C} : |z - a_{ii}| \le r_i\}$$
where
$$r_i \coloneqq \sum_{j \ne i} |a_{ij}|$$
- Gerschgorin's theoerm says that the n eigenvalues of A are all contained in the union of G_i's
- before we prove this, we need a result that is by itself useful
- a matrix $A \in \mathbb{C}^{n \times n}$ is called *strictly diagonally dominant* if
$$|a_{ii}| > \sum_{j \ne i} |a_{ij}|$$
- it is called *weakly diagonally dominant* if the '$>$' is replaced by '$\ge$'

Lemma 1. *A strictly diagonally dominant matrix is nonsingular.*

Proof. Let A be strictly diagonally dominant. Suppose $A\mathbf{x} = \mathbf{0}$ for some $\mathbf{x} \ne \mathbf{0}$. Let $k \in \{1, \dots, n\}$ be such that $|x_k| = \max_{i=1,\dots,n} |x_i|$. Since $\mathbf{x} \ne \mathbf{0}$, we must have $|x_k| > 0$. Now observe that
$$a_{kk} x_k = -\sum_{j \ne k} a_{kj} x_j$$
and so by the triangle inequality,
$$|a_{kk}||x_k| = \left| \sum\nolimits_{j \ne k} a_{kj} x_j \right| \le \sum\nolimits_{j \ne k} |a_{kj}||x_j| \le \left(\sum\nolimits_{j \ne k} |a_{kj}| \right) |x_k| < |a_{kk}||x_k|$$
where the last inequality is by strict diagonal dominance. This is a contradiction. In other words, there are no non-zero vector with $A\mathbf{x} = \mathbf{0}$. So $\ker(A) = \{\mathbf{0}\}$ and so A is nonsingular. $\square$

- we are going to use this to prove the first part of Gerschgorin theorem
- the second part requires a bit of topology

Theorem 1 (Gerschgorin)**.** *The spectrum of A is contained in the union of its Gerschgorin's discs, i.e.,*
$$\lambda(A) \subseteq \bigcup\nolimits_{i=1}^{n} G_i.$$
Furthermore, the number of eigenvalues (counted with multiplicity) in each connected component of $\bigcup_{i=1}^{n} G_i$ is equal to the number of Gerschgorin discs that constitute that component.

Date: October 9, 2019, version 1.0. Comments, bug reports: `lekheng@galton.uchicago.edu`.

Proof. Suppose $z \notin \bigcup_{i=1}^n G_i$. Then $A - zI$ is a strictly diagonal dominant matrix (check!) and therefore nonsingular by the above lemma. Hence $\det(A - zI) \neq 0$ and so z is not an eigenvalue of A. This proves the first part. For the second part, consider the matrix

$$A(t) := \begin{bmatrix} a_{11} & ta_{12} & \cdots & ta_{1n} \\ ta_{21} & a_{22} & \cdots & ta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ta_{n1} & ta_{n2} & \cdots & a_{nn} \end{bmatrix}$$

for $t \in [0,1]$. Note that $A(0) = \operatorname{diag}(a_{11}, \dots, a_{nn})$ and $A(1) = A$. We will let $G_i(t)$ be the ith Gerschgorin disc of $A(t)$. So

$$G_i(t) = \{z \in \mathbb{C} : |z - a_{ii}| \leq tr_i\}.$$

Clearly $G_i(t) \subseteq G_i$ for any $t \in [0,1]$. By the first part, all eigenvalues of all the matrices $A(t)$ are contained in $\bigcup_{i=1}^n G_i$. Since the set of eigenvalues of the matrices $A(t)$ depends continuously on the parameter t, $A(0)$ must have the same number of eigenvalues as $A(1)$ in each connected component of $\bigcup_{i=1}^n G_i$. Now just observe that the eigenvalues of $A(0)$ are simply the centers a_{kk} of each discs in a connected component. ☐

2. Schur decomposition

- suppose we want a decomposition for arbitrary matrices $A \in \mathbb{C}^{n \times n}$ like the EVD for normal matrices $A = Q\Lambda Q^*$, i.e., diagonalizing with unitary matrices
- the way to obtain such a decomposition is to relax the requirement of having a diagonal matrix Λ in $A = Q\Lambda Q^*$ but instead allow it to be upper-triangular
- this gives the *Schur decomposition*:

$$A = QRQ^* \tag{2.1}$$

- as in the EVD for normal matrices, Q is a unitary matrix but

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix}$$

- note that we the eigenvalues of A are precisely the diagonal entries of R: $r_{11}, \dots, r_{nn}$
- unlike the Jordan canonical form, the Schur decomposition is readily computable in finite-precision via the QR algorithm
- the QR algorithm is based on the QR decomposition, which we will discuss later in this course
- in its most basic form, QR algorithm does the following:

INPUT:	$A_0 = A$;	
STEP k:	$A_k = Q_k R_k$;	perform QR decomposition
STEP $k+1$:	$A_{k+1} = R_k Q_k$;	multiply QR factors in reverse order

- under suitable conditions, one may show that $Q_k \to Q$ and $R_k \to R$ where Q and R are as the requisite factors in (2.1)
- in most undergraduate linear algebra classes, one is taught to find eigenvalues by solving for the roots of the characteristic polynomial

$$p_A(x) = \det(xI - A) = 0$$

- this is almost never the case in practice
- for one, there is no finite algorithms for finding the roots of a polynomial when the degree exceeds four — by the famous impossibility result of Abel–Galois

- in fact what happens is the opposite — the roots of a univariate polynomial (divide by the coefficient of the highest degree term first so that it becomes a monic polynomial)
$$p(x) = c_0 + c_1 x + \cdots + c_{n-1}x^{n-1} + x^n$$
are usually obtained as the eigenvalues of its *companion matrix*
$$C_p = \begin{bmatrix} 0 & 0 & \dots & 0 & -c_0 \\ 1 & 0 & \dots & 0 & -c_1 \\ 0 & 1 & \dots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -c_{n-1} \end{bmatrix}$$
using the QR algorithm
- exercise: show that $\det(xI - C_p) = p(x)$

3. SINGULAR VALUE DECOMPOSITION

- the Schur decomposition exists for any square matrix but what if we want something analogous for rectangular matrices?
- let $A \in \mathbb{C}^{m\times n}$, we will see that we always have a decomposition
$$A = U\Sigma V^* \tag{3.1}$$
  - $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are both unitary matrices
$$U^*U = I_m = UU^*, \quad V^*V = I_n = VV^*$$
  - $\Sigma \in \mathbb{C}^{m\times n}$ is a diagonal matrix in the sense that $\sigma_{ij} = 0$ if $i \neq j$
  - if $m > n$, then Σ looks like
$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$
  - if $m < n$, then Σ looks like
$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 & \cdots & 0 \\ & \ddots & & \vdots & & \vdots \\ & & \sigma_m & 0 & \cdots & 0 \end{bmatrix}$$
  - if $m = n$, then Σ looks like
$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{bmatrix}$$
  - the diagonal elements of Σ, denoted σ_i, $i = 1, \dots, n$, are all nonnegative, and can be ordered such that
$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0, \quad \sigma_{r+1} = \cdots = \sigma_{\min(m,n)} = 0$$
  - r is the rank of A
- this decomposition of A is called the *singular value decomposition*, or SVD
  - the values σ_i, for $i = 1, 2, \dots, n$, are the *singular values* of A
  - the columns of U are the *left singular vectors*

- the columns of V are the *right singular vectors*

- an alternative decomposition of A omits the singular values that are equal to zero:

$$A = \tilde{U}\tilde{\Sigma}\tilde{V}^*$$

  - $\tilde{U} \in \mathbb{C}^{m\times r}$ is a matrix with orthonormal columns, i.e. satisfying $\tilde{U}^*\tilde{U} = I_r$ (but not $\tilde{U}\tilde{U}^* = I_m$!)
  - $\tilde{V} \in \mathbb{C}^{n\times r}$ is also a matrix with orthonormal columns, i.e. satisfying $\tilde{V}^*\tilde{V} = I_r$ (but again not $\tilde{V}\tilde{V}^* = I_n$!)
  - $\tilde{\Sigma}$ is an $r \times r$ diagonal matrix with diagonal elements $\sigma_1, \dots, \sigma_r$
  - again $r = \operatorname{rank}(A)$
  - the columns of $\tilde{U}$ are the left singular vectors corresponding to the nonzero singular values of A, and form an orthonormal basis for the image of A
  - the columns of $\tilde{V}$ are the right singular vectors corresponding to the nonzero singular values of A, and form an orthonormal basis for the coimage of A
  - this is called the *condensed* or *compact* or *reduced* SVD
  - note that in this case, $\tilde{\Sigma}$ is a square matrix
- the form in (3.1) is sometimes called the *full* SVD
- we may also write the reduced SVD as a sum of rank-1 matrices

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^* + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^* + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^*$$

  - $\tilde{U} = [\mathbf{u}_1, \dots, \mathbf{u}_r]$, i.e. $\mathbf{u}_1, \dots, \mathbf{u}_r \in \mathbb{C}^m$ are the left singular vectors of A
  - $\tilde{V} = [\mathbf{v}_1, \dots, \mathbf{v}_r]$, i.e. $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbb{C}^n$ are the right singular vectors of A
- it remains to show that the SVD always exist for any matrix

Theorem 2 (Existence of SVD)**.** *Every matrix has a singular value decomposition (condensed version).*

Proof. Let $A \in \mathbb{C}^{m\times n}$ and for simplicity we assume that all its nonzero singular values are distinct. We define the matrix

$$W = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \in \mathbb{C}^{(m+n)\times(m+n)}.$$

It is easy to verify that $W = W^*$ (after Wielandt, who's the first to consider this matrix) and by the spectral theorem for Hermitian matrices, W has an EVD,

$$W = Z\Lambda Z^*$$

where $Z \in \mathbb{C}^{(m+n)\times(m+n)}$ is a unitary matrix and $\Lambda \in \mathbb{R}^{(m+n)\times(m+n)}$ is a diagonal matrix with real diagonal elements. If $\mathbf{z}$ is an eigenvector of W, then we can write

$$W\mathbf{z} = \sigma\mathbf{z}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$$

and therefore

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \sigma \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$$

or, equivalently,

$$A\mathbf{y} = \sigma\mathbf{x}, \quad A^*\mathbf{x} = \sigma\mathbf{y}.$$

Now, suppose that we apply W to the vector obtained from $\mathbf{z}$ by negating $\mathbf{y}$. Then we have

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ -\mathbf{y} \end{bmatrix} = \begin{bmatrix} -A\mathbf{y} \\ A^*\mathbf{x} \end{bmatrix} = \begin{bmatrix} -\sigma\mathbf{x} \\ \sigma\mathbf{y} \end{bmatrix} = -\sigma \begin{bmatrix} \mathbf{x} \\ -\mathbf{y} \end{bmatrix}.$$

In other words, if $\sigma \neq 0$ is an eigenvalue, then $-\sigma$ is also an eigenvalue. So we may assume without loss of generality that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 = 0 \cdots = 0$$

where $r = \operatorname{rank}(A)$. So the diagonal matrix Λ of eigenvalues of W may be written as

$$\Lambda = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, -\sigma_1, -\sigma_2, \ldots, -\sigma_r, 0, \ldots, 0) \in \mathbb{C}^{(m+n)\times(m+n)}.$$

Observe that there is a zero block of size $(m+n-2r) \times (m+n-2r)$ in the bottom right corner of Λ.

We scale the eigenvector $\mathbf{z}$ of W so that $\mathbf{z}^*\mathbf{z} = 2$. Since W is symmetric, eigenvectors corresponding to the distinct eigenvalues σ and $-\sigma$ are orthogonal, so it follows that

$$\begin{bmatrix} \mathbf{x}^* & \mathbf{y}^* \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ -\mathbf{y} \end{bmatrix} = 0.$$

These yield the system of equations

$$\begin{aligned} \mathbf{x}^*\mathbf{x} + \mathbf{y}^*\mathbf{y} &= 2, \\ \mathbf{x}^*\mathbf{x} - \mathbf{y}^*\mathbf{y} &= 0, \end{aligned}$$

which has the unique solution

$$\mathbf{x}^*\mathbf{x} = 1, \quad \mathbf{y}^*\mathbf{y} = 1.$$

Now note that we can represent the matrix of normalized eigenvectors of W corresponding to nonzero eigenvalues (note that there are exactly $2r$ of these) as

$$\tilde{Z} = \frac{1}{\sqrt{2}} \begin{bmatrix} X & X \\ Y & -Y \end{bmatrix} \in \mathbb{C}^{(m+n)\times 2r}.$$

Note that the factor $1/\sqrt{2}$ appears because of the way we have chosen the norm of $\mathbf{z}$. We also let

$$\tilde{\Lambda} = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, -\sigma_1, -\sigma_2, \ldots, -\sigma_r) \in \mathbb{C}^{2r\times 2r}.$$

It is easy to see that

$$Z\Lambda Z^* = \tilde{Z}\tilde{\Lambda}\tilde{Z}^*$$

just by multiplying out the zero block in Λ. So we have

$$\begin{aligned} \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} &= W = Z\Lambda Z^* = \tilde{Z}\Lambda\tilde{Z}^* \\ &= \frac{1}{2} \begin{bmatrix} X & X \\ Y & -Y \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & -\Sigma_r \end{bmatrix} \begin{bmatrix} X^* & Y^* \\ X^* & -Y^* \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} X\Sigma_r & -X\Sigma_r \\ Y\Sigma_r & Y\Sigma_r \end{bmatrix} \begin{bmatrix} X^* & Y^* \\ X^* & -Y^* \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 0 & 2X\Sigma_r Y^* \\ 2Y\Sigma_r X^* & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & X\Sigma_r Y^* \\ Y\Sigma_r X^* & 0 \end{bmatrix} \end{aligned}$$

and therefore

$$A = X\Sigma_r Y^*, \quad A^* = Y\Sigma_r X^*$$

where X is an $m \times r$ matrix, Σ is $r \times r$, and Y is $n \times r$, and r is the rank of A. We have obtained the condensed SVD of A.

The last missing bit is the orthonormality of the columns of X and Y. This follows from the fact that distinct columns of

$$\begin{bmatrix} X & X \\ Y & -Y \end{bmatrix}$$

are mutually orthogonal since they correspond to distinct eigenvalues and so if we pick $\left[\begin{smallmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} \mathbf{x}_j \\ \mathbf{y}_j \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} \mathbf{x}_j \\ -\mathbf{y}_j \end{smallmatrix}\right]$ for $i \neq j$, and take their inner products, we get

$$\mathbf{x}_i^* \mathbf{x}_j + \mathbf{y}_i^* \mathbf{y}_j = 0,$$
$$\mathbf{x}_i^* \mathbf{x}_j - \mathbf{y}_i^* \mathbf{y}_j = 0.$$

Adding them gives $\mathbf{x}_i^* \mathbf{x}_j = 0$ and substracting them gives $\mathbf{y}_i^* \mathbf{y}_j = 0$ for all $i \neq j$, as required. □

- see Theorem 4.1 in Trefethen and Bau for an alternative non-constructive proof that does not require the use of spectral theorem

4. OTHER CHARACTERIZATIONS OF SVD

- the proof of the above theorem gives us two more characterizations of singular values and singular vectors:
  (i) in terms of eigenvalues and eigenvectors of an $(m+n) \times (m+n)$ Hermitian matrix:
$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} U & U \\ V & -V \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} \begin{bmatrix} U^* & V^* \\ U^* & -V^* \end{bmatrix}$$
  (ii) in terms of a coupled system of equations
$$\begin{cases} A\mathbf{v} = \sigma \mathbf{u}, \\ A^*\mathbf{u} = \sigma \mathbf{v} \end{cases}$$
- the following is yet another way to characterize them in terms of eigenvalues/eigenvectors of an $m \times m$ Hermitian matrix and an $n \times n$ Hermitian matrix

Lemma 2. *The suqare of the singular values of a matrix A are eigenvalues of AA^* and A^*A. The left singular vectors of A are the eigenvectors of AA^* and the right singular vectors of A are the eigenvectors of A^*A.*

Proof. From the relationships $A\mathbf{y} = \sigma\mathbf{x}$, $A^*\mathbf{x} = \sigma\mathbf{y}$, we obtain

$$A^*A\mathbf{y} = \sigma^2 \mathbf{y}, \quad AA^*\mathbf{x} = \sigma^2 \mathbf{x}.$$

Therefore, if $\pm\sigma$ are eigenvalues of W, then σ^2 is an eigenvalue of both AA^* and A^*A. Also

$$AA^* = (U\Sigma V^*)(V\Sigma^\mathsf{T} U^*) = U\Sigma\Sigma^\mathsf{T} U^*,$$
$$A^*A = (V\Sigma^\mathsf{T} U^*)(U\Sigma V^*) = V\Sigma^\mathsf{T}\Sigma V^*.$$

Note that $\Sigma^* = \Sigma^\mathsf{T}$ since singular values are real. The matrices $\Sigma^\mathsf{T}\Sigma$ and $\Sigma\Sigma^\mathsf{T}$ are respectively $n \times n$ and $m \times m$ diagonal matrices with diagonal elements σ_i^2 and 0. □

- the SVD is something like a swiss army knife of linear algebra, matrix theory, and numerical linear algebra, you can do a lot with it
- over the next few sections we will see that the singular value decomposition is a singularly powerful tool — once we have it, we could solve just about any problem involving matrices
  - given $A \in \mathbb{C}^{m \times n}$ and $\mathbf{b} \in \operatorname{im}(A) \subseteq \mathbb{C}^m$, find all solutions of $A\mathbf{x} = \mathbf{b}$
  - given $A \in \mathrm{GL}(n)$, find A^{-1}
  - given $A \in \mathbb{C}^{m \times n}$, find $\|A\|_2$ and $\|A\|_F$
  - given $A \in \mathbb{C}^{m \times n}$, find $\|A\|_{\sigma,p,k}$ for $p \in [0, \infty]$ and $k \in \mathbb{N}$

  - given $A \in \mathbb{C}^{n \times n}$, find $|\det(A)|$
  - given $A \in \mathbb{C}^{m \times n}$, find $A^{\dagger}$
  - given $A \in \mathbb{C}^{m \times n}$ and $\mathbf{b} \in \mathbb{C}^{m}$, find all solutions of the $\min_{\mathbf{x} \in \mathbb{C}^n} \|A\mathbf{x} - \mathbf{b}\|_2$
- the good news is that unlike the Jordan canonical form, the SVD is actually computable
- there are two main methods to compute it: Golub-Reinsch and Golub-Kahan, we will look at these briefly later, right now all you need to know is that you can call MATLAB to give you the SVD, both the full and compact versions
- in all of the following we shall assume that we have the full SVD of $A = U\Sigma V^*$
- furthermore $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values of A and $r = \operatorname{rank}(A)$