STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2019
LECTURE 2

1. Operator norms

- a very important class of matrix norms are the so called *operator* or *induced* or *natural norms* defined as
$$\|A\|_{\alpha,\beta} \coloneqq \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_\alpha}{\|\mathbf{x}\|_\beta} \tag{1.1}$$
for any $A \in \mathbb{C}^{m\times n}$ and any vector norms $\|\cdot\|_\alpha : \mathbb{C}^m \to \mathbb{R}$ and $\|\cdot\|_\beta : \mathbb{C}^n \to \mathbb{R}$ defined on the domain and codomain of A respectively
- the operator norm may also be written as
$$\|A\|_{\alpha,\beta} = \max\{\|A\mathbf{x}\|_\alpha : \|\mathbf{x}\|_\beta \le 1\} \tag{1.2}$$
or as
$$\|A\|_{\alpha,\beta} = \max\{\|A\mathbf{x}\|_\alpha : \|\mathbf{x}\|_\beta = 1\} \tag{1.3}$$
- in other words, the operator norm measures how far the operator A sends points in the unit disc (or the unit circle)
- proof is simple, for example, here's how you would prove (1.3):
$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_\alpha}{\|\mathbf{x}\|_\beta} = \max_{\mathbf{x}\neq\mathbf{0}} \left\| \frac{1}{\|\mathbf{x}\|_\beta} A\mathbf{x} \right\|_\alpha = \max_{\mathbf{x}\neq\mathbf{0}} \left\| A\left(\frac{\mathbf{x}}{\|\mathbf{x}\|_\beta}\right) \right\|_\alpha = \max_{\|\mathbf{v}\|_\beta=1} \|A\mathbf{v}\|_\alpha,$$
the first equality uses the property that $|\alpha|\|\mathbf{v}\|_\alpha = \|\alpha\mathbf{v}\|_\alpha$, the second equality uses $\alpha A\mathbf{x} = A(\alpha\mathbf{x})$, and the last equality uses the observation that $\mathbf{v} = \mathbf{x}/\|\mathbf{x}\|_\beta$ always has unit β-norm
- exercise: prove (1.3) and (1.2) are equal
- another exercise: prove that
$$\|A\mathbf{x}\|_\alpha \le \|A\|_{\alpha,\beta}\|\mathbf{x}\|_\beta \tag{1.4}$$
for any $\mathbf{x} \in \mathbb{C}^n$; this more restrictive form of submultiplicativity is called *consistency*
- a note on the use of *supremum* and *maximum*: for $S \subseteq \mathbb{C}^n$ and a real-valued function f whose domain includes S,
  - we write $\sup_{\mathbf{x}\in S} f(\mathbf{x})$ for the smallest $\mu \in \mathbb{R}$ such that $f(\mathbf{x}) \le \mu$ for every $\mathbf{x} \in S$ (and we set $\mu = +\infty$ if f is unbounded on S)
  - we write $\max_{\mathbf{x}\in S} f(\mathbf{x})$ if the supremum is attained by some element in S, i.e., there is an $\mathbf{x}_{\max} \in S$ such that $f(\mathbf{x}_{\max}) = \sup_{\mathbf{x}\in S} f(\mathbf{x})$
  - $\mathbf{x}_{\max}$ is called a *maximizer* of f on S
  - likewise for infimum and minimum (and minimizer)
  - by the extreme value theorem, if f is continuous and S is compact, then supremum and infimum are always attained
- in the above $S = \{\mathbf{x} \in \mathbb{C} : \|\mathbf{x}\|_\beta \le 1\}$ and $S = \{\mathbf{x} \in \mathbb{C} : \|\mathbf{x}\|_\beta = 1\}$ are compact and the function $f = \|\cdot\|_\alpha : \mathbb{C}^m \to \mathbb{R}$ is continuous
- in other words, we can always find an $\mathbf{x}_{\max}$ with $\|\mathbf{x}_{\max}\|_\beta = 1$ such that
$$\|A\mathbf{x}_{\max}\|_\alpha = \|A\|_{\alpha,\beta}$$

Date: October 7, 2019, version 1.1. Comments, bug reports: `lekheng@galton.uchicago.edu`.

- that's why we may always write max in (1.3) and (1.2), and therefore in (1.1); although strictly speaking we should have written (1.1)
$$\|A\|_{\alpha,\beta} \coloneqq \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_\alpha}{\|\mathbf{x}\|_\beta}$$
- the operator norm is *not* submultiplicative in general: take
$$A = \begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$
since every $\mathbf{x} \in \mathbb{R}^2$ with $\|\mathbf{x}\| = 1$ has the form $\mathbf{x} = (\cos\theta, \sin\theta)^\mathsf{T}$, we see that
$$\begin{aligned} \|A\|_{\infty,2} &= \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_\infty = \max_\theta |2\cos\theta + 2\sin\theta| = 2\sqrt{2} \\ \|B\|_{\infty,2} &= \max_{\|\mathbf{x}\|_2=1} \|B\mathbf{x}\|_\infty = \max_\theta |\cos\theta| = 1 \\ \|AB\|_{\infty,2} &= \max_{\|\mathbf{x}\|_2=1} \|AB\mathbf{x}\|_\infty = \max_\theta |4\cos\theta| = 4 \end{aligned}$$
but
$$\|AB\|_{\infty,2} = 4 > 2\sqrt{2} = \|A\|_{\infty,2}\|B\|_{\infty,2}$$
(thanks to Lijun Ding for this example)
- however given $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times p}$ it is always true that
$$\|AB\|_{\alpha,\gamma} \le \|A\|_{\alpha,\beta}\|B\|_{\beta,\gamma}$$
for any norms $\|\cdot\|_\gamma$ on $\mathbb{C}^p$, $\|\cdot\|_\beta$ on $\mathbb{C}^n$, $\|\cdot\|_\alpha$ on $\mathbb{C}^m$
- the most interesting operator norms are the ones obtained when $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are vector ℓ^p-norms, we write
$$\|A\|_{p,q} \coloneqq \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_q} \quad \text{and} \quad \|A\|_p \coloneqq \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$
for any $A \in \mathbb{C}^{m\times n}$ and $p, q \in [1, \infty]$
- we call $\|\cdot\|_{p,q}$ the matrix (p,q)-norm and $\|\cdot\|_p$ the matrix p-norm
- the matrix 2-norm
$$\|A\|_2 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$$
is very widely used and has its own special name, *spectral norm*, because of its relation to the spectrum of a matrix (i.e., the eigenvalues); we will discuss it in the next two lectures
- the matrix 1-norm and ∞-norm are also very widely used, largely because, they can be easily computed
- let $A = [a_{ij}]_{i,j=1}^{m,n} \in \mathbb{C}^{m\times n}$, then
$$\|A\|_1 = \max_{j=1,\dots,n} \left[\sum_{i=1}^m |a_{ij}|\right] \tag{1.5}$$
and
$$\|A\|_\infty = \max_{i=1,\dots,m} \left[\sum_{j=1}^n |a_{ij}|\right] \tag{1.6}$$
- an easy way to remember these is that $\|A\|_1$ is the maximum column sum and $\|A\|_\infty$ is the maximum row sum of A
- let us prove (1.6) and leave (1.5) as an exercise:

– we use (1.3), so

$$
\begin{aligned}
\|A\|_\infty &= \max\{\|A\mathbf{x}\|_\infty : \|\mathbf{x}\|_\infty = 1\} \\
&= \max_{\|\mathbf{x}\|_\infty=1} \left\{ \max_{i=1,\dots,m} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \right\} \\
&\le \max_{\|\mathbf{x}\|_\infty=1} \left\{ \max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}||x_j| \right] \right\} \\
&\le \max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}| \right] \qquad (1.7)
\end{aligned}
$$

where the last inequality follows because $\|\mathbf{x}\|_\infty = 1$ and so we must have $|x_j| \le 1$

– to show equality, we just need to exhibit one single $\mathbf{x}^*$ with $\|\mathbf{x}^*\|_\infty = 1$ so that

$$\|A\mathbf{x}^*\|_\infty \ge \max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}| \right]$$

– we know that the maximum in (1.7) is attained by some row $i = k \in \{1, \dots, m\}$, so

$$\max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}| \right] = \sum_{j=1}^{n} |a_{kj}|$$

– now we define $\mathbf{x}^* = [x_1^*, \dots, x_n^*] \in \mathbb{C}^n$ as the vector whose coordinates are given by

$$x_j^* = \begin{cases} |a_{kj}|/a_{kj} & \text{if } a_{kj} \ne 0, \\ 0 & \text{if } a_{kj} = 0, \end{cases}$$

for $j = 1, \dots, n$

– observe that $\mathbf{x}^*$ has $\|\mathbf{x}^*\|_\infty = 1$ as well as the effect of attaining the requisite bound

$$\|A\mathbf{x}^*\|_\infty = \max_{i=1,\dots,m} \left| \sum_{j=1}^{n} a_{ij} x_j^* \right| \ge \sum_{j=1}^{n} |a_{kj}| = \max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}| \right]$$

2. INNER PRODUCTS

- an *inner product* is a complex-valued function on a product of a vector space (over $\mathbb{C}$) with itself, denoted $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$, satisfying
  (1) $\langle v, v \rangle \ge 0$ for all $v \in V$
  (2) $\langle v, v \rangle = 0$ if and only if $v = 0_V$
  (3) $\langle v, \alpha_1 w_1 + \alpha_2 w_2, w \rangle = \alpha_1 \langle v, w_1 \rangle + \alpha_2 \langle v, w_2 \rangle$ for all $\alpha_1, \alpha_2 \in \mathbb{C}$ and $v, w_1, w_2 \in V$
  (4) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for any $v, w \in V$
- by virtue of the last two conditions $\langle \alpha_1 v_1 + \alpha_2 v_2, w \rangle = \overline{\alpha}_1 \langle v_1, w \rangle + \overline{\alpha}_2 \langle v_2, w \rangle$ for all $\alpha_1, \alpha_2 \in \mathbb{C}$ and $v_1, v_2, w \in V$
- for real vector spaces, an inner product is a real-valued function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ and as a result, the last two conditions become:
  (3) $\langle \alpha_1 v_1 + \alpha_2 v_2, w \rangle = \alpha_1 \langle v_1, w \rangle + \alpha_2 \langle v_2, w \rangle$ for all $\alpha_1, \alpha_2 \in \mathbb{R}$ and $v_1, v_2, w \in V$
  (4) $\langle v, w \rangle = \langle w, v \rangle$ for any $v, w \in V$
- just as norms are an abstraction of length, inner products are an abstraction of angles (or rather, inverse cosines of angles)

- the defining properties of an inner product tell us that
$$\|v\| \coloneqq \sqrt{\langle v, v\rangle}$$
defines a norm called the *norm induced by the inner product*
- Cauchy–Schwartz inequality in fact holds for any inner product and the norm induced by that inner product
$$|\langle v, w\rangle| \le \|v\|\|w\|$$
- given a norm $\|\cdot\|$, how can we tell if it is a norm induced by some inner product?
- the answer is: if and only if the norm satisfies the *parallelogram law*
$$\|v+w\|^2 + \|v-w\|^2 = 2\|v\|^2 + 2\|w\|^2$$
- the only inner products we care about in this course are the *Hermitian inner product* or l^2-inner product for vectors
$$\langle \mathbf{x}, \mathbf{y}\rangle \coloneqq \mathbf{x}^*\mathbf{y} = \sum_{i=1}^n \overline{x}_i y_i, \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n$$
and *trace inner product* for matrices
$$\langle X, Y\rangle \coloneqq \operatorname{tr}(X^*Y) = \sum_{i=1}^m \sum_{j=1}^n \overline{x}_{ij} y_{ij}, \qquad \text{for all } X, Y \in \mathbb{C}^{m\times n}$$
- over reals, we have
$$\langle \mathbf{x}, \mathbf{y}\rangle \coloneqq \mathbf{x}^{\mathsf{T}}\mathbf{y} = \sum_{i=1}^n \overline{x}_i y_i, \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$
called the *Euclidean* (instead of Hermitian) inner product, and
$$\langle X, Y\rangle \coloneqq \operatorname{tr}(X^{\mathsf{T}}Y) = \sum_{i=1}^m \sum_{j=1}^n \overline{x}_{ij} y_{ij}, \qquad \text{for all } X, Y \in \mathbb{R}^{m\times n}$$
- the norms induced by these inner products are precisely the Euclidean norm and Frobenius norm respectively since
$$(\mathbf{x}^*\mathbf{x})^2 = \|\mathbf{x}\|_2^2 \quad \text{and} \quad \operatorname{tr}(X^*X) = \|X\|_F^2$$
for any $\mathbf{x} \in \mathbb{C}^n$ and $X \in \mathbb{C}^{m\times n}$
- Cauchy–Schwartz inequality yields
$$|\mathbf{x}^*\mathbf{y}| \le \|\mathbf{x}\|_2\|\mathbf{y}\|_2 \quad \text{and} \quad |\operatorname{tr}(X^*Y)| \le \|X\|_F\|Y\|_F$$
- using the parallelogram law, we can show that no other vector p-norms or matrix Hölder p-norm are induced by inner products when $p \ne 2$
- the parallelogram law also tells us that matrix (p, q)-norm are not induced by inner products, whatever the value of p and q (including $p = q = 2$, so the spectral norm is not induced by an inner product either)

3. Outer product

- for $\mathbf{x} = [x_1, \dots, x_m]^{\mathsf{T}} \in \mathbb{C}^m$ and $\mathbf{y} = [y_1, \dots, y_n]^{\mathsf{T}} \in \mathbb{C}^n$, the product
$$\mathbf{x}\mathbf{y}^{\mathsf{T}} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix} \in \mathbb{C}^{m\times n}$$

or

$$\mathbf{x}\mathbf{y}^* = \begin{bmatrix} x_1\bar{y}_1 & x_1\bar{y}_2 & \cdots & x_1\bar{y}_n \\ x_2\bar{y}_1 & x_2\bar{y}_2 & \cdots & x_2\bar{y}_n \\ \vdots & \vdots & & \vdots \\ x_m\bar{y}_1 & x_m\bar{y}_2 & \cdots & x_m\bar{y}_n \end{bmatrix} \in \mathbb{C}^{m\times n}$$

is often called the *outer product* of $\mathbf{x}$ and $\mathbf{y}$

- if neither $\mathbf{x}$ nor $\mathbf{y}$ is the zero vector, then

$$\operatorname{rank}(\mathbf{x}\mathbf{y}^\mathsf{T}) = \operatorname{rank}(\mathbf{x}\mathbf{y}^*) = 1$$

- furthermore if $\operatorname{rank}(A) = 1$, then there exists $\mathbf{x} \in \mathbb{C}^m$ and $\mathbf{y} \in \mathbb{C}^n$ so that $A = \mathbf{x}\mathbf{y}^*$
- as such a matrix of this form is often called a *rank-1 matrix*

4. MATRIX PRODUCT

- the following are some useful observations regarding matrix-matrix and matrix-vector products
- by definition, multiplying two matrices

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{\alpha}_n^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{m\times n} \qquad \text{and} \qquad B = [\mathbf{b}_1, \dots, \mathbf{b}_p] \in \mathbb{R}^{n\times p}$$

is the same as forming the matrix of inner products of the row vectors $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m \in \mathbb{R}^n$ and the column vectors $\mathbf{b}_1, \dots, \mathbf{b}_p \in \mathbb{R}^n$,

$$AB = \begin{bmatrix} \boldsymbol{\alpha}_1^\mathsf{T}\mathbf{b}_1 & \boldsymbol{\alpha}_1^\mathsf{T}\mathbf{b}_2 & \cdots & \boldsymbol{\alpha}_1^\mathsf{T}\mathbf{b}_p \\ \boldsymbol{\alpha}_2^\mathsf{T}\mathbf{b}_1 & \boldsymbol{\alpha}_2^\mathsf{T}\mathbf{b}_2 & \cdots & \boldsymbol{\alpha}_2^\mathsf{T}\mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}_m^\mathsf{T}\mathbf{b}_1 & \boldsymbol{\alpha}_m^\mathsf{T}\mathbf{b}_2 & \cdots & \boldsymbol{\alpha}_m^\mathsf{T}\mathbf{b}_p \end{bmatrix} \in \mathbb{R}^{m\times p} \tag{4.1}$$

- multiplying two matrices

$$A = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbb{R}^{m\times n} \qquad \text{and} \qquad B = \begin{bmatrix} \boldsymbol{\beta}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{\beta}_n^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{n\times p}$$

is the same as taking the sum of outer products of the column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$ and the row vectors $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_n \in \mathbb{R}^n$,

$$AB = \mathbf{a}_1\boldsymbol{\beta}_1^\mathsf{T} + \cdots + \mathbf{a}_n\boldsymbol{\beta}_n^\mathsf{T} \in \mathbb{R}^{m\times p} \tag{4.2}$$

- multiplying two matrices

$$A \in \mathbb{R}^{m\times n} \qquad \text{and} \qquad B = [\mathbf{b}_1, \dots, \mathbf{b}_p] \in \mathbb{R}^{n\times p}$$

is the same as multiplying A to each of the column vectors $\mathbf{b}_1, \dots, \mathbf{b}_p \in \mathbb{R}^n$ of B,

$$AB = [A\mathbf{b}_1, \dots, A\mathbf{b}_p] \in \mathbb{R}^{m\times p}$$

- multiplying two matrices

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{\alpha}_n^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{m\times n} \qquad \text{and} \qquad B \in \mathbb{R}^{n\times p}$$

is the same as multiplying each of the row vectors $\boldsymbol{\alpha}_1^\mathsf{T}, \dots, \boldsymbol{\alpha}_n^\mathsf{T} \in (\mathbb{R}^m)^* = \mathbb{R}^{m\times 1}$ by B on the right:

$$AB = \begin{bmatrix} \boldsymbol{\alpha}_1^\mathsf{T} B \\ \vdots \\ \boldsymbol{\alpha}_n^\mathsf{T} B \end{bmatrix} \in \mathbb{R}^{m\times p}$$

- the following are special cases when one of the matrices is a vector or is a diagonal matrix
- multiplying $A = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$ on the right by $\mathbf{x} = [x_1, \dots, x_n]^\mathsf{T} \in \mathbb{R}^n$ is the same as taking linear combinations of the column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$,

$$A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \in \mathbb{R}^m$$

- multiplying

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{\alpha}_n^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

  on the left by $\mathbf{y}^\mathsf{T} = [y_1, \dots, y_m] \in \mathbb{R}^m$ is the same as taking linear combinations of the row vectors $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m \in \mathbb{R}^n$,

$$\mathbf{y}^\mathsf{T} A = y_1\boldsymbol{\alpha}_1^\mathsf{T} + \cdots + y_m\boldsymbol{\alpha}_m^\mathsf{T} \in \mathbb{R}^{n*}$$

  where $\mathbb{R}^{n*} = \mathbb{R}^{1\times n}$ is the dual space of $\mathbb{R}^n = \mathbb{R}^{n\times 1}$
- multiplying $A = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$ on the right by a diagonal matrix

$$D = \operatorname{diag}(d_1, \dots, d_n) = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \in \mathbb{R}^{n\times n}$$

  is the same as scaling the column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^n$ by $d_1, \dots, d_n \in \mathbb{R}$,

$$AD = [d_1\mathbf{a}_1, \dots, d_n\mathbf{a}_n] \in \mathbb{R}^{m\times n}$$

- multiplying

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{\alpha}_m^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

  on the left by a diagonal matrix

$$D = \operatorname{diag}(d_1, \dots, d_m) = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_m \end{bmatrix} \in \mathbb{R}^{m\times m}$$

  is the same as scaling the row vectors $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n \in \mathbb{R}^m$, by $d_1, \dots, d_m \in \mathbb{R}$,

$$DA = \begin{bmatrix} d_1\boldsymbol{\alpha}_1^\mathsf{T} \\ \vdots \\ d_m\boldsymbol{\alpha}_m^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

5. Eigenvalues and eigenvectors

- recall: $A \in \mathbb{C}^{n\times n}$, if there exists $\lambda \in \mathbb{C}$ and $\mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n$ such that

$$A\mathbf{x} = \lambda\mathbf{x},$$

  we call λ an *eigenvalue* of A and $\mathbf{x}$ an *eigenvector* of A corresponding to λ or λ-eigenvector
- we review some basic properties and terminologies for eigenvalues and eigenvectors

- real matrices can have complex eigenvalues and eigenvectors, an example is

$$S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

which has eigenvectors $[-i, 1]^\mathsf{T}$ and $[i, 1]^\mathsf{T}$ corresponding to eigenvalues i and $-i$ respectively
- eigenvector is a scale invariant notion, if $\mathbf{x}$ is a λ-eigenvector, then so is $c\mathbf{x}$ for any $c \in \mathbb{C}^\times$
- we usually, but not always, require that $\mathbf{x}$ be a unit vector, i.e., $\|\mathbf{x}\|_2 = 1$
- note that if $\mathbf{x}_1$ and $\mathbf{x}_2$ are λ-eigenvectors, then so is $\mathbf{x}_1 + \mathbf{x}_2$
- for an eigenvalue λ, the subspace

$$V_\lambda \coloneqq \{\mathbf{x} \in \mathbb{C}^n : A\mathbf{x} = \lambda\mathbf{x}\}$$

is called the *λ-eigenspace* of A and is the set of all λ-eigenvectors of A together with $\mathbf{0}$
- the set of all eigenvalues of A is called its *spectrum* and often denoted $\lambda(A)$, i.e.,

$$\lambda(A) \coloneqq \{\lambda \in \mathbb{C} : A\mathbf{x} = \lambda\mathbf{x} \text{ for some } \mathbf{x} \neq \mathbf{0}\}$$

- an $n \times n$ matrix always have n eigenvalues in $\mathbb{C}$ counted with multiplicty
- however an $n \times n$ matrix may not have n linear independent eigenvectors, an example is

$$J = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{5.1}$$

which has eigenvalue 0 with multiplicity 2 but only one eigenvector (up to scaling) $\mathbf{x} = [1, 0]^\mathsf{T}$
- normally we will sort the eigenvalues in descending order of magnitude

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$$

- λ_1, also denoted $\lambda_{\max}$, is called the *principle eigenvalue* of A and a $\lambda_{\max}$-eigenvector is called a *principal eigenvector*
- the eigenvalues of A and A^T are identical, i.e., $\lambda(A) = \lambda(A^\mathsf{T})$
- the eigenvectors of A^T are called *left eigenvectors* of A (and if one needs to make a distinction, the usual eigenvectors are called *right eigenvectors*) and sometimes defined directly via

$$\mathbf{y}^\mathsf{T} A = \lambda \mathbf{y}^\mathsf{T}$$

- in general, for a nonsymmetric matrix $A \in \mathbb{R}^{n \times n}$, left and right eigenvectors corresponding to the same eigenvalue λ are different