MATH36001 Norms 2015

If one has several vectors in $\mathbb{C}^n$ or several matrices in $\mathbb{C}^{n\times n}$, what might it mean to say that some are small or that others are large? Under what circumstances might we say that two vectors are close together or far apart?

One way to answer these questions is to study norms of vectors and matrices. The study of norms is necessary for a proper formulation of notions such as power series of matrices. It is essential in the analysis and assessment of algorithms for numerical computations. Bounds for eigenvalues often involve norms, as do bounds for possible changes in the eigenvalues when a matrix is perturbed.

1 Vector Norms

A vector norm is a function $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ satisfying

1. $\|x\| \geq 0$ with equality iff $x = 0$, (*positive definiteness*)
2. $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{C}$, $x \in \mathbb{C}^n$, (*absolute homogeneity*)
3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{C}^n$. (*triangle inequality*)

The three most practically useful norms are

$$\begin{aligned}
\|x\|_1 &= |x_1| + \cdots + |x_n| = \sum_{i=1}^{n} |x_i|, \qquad \text{“Manhattan” or “taxi cab” norm,}\\
\|x\|_2 &= (|x_1|^2 + \cdots + |x_n|^2)^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} = (x^*x)^{1/2}, \qquad \text{Euclidean length,}\\
\|x\|_\infty &= \max_{1\leq i\leq n} |x_i|, \qquad \text{maximum norm.}
\end{aligned}$$

These are all special cases of the p-norm:

$$\|x\|_p = \big(|x_1|^p + \cdots + |x_n|^p\big)^{1/p} = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \qquad p \geq 1.$$

An important inequality is the **Hölder inequality**

$$|x^*y| \leq \|x\|_p\|y\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1.$$

The special case with $p = q = 2$ is called the **Cauchy–Schwarz inequality**:

$$|x^*y| \leq \|x\|_2\|y\|_2.$$

Let $V = [\,x \quad y\,] \in \mathbb{C}^{n\times 2}$ then V^*V is Hermitian positive semidefinite since $u^*(V^*V)u = \|Vu\|_2^2 \geq 0$ for all $u \in \mathbb{C}^n$. Also, $\det(V^*V) \geq 0$ since $\det(V^*V) =$ product of eigenvalues and eigenvalues of Hermitian positive semidefinite matrices are real and nonnegative. Hence $0 \leq \det(V^*V) = \det \begin{bmatrix} x^*x & x^*y \\ y^*x & y^*y \end{bmatrix} = \|x\|_2^2\|y\|_2^2 - |x^*y|^2$ and taking square roots gives the result.

All norms on $\mathbb{C}^n$ are **equivalent**: if $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are norms on $\mathbb{C}^n$ then there exist positive constants ν_1 and ν_2 such that

$$\nu_1\|x\|_\alpha \le \|x\|_\beta \le \nu_2\|x\|_\alpha, \quad \forall x \in \mathbb{C}^n.$$

For example, for all $x \in \mathbb{C}^n$ (see Exercise 3)

$$\|x\|_\infty \le \|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2 \le n\|x\|_\infty.$$

Vector norms can be used to measure the convergence of a sequence of vectors. We say that the sequence $\{v^{(k)}\}$ of vectors in $\mathbb{C}^n$ **converges** to a vector $v \in \mathbb{C}^n$ with respect to the vector norm $\|\cdot\|$ if and only if $\lim_{k\to\infty}\|v^{(k)} - v\| = 0$. The choice of norms is irrelevant since all norms on $\mathbb{C}^n$ are equivalent.

2 Matrix Norms

A matrix norm is a function $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{C}$ satisfying obvious analogues of the three vector norm properties (1)–(3). The simplest example is the Frobenius norm

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2\right)^{1/2} = \big(\mathrm{trace}(A^*A)\big)^{1/2}.$$

A very important class of matrix norms is those **subordinate** to vector norms (we also say **induced** by a vector norm). Given a vector norm, the corresponding subordinate matrix norm is defined by

$$\|A\| = \max_{x\neq 0}\frac{\|Ax\|}{\|x\|}.$$

When the vector norm is the p-norm then the subordinate matrix norm $\|A\|_p = \max_{x\neq 0}\|Ax\|_p/\|x\|_p$ is called the **matrix p-norm**.

Theorem 1 *Let $A \in \mathbb{C}^{m\times n}$, then*

$$\|A\|_1 = \max_{x\neq 0}\frac{\|Ax\|_1}{\|x\|_1} = \max_{1\le j\le n}\sum_{i=1}^m |a_{ij}|, \qquad \text{“max column sum”},$$

$$\|A\|_\infty = \max_{x\neq 0}\frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{1\le i\le m}\sum_{j=1}^n |a_{ij}| = \|A^T\|_1, \qquad \text{“max row sum”},$$

$$\|A\|_2 = \max_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2} = \sqrt{\lambda_{\max}(A^*A)}, \qquad \text{spectral norm},$$

where $\lambda_{\max}$ denotes the largest eigenvalue.

Proof. We prove the expression for $\|\cdot\|_2$. For the 1-norm and ∞-norm, see Exercise 7.

Since A^*A is Hermitian positive semidefinite, there exists an eigendecomposition $A^*A = Q\Lambda Q^*$ with Q unitary and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and all $\lambda_i \geq 0$. Therefore

$$\begin{aligned}
\|A\|_2 &= \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{x\neq 0} \frac{(x^*A^*Ax)^{1/2}}{\|x\|_2} = \max_{x\neq 0} \frac{(x^*Q\Lambda Q^*x)^{1/2}}{\|x\|_2} \\
&= \max_{x\neq 0} \frac{((Q^*x)^*\Lambda(Q^*x)^{1/2})}{\|Qx\|_2} = \max_{y\neq 0} \frac{(y^*\Lambda y)^{1/2}}{\|y\|_2} = \max_{y\neq 0} \sqrt{\frac{\sum \lambda_i y_i^2}{\sum y_i^2}} \\
&\leq \max_{y\neq 0} \sqrt{\lambda_{\max}} \sqrt{\frac{\sum y_i^2}{\sum y_i^2}} = \sqrt{\lambda_{\max}}
\end{aligned}$$

which is attained by choosing y to be the appropriate column of the identity matrix. $\square$

A norm is **consistent** if it satisfies $\|AB\| \leq \|A\|\|B\|$ whenever the product AB is defined. The Frobenius norm and all subordinate norms are consistent.

For example the "mixed" norm

$$\|A\|_{\infty,1} := \max_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_1} = \max_{i,j} |a_{ij}|$$

is a matrix norm but is not consistent. If $A = B = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$ then $\|AB\|_{\infty,1} > \|A\|_{\infty,1}\|B\|_{\infty,1}$.

Both the 2-norm and the Frobenius norm are invariant under unitary transformations: for unitary $Q \in \mathbb{C}^{m\times m}$ and $Z \in \mathbb{C}^{n\times n}$,

$$\|QAZ\|_2 = \|A\|_2, \qquad \|QAZ\|_F = \|A\|_F.$$

This property has implications for error analysis, for it means that multiplication by unitary matrices does not magnify errors. For example, if $A \in \mathbb{C}^{n\times n}$ is contaminated by errors E and U is unitary (i.e., $U^* = U^{-1}$), then

$$U(A + E)U^* = UAU^* + F,$$

and $\|F\|_2 = \|UEU^*\|_2 = \|E\|_2$. In contrast, if we do a general, nonsingular similarity transformation

$$X(A + E)X^{-1} = XAX^{-1} + G,$$

then $\|G\|_2 = \|XEX^{-1}\|_2 \leq \kappa_2(X)\|E\|_2$, where $\kappa_2(X) = \|X\|_2\|X^{-1}\|_2 \geq 1$ is the **condition number** of X and can be arbitrarily large.

The **spectral radius** $\rho(A)$ of a matrix A is $\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$.

Theorem 2 *Let $A \in \mathbb{C}^{n\times n}$. For any consistent matrix norm $\|\cdot\|$,*

$$\rho(A) \leq \|A\|.$$

Proof. If $Ax = \lambda x$, $x \neq 0$ and if $|\lambda| = \rho(A)$, then form the matrix $X = [\, x, \ldots, x\,]$ and observe that $AX = \lambda X$. For any consistent matrix norm $\|\cdot\|$,

$$|\lambda|\|X\| = \|\lambda X\| = \|AX\| \leq \|A\|\|X\|$$

and therefore $|\lambda| = \rho(A) \leq \|A\|$.

A sequence $\{A^{(k)}\} \in \mathbb{C}^{n\times n}$ converges to a matrix A if $\lim_{k\to\infty} \|A^{(k)} - A\| = 0$. The choice of norms is irrelevant since all norms on $\mathbb{C}^{n\times n}$ are equivalent.

The next theorem characterizes matrices A such that $A^k \to O$ as $k \to \infty$.

Theorem 3 *Let $A \in \mathbb{C}^{n\times n}$. Then*

$$\lim_{k\to\infty} A^k = O \;\Leftrightarrow\; \rho(A) < 1.$$

Proof. Let $Q^*AQ = D + N$ be a Schur decomposition of A, where Q is unitary, D is diagonal and N is strictly upper triangular. For $\theta \geq 1$, define the matrix

$$\Delta = \mathrm{diag}(1, \theta, \theta^2, \ldots, \theta^{n-1}),$$

for which $\kappa_2(\Delta) = \|\Delta\|_2\|\Delta^{-1}\|_2 = \theta^{n-1}$. Since N is strictly upper triangular it is easy to show that $\|\Delta N \Delta^{-1}\|_F \leq \|N\|_F/\theta$. Thus

$$\begin{aligned}
\|A^k\|_2 &= \|(D+N)^k\|_2 \\
&= \|\Delta^{-1}(D + \Delta N \Delta^{-1})^k \Delta\|_2 \\
&\leq \kappa_2(\Delta)\big(\|D\|_2 + \|\Delta N \Delta^{-1}\|_2\big)^k \\
&\leq \theta^{n-1}\left(\rho(A) + \frac{\|N\|_F}{\theta}\right)^k,
\end{aligned}$$

where we used the fact that $\|B\|_2 \leq \|B\|_F$ (see Exercise 6). Since $\rho(A) < 1$, we can choose θ so that $\rho(A) + \|N\|_F/\theta < 1$; then, letting $k \to \infty$ and noting that n is fixed, it follows that $\|A^k\|_2 \to 0$ and hence that $A^k \to O$.

Conversely, if $A^k \to O$ and if $x \neq 0$ is a vector such that $Ax = \lambda x$ then $A^k x = \lambda^k x \to 0$ only if $|\lambda| < 1$. Since this inequality must hold for every eigenvalue of A we conclude that $\rho(A) < 1$.
□

3 Bounds for Eigenvalues

One important area of application of matrix norms is in giving bounds for the spectrum of a matrix. We already know from Theorem 2 that

$$|\lambda| \leq \|A\|$$

for any consistent norm. This means that all the eigenvalues lie in a disk centered at the origin with radius $\|A\|$. A more powerful result is:

Theorem 4 (Gershgorin's theorem, 1931) *The eigenvalues of $A \in \mathbb{C}^{n\times n}$ lie in the union of the n discs in the complex plane*

$$D_i = \Big\{ z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \Big\}, \qquad i = 1, \ldots, n.$$

Proof. Let λ be an eigenvalue of A and x a corresponding eigenvector, and let $|x_k| = \|x\|_\infty$. From the kth equation in $Ax = \lambda x$ we have

$$\sum_{\substack{j=1\\j\neq k}}^{n} a_{kj}x_j = (\lambda - a_{kk})x_k.$$

Hence

$$|\lambda - a_{kk}| \le \sum_{\substack{j=1\\j\neq k}}^{n} |a_{kj}||x_j|/|x_k|,$$

and since $|x_j|/|x_k| \le 1$ it follows that λ belongs to the kth disk, D_k. $\square$

Example 1 Let $A = \begin{bmatrix} 8 & 1 & 0 \\ 1 & 10 & 1 \\ 0 & 1 & 12 \end{bmatrix}$. What can be said about the location of A's eigenvalues?

Since A is symmetric its eigenvalues are real. Since $\|A\|_\infty = \|A\|_1 = 13$ we know that $\lambda \in [-13, 13]$. Gershgorin theorem gives a sharper lower bound: all eigenvalues lie in the interval $[7, 13]$. Actually we can obtain a smaller interval by doing the following. Let $D = \mathrm{diag}(d, 1, 1)$. Then

$$DAD^{-1} = \begin{bmatrix} 8 & d & \\ d^{-1} & 10 & 1 \\ & 1 & 12 \end{bmatrix}$$

is similar to A and has the same eigenvalues. If we apply Gershgorin's theorem to DAD^{-1} we obtain $\min\{8 - d, 10 - 1 - d^{-1}\}$ as the lower bound of the eigenvalues. Thus the lower bounds will be as large as possible when $8 - d = 10 - 1 - d^{-1}$ or when $d = (-1 + \sqrt{5})/2$. Hence all the eigenvalues are larger than 7.38. A similar improvement of the upper bound limit can be made. Note that $\Lambda(A) = \{7.55, 10, 12.45\}$.

Exercises

1. Let $\|\cdot\|$ be a vector norm on $\mathbb{R}^m$ and let $A \in \mathbb{R}^{m\times n}$. Show that if $\mathrm{rank}(A) = n$ then $\|x\|_A = \|Ax\|$ is a vector norm.

2. Define the function $\nu : \mathbb{C}^n \to \mathbb{R}$ by $\nu(x) = \sum_{i=1}^{n}\big(|\operatorname{Re} x_i| + |\operatorname{Im} x_i|\big)$. Is ν a vector norm on $\mathbb{C}^n$?

3. Show that for all $x \in \mathbb{C}^n$, $\|x\|_\infty \le \|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2 \le n\|x\|_\infty$. Show that each inequality is attained for a particular vector.

4. Show that the vector 2-norm $\|\cdot\|_2$ is **unitarily invariant** but that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are not.

5. Verify that $\|xy^*\|_F = \|xy^*\|_2 = \|x\|_2\|y\|_2$ for any $x, y \in \mathbb{C}^n$.

6. Let $A \in \mathbb{C}^{n\times m}$. Show that $\|A\|_2 \le \|A\|_F$. (Recall that $\mathrm{trace}(B) = \sum_i \lambda_i(B)$).

7. For $A \in \mathbb{C}^{m\times n}$, show that

$$\|A\|_1 = \max_{x\neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|, \qquad \|A\|_\infty = \max_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|.$$

8.* Let $A \in \mathbb{C}^{m\times n}$. Show that $\|A\|_2 \le \sqrt{\|A\|_1\|A\|_\infty}$.

9. Let $A \in \mathbb{C}^{n\times n}$ and $\|\cdot\|$ be any consistent matrix norm. Prove that if $\|A\| < 1$ then $I - A$ is nonsingular and $(I-A)^{-1} = \sum_{k=0}^{\infty} A^k$.

10. (i) Show that a Hermitian matrix A is positive definite (i.e. $x^*Ax > 0$ for all nonzero vector x) if and only if all its eigenvalues are real and positive.

(ii) Use Gershgorin's theorem to show that the tridiagonal matrix

$$A = \begin{bmatrix} 3 & i & & & \\ -i & 3 & i & & \\ & -i & \ddots & \ddots & \\ & & \ddots & \ddots & i \\ & & & -i & 3 \end{bmatrix} \in \mathbb{C}^{n\times n}$$

is Hermitian positive definite.

11. Consider the monic scalar polynomial $p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ with $a_0 \neq 0$.

(i) Show that p is the characteristic polynomial of the companion matrix

$$C(p) = \begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_1 & -a_0 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & & 1 & 0 \end{bmatrix}.$$

(ii) Show that if $\widetilde{z}$ is a root of $p(z) = 0$ and if $\|\cdot\|$ is any consistent matrix norm on $\mathbb{C}^{n\times n}$ then $|\widetilde{z}| \le \|C(p)\|$.

(iii) Using $\|\cdot\|_1$ and $\|\cdot\|_\infty$, derive *Cauchy's bound*

$$|\widetilde{z}| \le 1 + \max_{0\le i\le n-1} |a_i|,$$

and *Montel's bound*

$$|\widetilde{z}| \le 1 + \sum_{i=0}^{n-1} |a_i|.$$

12.* A magic square M_n is an $n \times n$ matrix containing the integers from 1 to n^2 whose row and column sums are all the same. For example

$$M_4 = \begin{bmatrix} 16 & 2 & 3 & 13 \\ 5 & 11 & 10 & 8 \\ 9 & 7 & 6 & 12 \\ 4 & 14 & 15 & 1 \end{bmatrix}.$$

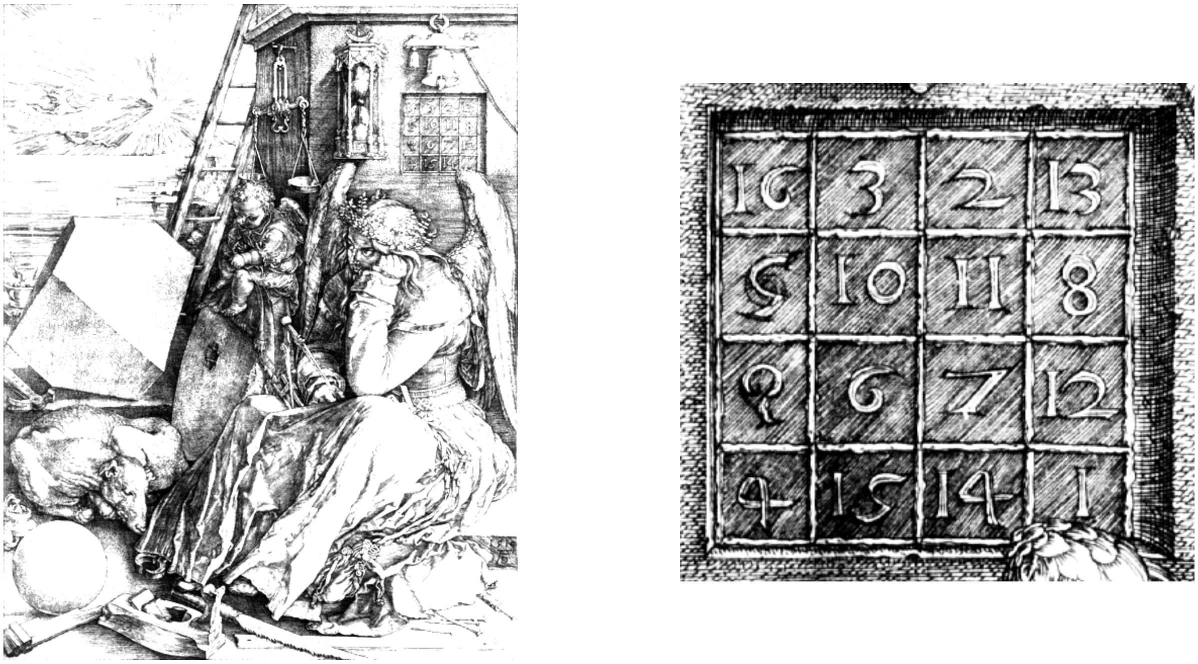

Figure 1: Left: Albrecht Durer's Melancolia. Can you find the matrix? Right: Details of Dürer's magic square. Dürer slipped the date of the painting, 1514, into the bottom row!

This magic square appears in the Renaissance engraving *Melencolia I* by the German painter, engraver and amateur mathematician Albrecht Dürer (1471–1528); see Figure 1.

Let μ_n denote the magic constant of M_n, so that $\mu_n = n(n^2+1)/2$. Let e denote the vector of all 1s.

(i) Determine $M_n e$ and $e^T M_n$. Conclude that μ_n is an eigenvalue of M_n.

(ii) Show that the row and column sums of M_n^2 are all the same. Is M_n^2 a magic square?

(iii) Determine $\|M_n\|_1$, $\|M_n\|_\infty$, $\|M_n\|_2$ and $\rho(M_n)$.

(iv) Assume that M_n has distinct eigenvalues. Show that $\lim_{k\to\infty} M_n^k/\mu_n^k = ee^T/n$, that is, the powers of M_n converge to a rank-one matrix.