CME 302: NUMERICAL LINEAR ALGEBRA
FALL 2005/06
LECTURE 2

GENE H. GOLUB

1. Vector Norms

A real-valued function $f(\mathbf{x})$ defined on a vector space is said to be a *norm* if

(1) For any vector $\mathbf{x}$, $f(\mathbf{x}) \geq 0$, and $f(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
(2) $f(\alpha\mathbf{x}) = |\alpha| f(\mathbf{x})$ for any scalar α.
(3) $f(\mathbf{x}+\mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ for any two vectors $\mathbf{x}$ and $\mathbf{y}$.

We indicate $f(\mathbf{x}) = \|\mathbf{x}\|$.
Consider

$$\nu(\mathbf{x}) = \sum_{i=1}^{n} |x_i|.$$

Is this a norm?

(1) Clearly $\nu(\mathbf{x}) \geq 0$, and the only way that $\nu(\mathbf{x}) = 0$ is if $\mathbf{x} = \mathbf{0}$.
(2) We have
$$\nu(\alpha\mathbf{x}) = \sum_{i=1}^{n} |\alpha x_i| = |\alpha| \sum_{i=1}^{n} |x_i| = |\alpha|\nu(\mathbf{x}).$$
(3) Using the triangle inequality for scalars, we obtain
$$\nu(\mathbf{x}+\mathbf{y}) = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} |x_i| + |y_i| \leq \nu(\mathbf{x}) + \nu(\mathbf{y}).$$

We define the p-norm $\|\mathbf{x}\|_p$ by

$$\|\mathbf{x}\|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p}.$$

The 1-norm

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

is the same as the function $\nu(\mathbf{x})$ discussed above. Another common norm is the 2-norm

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}.$$

It can be shown that for any p, we have

$$(\max_i |x_i|^p)^{1/p} \leq \|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \leq (n \max_i |x_i|)^{1/p}$$

from which it follows that

$$\max_i |x_i| \leq \|\mathbf{x}\|_p \leq n^{1/p} \max_i |x_i|.$$

Date: October 26, 2005, version 1.0.
Notes originally due to James Lambers. Edited by Lek-Heng Lim.

Therefore, as $p \to \infty$, we obtain the *infinity norm*

$$\|\mathbf{x}\|_\infty = \lim_{p\to\infty} \|\mathbf{x}\|_p = \max_i |x_i|.$$

This norm is also known as the *Chebyshev norm.* It is easy to verify that it is in fact a norm.

A variation of the p-norm is the *weighted p-norm*, defined by

$$\|\mathbf{x}\|_{p,\mathbf{w}} = \left(\sum_{i=1}^n w_i |x_i|^p\right)^{1/p}.$$

It can be shown that this is a norm as long as the *weights* w_i, $i = 1, \ldots, n$, are strictly positive. Another common norm is the A-norm, defined in terms of a matrix A by

$$\|\mathbf{x}\|_A = (\mathbf{x}^\top A\mathbf{x})^{1/2},$$

which is a norm provided that the matrix A is positive definite.

We now highlight some additional, and useful, relationships involving norms. First of all, the triangle inequality generalizes directly to sums of more than two vectors:

$$\|\mathbf{x}+\mathbf{y}+\mathbf{z}\| \leq \|\mathbf{x}+\mathbf{y}\| + \|\mathbf{z}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| + \|\mathbf{z}\|,$$

and in general,

$$\left\|\sum_{i=1}^m \mathbf{x}_i\right\| \leq \sum_{i=1}^m \|\mathbf{x}_i\|.$$

What can we say about the norm of the difference of two vectors? While we know that $\|\mathbf{x}-\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, we can obtain a more useful relationship as follows: From

$$\|\mathbf{x}\| = \|(\mathbf{x}-\mathbf{y}) + \mathbf{y}\| \leq \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{y}\|$$

we obtain

$$\|\mathbf{x}-\mathbf{y}\| \geq \|\mathbf{x}\| - \|\mathbf{y}\|.$$

Similarly, from

$$\|\mathbf{y}\| = \|\mathbf{y}-\mathbf{x}+\mathbf{x}\| = \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{x}\|$$

it follows that

$$\|\mathbf{x}-\mathbf{y}\| \geq \|\mathbf{y}\| - \|\mathbf{x}\|$$

and therefore

$$|\|\mathbf{x}\| - \|\mathbf{y}\|| \leq \|\mathbf{x}-\mathbf{y}\|.$$

There are also interesting relationships among different norms. First and foremost, all norms are, in some sense, equivalent. In particular, given two norms $\|\mathbf{x}\|_\alpha$ and $\|\mathbf{x}\|_\beta$, there exist constants c_1 and c_2, independent of $\mathbf{x}$, such that

$$c_1\|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\beta \leq c_2\|\mathbf{x}\|_\alpha.$$

For example, from the definition of the ∞-norm, we have

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty.$$

It is also not hard to show that

$$\frac{1}{n}\|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1.$$

We also have a relationship that applies to products of norms, the *Hölder inequality*

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

A well-known corollary arises when $p = q = 2$, the *Cauchy-Schwarz inequality*

$$|\mathbf{x}^\top\mathbf{y}| \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2.$$

It is interesting to note that by setting $\mathbf{x} = (1, 1, \ldots, 1)$, the Hölder inequality yields the relationships

$$\left|\sum_{i=1}^{n} y_i\right| \leq \sum_{i=1}^{n} |y_i|,$$

$$\left|\sum_{i=1}^{n} y_i\right| \leq n \max_i |y_i|,$$

and

$$\left|\sum_{i=1}^{n} y_i\right| \leq \sqrt{n}\left(\sum_{i=1}^{n} |y_i|^2\right)^{1/2}.$$

A very important property of norms is that they are all continuous functions of the entries of their arguments. It follows that a sequence of vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots$ converges to a vector $\mathbf{x}$ if and only if

$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}\| = 0$$

for any norm. For this reason, norms are very useful to measure the error in an approximation. Three commonly used measures of the error in an approximation $\hat{\mathbf{x}}$ to a vector $\mathbf{x}$ are the *absolute error*

$$\varepsilon_{\text{abs}} = \|\mathbf{x} - \hat{\mathbf{x}}\|,$$

the *relative error*

$$\varepsilon_{\text{rel}} = \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\|\mathbf{x}\|},$$

and the *point-wise error*

$$\varepsilon_{\text{elem}} = \|\mathbf{y}\|, \quad y_i = \frac{\hat{x}_i - x_i}{x_i}.$$

2. Matrix Norms

A real-valued function $f(A)$ defined on the space of $m \times n$ matrices is called a *norm* if

(1) $f(A) \geq 0$, and $f(A) = 0$ if and only if $A = 0$,
(2) $f(A + B) \leq f(A) + f(B)$,
(3) $f(\alpha A) = |\alpha| f(A)$.

Often, we add the condition that $f(A)$ satisfy the *submultiplicative property*

$$f(AB) \leq f(A)f(B).$$

We write $\|A\| = f(A)$. An example of a matrix norm is the *Frobenius norm*

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

On the other hand, the function

$$f(A) = \max_{i,j} |a_{ij}|$$

is not a norm because it does not satisfy the submultiplicative property, but for an appropriate choice of a constant c, the function $f(A) = c \max_{i,j} |a_{ij}|$ is a norm.

Note that the submultiplicative property implies that

$$\|A^n\| \leq \|A\|^n,$$

from which it follows that if $\|A\| < 1$, then, as $n \to \infty$, $\|A^n\| \to 0$, and therefore $A^n \to 0$ as well, due to the continuity of the matrix norm.

Note that the condition $\|A\| < 1$ is not necessary for $A^n \to 0$. For example,

$$A = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \end{bmatrix}$$

has $\|A\|_\infty = 1$ but $A^n \to 0$. One way to see this is that A has an eigendecomposition $A = QDQ^\top$, where $D = \mathrm{diag}[1/\sqrt{2}, 0, -1/\sqrt{2}]$ is a diagonal matrix of eigenvalues and Q an orthogonal matrix of eigenvectors. Since $A^n = QD^nQ^\top$ and $D^n = \mathrm{diag}[2^{-n/2}, 0, -2^{-n/2}] \to 0$, we have that $A^n \to 0$.

How can we define a matrix norm? It is natural to want to define matrix norms in terms of vector norms, but we must be careful in doing so. For example, if we choose to view an $m \times n$ matrix A as a vector $\boldsymbol{\alpha}$ with mn elements, then we have $\|A\|_F = \|\boldsymbol{\alpha}\|_2$. However, if we define a norm $\|A\|_\alpha = \|\boldsymbol{\alpha}\|_\infty$, then the resulting matrix norm does not satisfy the submultiplicative property. To see this, take

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix},$$

but $\|A\|_\alpha = \|B\|_\alpha = 1$, while $\|AB\|_\alpha = 2$.

Instead, we take the approach of defining the *natural norm*

$$\|A\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}.$$

where $\|\mathbf{x}\|$ is any given vector norm. We say that the vector norm $\|\mathbf{x}\|$ *induces* the matrix norm $\|A\|$. Note that

$$\sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \sup_{\mathbf{x} \neq \mathbf{0}} \left\| \frac{1}{\|\mathbf{x}\|} A\mathbf{x} \right\| = \sup_{\mathbf{x} \neq \mathbf{0}} \left\| A \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\| = \sup_{\|\mathbf{y}\|=1} \|A\mathbf{y}\|$$

since $\mathbf{x}/\|\mathbf{x}\|$ always have norm 1 for any $\mathbf{x} \neq \mathbf{0}$.

The set of unit vectors $\{\mathbf{y} \mid \|\mathbf{y}\| = 1\}$ is compact and both the functions $\mathbf{y} \mapsto A\mathbf{y}$ and $\mathbf{z} \mapsto \|\mathbf{z}\|$ are continuous (and thus so is $\|\mathbf{y}\| \mapsto \|A\mathbf{y}\|$). Hence

$$\|A\| = \sup_{\|\mathbf{y}\|=1} \|A\mathbf{y}\|$$

is finite and the supremum is attained, ie. there is some vector $\mathbf{y}_0$ such that $\|\mathbf{y}_0\| = 1$ and $\|A\| = \|A\mathbf{y}_0\|$. Consequently, we may write

$$\|A\| = \max_{\|\mathbf{y}\|=1} \|A\mathbf{y}\|.$$

We will now verify that the natural norm does in fact define a valid matrix norm.

(1) If $A = 0$, then clearly $A\mathbf{x} = \mathbf{0}$ for any $\mathbf{x}$, so we must have $\|A\| = 0$. Otherwise, $a_{ij} \neq 0$ for some i, j. If we let $\mathbf{e}_j$ be the vector containing all zero elements except for a 1 in the jth position, then $\|\mathbf{e}_j\| > 0$ and $\|A\mathbf{e}_j\| > 0$, since $A\mathbf{e}_j$ is the jth column of A, which is a nonzero vector. Therefore

$$\|A\| \geq \frac{\|A\mathbf{e}_j\|}{\|\mathbf{e}_j\|} > 0.$$

(2) Let α be a scalar. By the scaling property of the vector norm, we know that

$$\|\alpha A\mathbf{y}\| = |\alpha| \|A\mathbf{y}\|$$

for any $\mathbf{y}$. Hence we may take supremum over all $\mathbf{y}$ with $\|\mathbf{y}\| = 1$ to get

$$\sup_{\|\mathbf{y}\|=1} \|\alpha A\mathbf{y}\| = |\alpha| \sup_{\|\mathbf{y}\|=1} \|A\mathbf{y}\|.$$

It then follows that $\|\alpha A\| = |\alpha| \|A\|$.

(3) Before we prove that the triangle inequality holds, it is useful to note that for any matrix A and vector $\mathbf{z}$,

$$\begin{aligned}\|A\mathbf{z}\| &= \left\| A\|\mathbf{z}\|\frac{\mathbf{z}}{\|\mathbf{z}\|}\right\| \\ &= \|\mathbf{z}\|\left\| A\frac{\mathbf{z}}{\|\mathbf{z}\|}\right\| \\ &\leq \|\mathbf{z}\|\|A\|.\end{aligned}$$

Then, for some vector $\mathbf{y}_0$ satisfying $\|\mathbf{y}_0\| = 1$, we have

$$\begin{aligned}\|A+B\| &= \|(A+B)\mathbf{y}_0\| \\ &\leq \|A\mathbf{y}_0\| + \|B\mathbf{y}_0\| \\ &\leq \|A\| + \|B\|.\end{aligned}$$

(4) Using the property $\|A\mathbf{z}\| \leq \|A\|\|\mathbf{z}\|$, it follows that for some vector $\mathbf{y}_0$ satisfying $\|\mathbf{y}_0\| = 1$, we have

$$\begin{aligned}\|AB\| &= \|A(B\mathbf{y}_0)\| \\ &\leq \|A\|\|B\mathbf{y}_0\| \\ &\leq \|A\|\|B\|\|\mathbf{y}_0\| = \|A\|\|B\|.\end{aligned}$$

We will now try to arrive at an explicit expression for $\|A\|_\infty$. We have

$$\begin{aligned}\|A\|_\infty &= \max_{\|\mathbf{y}\|_\infty=1} \|A\mathbf{y}\| \\ &= \max_{\|\mathbf{y}\|_\infty=1} \max_i \left|\sum_{j=1}^n a_{ij}y_j\right| \\ &\leq \max_{\|\mathbf{y}\|_\infty=1} \max_i \sum_{j=1}^n |a_{ij}||y_j| \\ &\leq \max_{\|\mathbf{y}\|_\infty=1} \max_i \max_j |y_j| \sum_{j=1}^n |a_{ij}| \\ &\leq \max_i \sum_{j=1}^n |a_{ij}|.\end{aligned}$$

Now, suppose that

$$\max_i \sum_{j=1}^n |a_{ij}| = \sum_{j=1}^n |a_{Ij}|.$$

Let $\mathbf{y}$ be the vector with elements

$$y_j = \begin{cases} +1 & a_{Ij} \geq 0, \\ -1 & a_{Ij} < 0. \end{cases}$$

Then $\|\mathbf{y}\|_\infty = 1$ and

$$\|A\mathbf{y}\|_\infty = \sum_{j=1}^n |a_{Ij}| = \max_i \sum_{j=1}^n |a_{ij}|.$$

Therefore the upper bound we obtained for $\|A\|_\infty$ is actually assumed for some unit vector $\mathbf{y}$, from which it follows that

$$\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|.$$

3. Quick summary

Vector norms have the following defining properties:

(1) $\|\mathbf{x}\| > 0$ for all $\mathbf{x} \neq \mathbf{0}$, and $\|\mathbf{0}\| = 0$,
(2) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$,
(3) $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$.

Matrix norms, in addition to the properties above, frequently have the submultiplicative property:

(1) $\|A\| > 0$ for all $A \neq 0$, and $\|0\| = 0$,
(2) $\|A + B\| \leq \|A\| + \|B\|$,
(3) $\|\alpha A\| = |\alpha| \|A\|$,
(4) $\|AB\| \leq \|A\| \|B\|$.

We rarely consider matrix norms that do not have the submultiplicative property.

4. The Matrix 2-Norm

We have establisheded that the vector ∞-norm

$$\|\mathbf{x}\|_\infty = \max_i |x_i|$$

induced the matrix ∞-norm

$$\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|.$$

Similarly, (ie. we leave it as an exercise) the vector 1-norm

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$$

induces the matrix 1-norm

$$\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}|.$$

It is natural to ask whether there is a similar explicit expression for the matrix 2-norm induced by the vector 2-norm

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2}.$$

One might suggest the Frobenius norm

$$\|A\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2},$$

but that is incorrect. We will now derive an explicit expression for the induced matrix 2-norm.

Recalling the definition of the matrix 2-norm,

$$\|A\|_2 = \max_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2},$$

we examine the expression

$$\|A\mathbf{x}\|_2^2 = x^\top A^\top A x.$$

The matrix $A^\top A$ is symmetric and positive semi-definite, i.e. $\mathbf{x}^\top (A^\top A) \mathbf{x} \geq 0$ for all nonzero $\mathbf{x}$. As such, it has the decomposition

$$A^\top A = U \Sigma U^\top$$

where U is an orthogonal matrix whose columns are the eigenvectors of $A^\top A$, and Σ is a diagonal matrix of the form

$$\begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix}$$

where each σ_i is nonnegative and an eigenvalue of $A^\top A$. These eigenvalues can be ordered such that

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0, \quad \sigma_{r+1}^2 = \cdots = \sigma_n^2 = 0,$$

where r is the rank of A. If we define $\mathbf{w} = U^\top \mathbf{x}$, then we obtain

$$\begin{aligned} \frac{\|A\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} &= \frac{\mathbf{x}^\top A^\top A\mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \\ &= \frac{\mathbf{x}^\top U\Sigma U^\top \mathbf{x}}{\mathbf{x}^\top U^\top U\mathbf{x}} \\ &= \frac{\mathbf{w}^\top \Sigma\mathbf{w}}{\mathbf{w}^\top \mathbf{w}} \\ &= \frac{\sum_{i=1}^n \sigma_i^2 |w_i|^2}{\sum_{i=1}^n |w_i|^2} \\ &\leq \sigma_1^2. \end{aligned}$$

It follows that

$$\frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \sigma_1$$

for all nonzero $\mathbf{x}$, but we must now determine whether this inequality is actually an equality for any $\mathbf{x}$. Since U is an orthogonal matrix, it follows that there exists an $\mathbf{x}$ such that

$$\mathbf{w} = U^\top \mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{e}_1,$$

in which case

$$\mathbf{x}^\top A^\top A\mathbf{x} = \mathbf{e}_1^\top \Sigma \mathbf{e}_1 = \sigma_1^2.$$

In fact, this vector $\mathbf{x}$ is the eigenvector of $A^\top A$ corresponding to the eigenvalue σ_1^2.

We conclude that

$$\|A\|_2 = \sigma_1.$$

5. The Spectral Radius

The matrix 2-norm is also known as the *spectral norm*. This name is connected to the fact that the norm is given by the square root of the largest eigenvalue of $A^\top A$, and, in general, the *spectral radius* $\rho(A)$ of a matrix A is defined in terms of its largest eigenvalue:

$$\rho(A) = \max_i |\lambda_i|, \quad A\mathbf{x}_i = \lambda_i \mathbf{x}_i, \quad \mathbf{x}_i \neq \mathbf{0}.$$

We now discuss some relationships between the norm of a matrix and its spectral radius. First, suppose that

$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i.$$

Then, for any matrix norm,

$$\|A\mathbf{x}_i\| = \|\lambda_i \mathbf{x}_i\| = |\lambda_i| \|\mathbf{x}_i\|.$$

Therefore

$$|\lambda_i| = \frac{\|A\mathbf{x}_i\|}{\|\mathbf{x}_i\|} \leq \|A\|.$$

Since this holds for *any* eigenvalue of A, it follows that

$$\max_i |\lambda_i| = \rho(A) \leq \|A\|.$$

We also have the following result:

Theorem 1. *For every $\epsilon > 0$, there exists a matrix norm $\|A\|_\alpha$ such that*

$$\|A\|_\alpha \leq \rho(A) + \epsilon.$$

The norm is dependent on A and ϵ.

This result suggests that the largest eigenvalue of a matrix can be easily approximated. A simple example is the case of the identity matrix I, whose only eigenvalue is 1, and whose norm is equal to 1 for any natural matrix norm.

As another example, let

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & 1 \\ & & & -1 & 2 \end{bmatrix}.$$

The eigenvalues of this matrix, which arises frequently in numerical methods for solving differential equations, are known to be

$$\lambda_j = 2 + 2\cos\frac{j\pi}{n+1}, \quad j = 1, 2, \ldots, n.$$

The largest eigenvalue is

$$|\lambda_1| = 2 + 2\cos\frac{\pi}{n+1} \leq 4,$$

and $\|A\|_\infty = 4$, so in this case, the ∞-norm provides an excellent approximation.

On the other hand, suppose

$$A = \begin{bmatrix} 1 & 10^6 \\ 0 & 1 \end{bmatrix}.$$

We have $\|A\|_\infty = 10^6 + 1$, but $\rho(A) = 1$, so in this case the norm yields a poor approximation. However, suppose

$$D = \begin{bmatrix} \epsilon & 0 \\ 0 & 1 \end{bmatrix}.$$

Then

$$DAD^{-1} = \begin{bmatrix} 1 & 10^6\epsilon \\ 0 & 1 \end{bmatrix},$$

and $\|DAD^{-1}\|_\infty = 1 + 10^{-6}\epsilon$,, which for sufficiently small ϵ, yields a much better approximation to $\rho(DAD^{-1}) = \rho(A)$.

We can also show that all matrix norms are equivalent: for any two matrix norms $\|A\|_\alpha$ and $\|A\|_\beta$, there exist constants c_1 and c_2 such that

$$c_2\|A\|_\alpha \leq \|A\|_\beta \leq c_1\|A\|_\alpha,$$

and the constants c_1 and c_2 are independent of the matrix A. For example, for any $m \times n$ matrix A,

$$\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\|A\|_\infty.$$

6. Gerschgorin's Theorem

Suppose that λ is an eigenvalue of A with corresponding eigenvector $\mathbf{x}$; i.e.

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

Then, for $i = 1, 2, \ldots, n$, we have

$$\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i.$$

Rearranging, we obtain

$$(a_{ii} - \lambda)x_i = -\sum_{j\neq i} a_{ij}x_j,$$

and, taking absolute values yields

$$|a_{ii} - \lambda||x_i| \leq \sum_{j\neq i} |a_{ij}||x_j|.$$

Now, suppose that $|x_I| \geq |x_i|$, for $i = 1, 2, \ldots, n$. Then

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

Since this bound applies to any eigenvalue of A, we can conclude that each eigenvalue of A lies within the union of the *Gerschgorin disks* defined by

$$|\lambda - a_{ii}| \leq r_i, \quad r_i = \sum_{j\neq i} |a_{ij}|, \quad i = 1, 2, \ldots, n.$$

This result is known as *Gerschgorin's Theorem*.

Suppose that $A = D + K$, where D is a diagonal matrix with diagonal entries $d_{ii} = a_{ii}$, and K represents the off-diagonal portion of A, with entries

$$K_{ij} = \begin{cases} a_{ij} & i \neq j \\ 0 & i = j \end{cases}$$

Then, define $A(\epsilon) = D+\epsilon K$. Then $A(0) = D$ and $A(1) = A$. The eigenvalues of $A(\epsilon)$ are continuous functions of ϵ, so we can approximate the eigenvalues of A by examining how the Gerschgorin disks change as ϵ changes. In particular, we can determine how many eigenvalues lie within individual disks or unions of disks.

Department of Computer Science, Gates Building 2B, Room 280, Stanford, CA 94305-9025
E-mail address: golub@stanford.edu