Topics: The vector norm of a matrix

Let \( \| \| \) denote a norm on \( \mathbb{R}_m \) and \( \mathbb{R}_n \). Typically, we think of \( \| x \| = \| x \|_\infty = \max_i |x_i| \), but it can be any norm.

We define the vector norm of a matrix \( A \) by

\[
\| A \| = \max_{\| x \| = 1} \| Ax \|.
\]

We say that the vector norm \( \| A \| \) is “induced” by the norm \( \| \| \). It is a measure of the “size” of the operator.

It is straightforward to show that this definition yields a norm on the vector space of all \( m \times n \) matrices. Moreover, for any vector \( x \neq 0 \) we have that

\[
\| Ax \| = \left\| A \frac{x}{\| x \|} \right\| \| x \| \leq \| A \| \| x \|,
\]

and

\[
\| ABx \| \leq \| A \| \| Bx \| \leq \| A \| \| B \| \| x \|,
\]

so that

\[
\| AB \| \leq \| A \| \| B \|.
\]

Finding the actual number for the norm of a matrix may be complicated for some norms on \( \mathbb{R}_n \). However, for the infinity norm it is easy.

Terminology: Given an \( m \times n \) matrix \( A \) its maximum row sum is the number

\[
R = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|.
\]

Theorem: The vector norm of a matrix \( A \) induced by the infinity norm is equal to its maximum row sum.

Proof: Let \( \| x \|_\infty = 1 \) then by definition \( |x_i| \leq 1 \) and \( |x_k| = 1 \) for some \( k \). Then

\[
\| Ax \|_\infty = \max_i \left| \sum_{j=1}^{n} a_{ij} x_j \right| \leq \max_i \sum_{j=1}^{n} |a_{ij}| = R.
\]
Hence the maximum row sum is always greater than or equal to the infinity vector norm of $A$.

Conversely, suppose the maximum row sum is obtained from row $k$ of the matrix $A$. Then choose the vector $x$ defined by

$$x_j = 1 \quad \text{if } a_{kj} \geq 0$$

$$x_j = -1 \quad \text{if } a_{kj} < 0.$$

Then $\|x\|_\infty = 1$ and

$$\|A\|_\infty \geq \|Ax\|_\infty \geq \left| \sum_{j=1}^{n} a_{kj}x_j \right| = \sum_{j=1}^{n} |a_{kj}| = R.$$

Hence here we have a specific vector of length 1 for which the vector norm $A$ dominates the maximum row sum. Therefore,

$$\|A\|_\infty = R.$$

**Application:** Suppose that the $n \times n$ matrix $A$ is strictly diagonally dominant, i.e.

$$|a_{ii}| > \sum_{j=1, j \neq i}^{n} |a_{ij}| \quad \text{for } i = i, \ldots, n.$$

Suppose we want to solve $Ax = b$.

First we observe that if $\lambda$ is an eigenvalue of $A$ with eigenvector $x$ then we can scale the eigenvector so that its maximum component is equal to +1 for some component $k$. Then it follows from the eigenvalue equation

$$(a_{kk} - \lambda)x_k + \sum_{j=1, j \neq k}^{n} a_{kj}x_j = 0$$

and strict diagonal dominance that

$$|a_{kk} - \lambda| \leq \sum_{j=1, j \neq k}^{n} |a_{kj}x_j| \leq \sum_{j=1, j \neq k}^{n} |a_{kj}| < |a_{kk}|.$$

This strict inequality implies that $\lambda \neq 0$. Hence the null space of $A$ contains only the zero vector so that

$$Ax = b$$
has a unique solution.

The solution can be found iteratively. We write

\[ A = D - B \]

where \( D \) is the diagonal of \( A \) and \( B = A - D \). The solution \( x^* \) of \( Ax = b \) satisfies the equation

\[ x = D^{-1}Bx + D^{-1}b \]

where \( D^{-1} = \text{diag}\{1/a_{11}, 1/a_{22}, \ldots, 1/a_{nn}\} \).

We shall find it from the so-called Jacobi iteration

\[ x^{k+1} = D^{-1}Bx^k + D^{-1}b \]

where \( x^0 \) is an initial guess. The advantage of such iterative solution is its applicability to huge linear systems where the entries of \( A \) are mostly zero and do not enter into the actual computation. However, the iteration will not always converge. The next theorem gives insight when the iteration will work.

**Theorem:** For a strictly diagonally dominant matrix \( A \) the Jacobi iteration converges to the unique solution \( x^* \) of \( Ax = b \).

**Proof:** Let \( e^k \) denote the error in iteration \( k \)

\[ e^k = x^k - x^*. \]

Then

\[ e^{k+1} = D^{-1}Be^k = (D^{-1}B)^{k+1}e^0 \]

so that

\[ \|e^k\| = \|(D^{-1}B)^ke^0\| \leq \|(D^{-1}B)^k\|\|e^0\|. \]

If \( (D^{-1}B)^k \to 0 \) as \( k \to \infty \) then \( e^k \to 0 \) so that

\[ \lim_{k \to \infty} x^k = x^*. \]
It follows from the vector norm properties of matrices that

\[ \|(D^{-1}B)^k\| \leq \|D^{-1}B\|^k. \]

From the above theorem we know that

\[ D^{-1}B = \max_i \left| \sum_{j=1}^{n} \frac{a_{ij}}{a_{ii}} \right|. \]

Strict diagonal dominance implies that

\[ \|D^{-1}B\|_\infty \leq R < 1. \]

Thus

\[ \|e^k\| < R^k\|e^0\| \to 0 \quad \text{as} \quad k \to \infty. \]

The condition \( \|A\| < 1 \) for a square matrix \( A \) in some vector norm insures that

\[ \lim_{k \to \infty} A^k = 0 \quad \text{(the zero matrix)} \]

because

\[ 0 \leq \|A^k\| < \|A\|^k. \]

However, it may well be that in a particular norm

\[ \|A\| > 1 \quad \text{and yet that} \quad \lim_{k \to \infty} A^k \to 0. \]

For example,

\[ A = \begin{pmatrix} 0 & 10^{10} \\ 0 & 0 \end{pmatrix} \]

satisfies \( \|A\|_\infty = 10^{10} \)

but \( A^k = 0 \) for \( k \geq 2. \)

The relevant question for applications is: Given a square matrix \( A \) what is the choice of vector norm for which \( \|A\| \) is smallest?

We observe first that for any eigenvalue and eigenvector of \( A \) we have

\[ \|Ax\| = |\lambda|\|x\| \]

76
regardless of the norm chosen. Moreover, if in this norm the eigenvector is scaled to have length 1 then it follows that

\[ |\lambda| = \|Ax\| \leq \|A\|. \]

Hence in any vector norm chosen for \( R^n \) the associated norm of \( A \) must satisfy

\[ \rho(A) = \max_i |\lambda_i| \leq \|A\|. \]

\( \rho(A) \) is known as the spectral radius of the matrix \( A \) and represents a lower bound on any vector norm of \( A \).

There are infinitely many norms which can be imposed on \( R^n \) and each induces a vector norm on the \( n \times n \) matrices. For example, if we define \( \|x\| = \|Cx\|_p \) for \( p \in [1, \infty] \) for a given non-singular matrix \( C \) then \( \| \| \) is a norm on \( R^n \). Now, if \( A \) is an \( n \times n \) matrix and we consider the vector norm induced by \( \| \| \), then we have by definition

\[ \|A\| = \max_{\|x\|=1} \|Ax\| = \max_{\|x\|=1} \|CAx\|_p = \max_{\|Cx\|_p=1} \|CAx\|_p. \]

But for each such \( x \) there is a unit vector \( y \) such that \( Cx = y \). Then

\[ \|A\| = \max_{\|y\|_p=1} \|CA^{-1}y\|_p \]

so that

\[ \|A\| = \|CA^{-1}\|_p. \]

The transformation of \( A \) into \( CA^{-1} \) is called a similarity transformation. We have seen above if the \( n \times n \) matrix \( A \) has \( n \) linearly independent eigenvectors then it follows from the eigenvector equations

\[ AX = X\Lambda \]

where the \( j \)th column of \( X \) is the eigenvector corresponding to the eigenvalue \( \lambda_j \) that in the norm

\[ \|x\| = \|X^{-1}x\|_p \]

the induced matrix norm \( \|A\| \) is

\[ \|A\| = \|X^{-1}AX\|_p = \|\Lambda\|_p. \]
In particular, if $p = \infty$ then $\|A\|$ is equal to the maximum row sum of $A$ so that

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

In this case the size of $A$ is equal to its spectral radius and $A^k \to 0$ as $k \to \infty$ whenever $\rho(A) < 1$. This analysis applies to matrices with $n$ distinct eigenvalues and to Hermitian matrices.

Not every square matrix is similar to a diagonal matrix. However, it is possible, but not easy, to prove via the Jordan canonical form that for any $\epsilon > 0$ the matrix $A$ can be transformed with a similarity transformation into a matrix whose diagonal entries are the $n$ eigenvalues of $A$, and whose entries $a_{i-1,i}$ are either $0$ or $\epsilon$. Thus, even when $A$ is not diagonalizeable there is a norm on $\mathbb{R}^n$ such that

$$\|A\| = \rho(A) + \epsilon.$$ 

This implies that if $\rho(A) < 1$ then for sufficiently small $\epsilon > 0$

$$\|A\| < 1$$

so that again $A^k \to 0$ as $k \to \infty$. The matrix $C$ in this similarity transformation is generally not available but also not needed for the conclusion that $A^k \to 0$.

Hence the smallest possible vector norm of a matrix $A$ is roughly equal to its spectral radius.