Matrix Norms

P. Sam Johnson

National Institute of Technology Karnataka (NITK)
Surathkal, Mangalore, India

Introduction

The analysis of matrix algorithms frequently requires use of matrix norms.

For example, the quality of a linear system solver may be poor if the matrix of coefficients is "Dearly singular."

To quantify the notion of near-singularity we need a measure of distance on the space of matrices. Matrix norms provide that measure.

Definitions

Since $\mathbb{R}^{m\times n}$ is isomorphic to $\mathbb{R}^{mn}$, the definition of a matrix norm should be equivalent to the definition of a vector norm. In particular, $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ is a matrix norm if the following three properties hold:

$$\begin{array}{lll} f(A) \geq 0 & A \in \mathbb{R}^{m\times n}, & (f(A) = 0 \text{ iff } A = 0) \\ f(A+B) \leq f(A) + f(B) & A, B \in \mathbb{R}^{m\times n}, & \\ f(\alpha A) = |\alpha| f(A) & \alpha \in \mathbb{R}, A \in \mathbb{R}^{m\times n}. & \end{array}$$

As with vector norms, we use a double bar notation with subscripts to designate matrix norms, i.e., $\|A\| = f(A)$.

Definitions (Contd...)

The most frequently used matrix norms in numerical linear algebra are the Frobenius norm,

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} \tag{1}$$

and the p-norms

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}. \tag{2}$$

Note that the matrix p-norms are defined in terms of the vector p-norms that we discussed in the previous section. The verification that (1) and (2) are matrix norms is left as an exercise. It is clear that $\|A\|_p$ is the p-norm of the largest vector obtained by applying A to a unit p-norm vector:

$$\|A\|_p = \sup_{x \neq 0} \left\| A \left(\frac{x}{\|x\|_p} \right) \right\|_p = \max_{\|x\|_p = 1} \|Ax\|_p.$$

Definitions (Contd...)

It is important to understand that (1) and (2) define families of norms-the 2-norm on $\mathbb{R}^{3\times 2}$ is a different function from the 2-norm on $\mathbb{R}^{5\times 6}$.

Thus, the easily verified inequality

$$\|AB\|_p \leq \|A\|_p \|B\|_p \qquad A \in \mathbb{R}^{m\times n},\ B \in \mathbb{R}^{n\times q} \tag{3}$$

is really an observation about the relationship between three different norms. Formally, we say that norms f_1, f_2, and f_3 on $\mathbb{R}^{m\times q}, \mathbb{R}^{m\times n}$, and $\mathbb{R}^{n\times q}$ are mutually consistent if for all $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times q}$ we have $f_1(AB) \leq f_2(A) f_3(B)$.

Definitions (Contd...)

Not all matrix norms satisfy the submultiplicative property

$$\|AB\| \leq \|A\| \, \|B\|. \tag{4}$$

For example, if $\|A\|_\Delta = \max |a_{ij}|$ and

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

then $\|AB\|_\Delta > \|A\|_\Delta \|B\|_\Delta$. For the most part we work with norms that satisfy (4).

Definitions (Contd...)

The p-norms have the important property that for every $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$ we have $\|Ax\|_p \leq \|A\|_p \|x\|_p$. More generally, for any vector norm $\|\cdot\|_\alpha$ on $\mathbb{R}^n$ and $\|\cdot\|_\beta$ on $\mathbb{R}^m$ we have $\|Ax\|_\beta \leq \|A\|_{\alpha,\beta} \|x\|_\alpha$ where $\|A\|_{\alpha,\beta}$ is a matrix norm defined by

$$\|A\|_{\alpha,\beta} = \sup_{x \neq 0} \frac{\|Ax\|_\beta}{\|x\|_\alpha}. \tag{5}$$

We say that $\|\cdot\|_{\alpha,\beta}$ is subordinate to the vector norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$. Since the set $\{x \in \mathbb{R}^n : \|x\|_\alpha = 1\}$ is compact and $\|\cdot\|_\beta$ is continuous, it follows that

$$\|A\|_{\alpha,\beta} = \max_{\|x\|_\alpha = 1} \|Ax\|_\beta = \|Ax^*\|_\beta \tag{6}$$

for some $x^* \in \mathbb{R}^n$ having unit α-norm.

Some Matrix Norm Properties

The Frobenius and p-norms (especially $p = 1, 2, \infty$) satisfy certain inequalities that are frequently used in the analysis of matrix computations. For $A \in \mathbb{R}^{m \times n}$ we have

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2 \tag{7}$$

$$\max_{i,j} |a_{ij}| \leq \|A\|_2 \leq \sqrt{mn} \max_{ij} |a_{ij}| \tag{8}$$

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}| \tag{9}$$

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}| \tag{10}$$

$$\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\|A\|_\infty \tag{11}$$

$$\frac{1}{\sqrt{m}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\|A\|_1 \tag{12}$$

Some Matrix Norm Properties (Contd...)

If $A \in \mathbb{R}^{m \times n}$, $1 \leq i_1 \leq i_2 \leq m$, and $1 \leq j_1 \leq j_2 \leq n$, then

$$\|A(i_1 : i_2, j_1 : j_2)\|_p \leq \|A\|_p \tag{13}$$

The proofs of these relations are not hard and are left as exercises.

A sequence $\{A^{(k)}\} \in \mathbb{R}^{m \times n}$ converges if $\lim_{k \to \infty} \|A^{(k)} - A\| = 0$. Choice of norm is irrelevant since all norms on $\mathbb{R}^{m \times n}$ are equivalent.

The Matrix 2-Norm

A nice feature of the matrix 1-norm and the matrix ∞-norm is that they are easily computed from (9) and (10). A characterization of the 2-norm is considerably more complicated.

Theorem 1.

If $A \in \mathbb{R}^{m \times n}$, then there exists a unit 2-norm n-vector z such that $A^T A z = \mu^2 z$ where $\mu = \|A\|_2$.

Proof: Suppose $z \in \mathbb{R}^n$ is a unit vector such that $\|Az\|_2 = \|A\|_2$. Since z maximizes the function

$$g(x) = \frac{1}{2}\frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{1}{2}\frac{x^T A^T A x}{x^T x}$$

it follows that it satisfies $\nabla g(z) = 0$ where ∇g is the gradient of g. But a tedious differentiation shows that for $i = 1 : n$.

The Matrix 2-Norm (Contd...)

$$\frac{\partial g(z)}{\partial z_i} = \left[(z^T z) \sum_{j=1}^{n} (A^T A)_{ij} z_j - (z^T A^T A z) z_i \right] / (z^T z)^2.$$

In vector notation this says $A^T A z = (z^T A^T A z) z$. The theorem follows by setting $\mu = \|Az\|_2$.

The theorem implies that $\|A\|_2^2$ is a zero of the polynomial $p(\lambda) = \det(A^T A - \lambda I)$. In particular, the 2-norm of A is the square root of the largest eigenvalue of $A^T A$.

For now, we merely observe that 2-norm computation is iterative and decidedly more complicated than the computation of the matrix 1-norm or ∞-norm. Fortunately, if the object is to obtain an order-of-magnitude estimate of $\|A\|_2$, then (7), (11), or (12) can be used.

The Matrix 2-Norm (Contd...)

As another example of "norm analysis," here is a handy result for 2-norm estimation.

Corollary 2.

If $A \in \mathbb{R}^{m \times n}$, *then* $\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$.

Proof: If $z \neq 0$ is such that $A^T A z = \mu^2 z$ with $\mu = \|A\|_2$, then $\mu^2 \|z\|_1 = \|A^T A z\|_1 \leq \|A^T\|_1 \|A\|_1 \|z\|_1 = \|A\|_\infty \|A\|_1 \|z\|_1$.

Perturbations and the Inverse

We frequently use norms to quantify the effect of perturbations or to prove that a sequence of matrices converges to a specified limit. As an illustration of these norm applications, let us quantify the change in A^{-1} as a function of change in A.

Lemma 3.

If $F \in \mathbb{R}^{n\times n}$ and $\|F\|_p < 1$, then $I - F$ is nonsingular and

$$(I - F)^{-1} = \sum_{k=0}^{\infty} F^k$$

with

$$\|(I - F)^{-1}\|_p \leq \frac{1}{1 - \|F\|_p}.$$

Perturbations and the Inverse (Contd...)

Proof: Suppose $I - F$ is singular. It follows that $(I - F)x = 0$ for some nonzero x. But then $\|x\|_p = \|Fx\|_p$ implies $\|F\|_p \geq 1$, a contradiction. Thus, $I - F$ is nonsingular. To obtain an expression for its inverse consider the identity

$$\left(\sum_{k=0}^{N} F^k\right)(I - F) = I - F^{N+1}.$$

Since $\|F\|_p < 1$ it follows that $\lim_{k \to \infty} F^k = 0$ because $\|F^k\|_p \leq \|F\|_p^k$. Thus,

$$\left(\lim_{N \to \infty} \sum_{k=0}^{N} F^k\right)(I - F) = I.$$

Perturbations and the Inverse (Contd...)

It follows that $(I - F)^{-1} = \lim_{N\to\infty} \sum_{k=0}^{N} F^k$. From this it is easy to show that

$$\|(I - F)^{-1}\|_p \leq \sum_{k=0}^{\infty} \|F\|_p^k = \frac{1}{1 - \|F\|_p}.$$

Note that $\|(I - F)^{-1} - I\|_p \leq \|F\|_p/(1 - \|F\|_p)$ as a consequence of the lemma.

Thus, if $\varepsilon << 1$, then $O(\varepsilon)$ perturbations in I induce $O(\varepsilon)$ perturbations in the inverse. We next extend this result to general matrices.

Perturbations and the Inverse (Contd...)

Theorem 4.

If A is nonsingular and $r = \|A^{-1}E\|_p < 1$, then $A + E$ is nonsingular and $\|(A+E)^{-1} - A^{-1}\|_p \leq \|E\|_p \|A^{-1}\|_p^2/(1-r)$.

Proof: Since A is nonsingular $A + E = A(I - F)$ where $F = -A^{-1}E$. Since $\|F\|_p = \tau < 1$ it follows from Lemma 3 that $I - F$ is nonsingular and $\|(I-F)^{-1}\|_p < 1/(1-r)$. Now $(A+E)^{-1} = (I-F)^{-1}A^{-1}$ and so

$$\|(A+E)^{-1}\|_p \leq \frac{\|A^{-1}\|_p}{1-\tau}.$$

Equation (2.1.3) says that $(A+E)^{-1} - A^{-1} = -A^{-1}E(A+E)^{-1}$ and so by taking norms we find

$$\begin{aligned} \|(A+E)^{-1} - A^{-1}\|_p &\leq \|A^{-1}\|_p \|E\|_p \|(A+E)^{-1}\|_p \\ &\leq \frac{\|A^{-1}\|_p^2 \|E\|_p}{1-\tau}. \end{aligned}$$

Exercises 5.

1. *Show* $\|AB\|_p \leq \|A\|_p \|B\|_p$ *where* $1 \leq p \leq \infty$.
2. *Let* B *be any submatrix of* A. *Show that* $\|B\|_p \leq \|A\|_p$.
3. *Show that if* $D = diag(\mu_1, \ldots, \mu_k) \in \mathbb{R}^{m \times n}$ *with* $k = \min\{m, n\}$, *then* $\|D\|_p$
4. *Verify* (7),(8), (9), (10), (11), (12) *and* (13).
5. *Show that if* $0 \neq s \in \mathbb{R}^n$ *and* $E \in \mathbb{R}^{n \times n}$, *then*

$$\left\| E \left(1 - \frac{ss^T}{s^T s} \right) \right\|_F^2 = \|E\|_F^2 - \frac{\|Es\|_2^2}{s^T s}.$$

Exercises 6.

6. *Suppose* $u \in \mathbb{R}^m$ *and* $v \in \mathbb{R}^n$. *Show that if* $E = uv^T$ *then* $\|E\|_F = \|E\|_2 = \|u\|_2\|v\|_2$ *and that* $\|E\|_\infty \leq \|u\|_\infty\|v\|_1$.
7. *Suppose* $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$, *and* $0 \neq s \in \mathbb{R}^n$. *Show that* $E = (y - As)s^T/s^Ts$ *has the smallest 2-norm of all m-by-n matrices* E *that satisfy* $(A + E)s = y$.

Reference Books

1. Gene H. Golub and Charles F. Van Loan, Matrix Computations, 3rd Edition, Hindustan book agency, 2007.
2. A.R. Gourlay and G.A. Watson, Computational methods for matrix eigen problems, John Wiley & Sons, New York, 1973.
3. W.W. Hager, Applied numerical algebra, Prentice-Hall, Englewood Cliffs, N.J, 1988.
4. D.S. Watkins, Fundamentals of matrix computations, John Wiley and sons, N.Y, 1991.
5. C.F. Van Loan, Introduction to scientific computing: A Matrix vector approach using Matlab, Prentice-Hall, Upper Saddle River, N.J, 1997.