CSE/Math 455
Lecture # 12

Matrix Norms and Condition Numbers

Last time we defined the notion of a matrix norm. Three examples of matrix norms are

$$\begin{aligned}
\|A\|_F &= \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}, \quad \text{The Frobenius Norm} \\
\|A\|_1 &= \max_{1\le j\le n}\sum_{i=1}^{m}|a_{ij}|, \quad \text{The One-Norm} \\
\mathbf{x}_1^* &= \mathbf{e}_{j_{max}}, \quad j_{max} \text{ Max column} \\
\|A\|_\infty &= \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|, \quad \text{The Infinity-Norm} \\
\mathbf{x}_\infty^* &= (x_1^*,\ldots,x_n^*)^T, \quad x_j^* = \text{sign}(a_{i_{max},j}).
\end{aligned}$$

In the last equation i_{max} is the "maximum row" for the ∞-norm. In MATLAB these are **norm**$(A,'fro')$, **norm**$(A,1)$, and **norm**$(A,'inf')$.

The last two norms are *operator norms.* An operator norm is defined by

$$\|A\|_\alpha = \max_{\|\mathbf{x}\|_\alpha=1}\|A\mathbf{x}\|_\alpha$$

for some corresponding vector norm $\|\cdot\|_\alpha$. One important property of operator norms is that

$$\|I\|_\alpha = 1, \quad I \text{ identity matrix .}$$

Since $\|I\|_F = \sqrt{n}$, the Frobenius is not an operator norm for $n > 1$.

We have not mentioned the matrix two-norm yet. Well, it is

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1}\|A\mathbf{x}\|_2. \tag{1}$$

That's it! In MATLAB is is **norm**(A) or **norm**(A,2). There is no closed form expression for either $\|A\|_2$ or its "magic" vector $\mathbf{x}_2^*$ such that

$$\|A\mathbf{x}_2^*\|_2 = \|A\|_2, \quad \|\mathbf{x}_2^*\|_2 = 1.$$

There is an important mathematical equivalence, however.

Take

$$\begin{aligned} \|A\|_2^2 &= \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2^2 \\ &= \max_{\|\mathbf{x}\|_2=1} (A\mathbf{x})^T (A\mathbf{x}) \\ &= \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^T A^T A\mathbf{x}. \end{aligned}$$

Since $M = A^T A$ is a symmetric matrix ($M = M^T$), by a well known theorem (Fisher (1905)),

$$\begin{aligned} \lambda_{max}(A^T A) &= \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^T A^T A\mathbf{x} \\ &= \|A\|_2^2. \end{aligned}$$

Here $\lambda_{max}(M) = \lambda_{max}(A^T A)$ is largest eigenvalue of M and $\mathbf{x}_2^*$ is the associated eigenvector. Since there are good techniques to compute the largest eigenvalue of a matrix, we a way to compute $\|A\|_2 = \sqrt{\lambda_{max}(A^T A)}$.

Other properties of the two norm include

$$\|A\|_2 = \|A^T\|_2.$$

It satisfies the following bounds with the other norms.

$$\begin{aligned} \frac{1}{(mn)^{1/4}}\sqrt{\|A\|_1\|A\|_\infty} &\le \|A\|_2 &\le \sqrt{\|A\|_1\|A\|_\infty} \\ \frac{1}{\min\{\sqrt{m},\sqrt{n}\}}\|A\|_F &\le \|A\|_2 &\le \|A\|_F \end{aligned}$$

Why do we need matrix norms? It turns out that operator norms are particularly useful. For those we have

$$\|A\mathbf{x}\| \le \|A\|\|\mathbf{x}\|.$$

This allows us to "quantify" the effect of a matrix.

In fact, $A \in \mathrm{R}^{m\times n}$,$B \in \mathrm{R}^{n\times p}$ implies

$$\|AB\| \le \|A\|\|B\|. \tag{2}$$

The property (2) is called *consistency*. Some textbooks list (2) as the fourth axiom for a matrix norm. The reason we do not is semantic, as this is actually a relationship among three different norms (on the spaces $\mathrm{R}^{m\times n}$, $\mathrm{R}^{n\times p}$ and $\mathrm{R}^{m\times p}$).

The following is an example of a norm that is not consistent.

Example 1 *Consider the norm* $\|\cdot\|_\beta$ *on* $\mathrm{R}^{m\times n}$ *given by*

$$\|A\|_\beta = \max_{(i,j)} |a_{ij}|.$$

This is simply the ∞*-norm applied to* A *written out as vector in* R^{mn}*. For* $m = n = 2$*, consider*

$$A = B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Note that

$$AB = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

and thus $\|AB\|_\beta = 2 > \|A\|_\beta \|B\|_\beta = 1$*. Clearly,* $\|\cdot\|_\beta$ *norm is not consistent.*

For almost every application of a matrix norm that I know of, it is wise to stay with consistent norms, and we shall always do so. All of the norms we have discussed (including the Frobenius norm) are consistent.

The Frobenius norm also satisfies a consistency property with the two-norm given below.

$$\begin{aligned} \|AB\|_F &\leq \|A\|_2 \|B\|_F \\ \|AB\|_F &\leq \|A\|_F \|B\|_2 \end{aligned}$$

Now we know what norms are, let's make some use of them.

Back to

$$A\mathbf{x} = \mathbf{b}$$

where A^{-1} exists. The solution from Gaussian elimination with partial pivoting satisfies

$$(A + E)\hat{\mathbf{x}} = \mathbf{b} + \mathbf{r}.$$

Except for pathological cases (discussed in the previous lecture), for any operator norm, we may assume that

$$\begin{aligned} \|E\| &\leq \varepsilon \|A\|, \\ \|\mathbf{r}\| &\leq \varepsilon \|\mathbf{b}\| \leq \|A\| \|\mathbf{x}\|. \end{aligned}$$

Here $\varepsilon = f(n)\varepsilon_M$ where $f(n)$ is "modest" function (e.g. a polynomial). Later we will use the one-norm, but any operator norm works in the following derivation.

First, I will simplify to the case where $E = 0$. Thus we have

$$\begin{aligned} A\hat{\mathbf{x}} &= \mathbf{b} + \mathbf{r}, \\ A\mathbf{x} &= \mathbf{b}. \end{aligned}$$

Subtracting, we get

$$A(\hat{\mathbf{x}} - \mathbf{x}) = \mathbf{r}.$$

Using the assumption that A^{-1} exists

$$\hat{\mathbf{x}} - \mathbf{x} = A^{-1}\mathbf{r}.$$

So

$$\begin{aligned} \|\hat{\mathbf{x}} - \mathbf{x}\| &= \|A^{-1}\mathbf{r}\| \\ &\leq \|A^{-1}\|\|\mathbf{r}\| \\ &\leq \varepsilon\|A^{-1}\|\|A\|\|\mathbf{x}\|. \end{aligned}$$

Therefore,

$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|}{\|\mathbf{x}\|} \leq \varepsilon\|A^{-1}\|\|A\| = \varepsilon\kappa(A)$$

where

$$\kappa(A) \stackrel{def}{=} \|A^{-1}\|\|A\|$$

is called the *condition number* of A. It is a bound on how sensitive a linear system is to perturbation. Often, it is a very good bound.

Example 2

$$A = \begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{pmatrix}$$

This is the Hilbert matrix of size 4. The Hilbert matrix of dimension n is generated by the MATLAB command ***hilb****(n).*

Its inverse is generated by the command ***invhilb****(n) (there are formulas for it).*

For this case,

$$A^{-1} = \begin{pmatrix} 16 & -120 & 240 & -140 \\ -120 & 1200 & -2700 & 1680 \\ 240 & -2700 & 6480 & -4200 \\ -400 & 1680 & -4200 & 2800 \end{pmatrix}.$$

We have $\|A\|_1 = 1 + 1/2 + 1/3 + 1/4 \approx 2.0833$ *(absolute sum of the first column), and* $\|A^{-1}\|_1 = 13620$ *(absolute sum of the third column).*

Thus

$$\kappa_1(A) = \|A^{-1}\|_1 \|A\|_1 = 2.8374 \times 10^4.$$

If I take

$$\mathbf{b} = A\mathbf{x}_0$$

where

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

and then solve

$$A\mathbf{x} = \mathbf{b}$$

by Gaussian elimination with partial pivoting, then

$$\frac{\|\mathbf{x} - \mathbf{x}_0\|_1}{\|\mathbf{x_0}\|_1} = 5.5750 \times 10^{-13}.$$

Now we have that

$$\varepsilon_M \cdot \kappa_1(A) = 6.3005 \times 10^{-12},$$

so the condition number gives us a good idea of how well we can expect to solve this system.

The condition numbers of the Hilbert matrices increase exponentially with the dimension. The Hilbert matrix of dimension 12 is singular in IEEE double precision.

Back to

$$\begin{aligned} (A+E)\hat{\mathbf{x}} &= \mathbf{b} + \mathbf{r} && (3) \\ \|E\| &\leq \varepsilon\|A\|, && (4) \\ \|\mathbf{r}\| &\leq \varepsilon\|\mathbf{b}\| \leq \|A\|\|\mathbf{x}\| && (5) \end{aligned}$$

Essentially, the bound on $\|\mathbf{x} - \hat{\mathbf{x}}\|$ is given in the following theorem (its proof requires another lecture).

Theorem 1 *Assume that* $\hat{\mathbf{x}}$ *satisfies* (3)–(5) *and and that* A^{-1} *exists. If* $\eta = \|A^{-1}E\| < 1$, *then* $(A+E)^{-1}$ *exists and*

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\|\mathbf{x}\|} \le 2\varepsilon \frac{\kappa(A)}{1-\eta}. \tag{6}$$

In practice, η should be much smaller than 1. Note that

$$\eta = \|A^{-1}E\| \le \|A^{-1}\|\|E\| \le \varepsilon\kappa(A).$$

Since $\varepsilon \approx \varepsilon_M$, if $\kappa(A)\varepsilon_M \ll 1$, the matrix A is considered nonsingular to machine precision. If $\kappa(A)\varepsilon_M \approx O(1)$, it is considered singular to machine precision.

MATLAB uses an estimate of $\kappa_1(A)$. If you attempt

$$x = A\backslash b.$$

MATLAB sends out a warning if $\kappa_1(A)\varepsilon_M > 1$. (For instance, take A to be the Hilbert matrix of size 15.)

In some cases, the bound (6) is very pessimistic. For instance, if A is a diagonal matrix, say

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10^{10} & 0 \\ 0 & 0 & 10^{-20} \end{pmatrix}.$$

Even though the condition number of A in the 1,2 and ∞ norms is 10^{30}, you should expect to get very accurate results when solving linear systems with it.