CSE/Math 455
Lecture # 13B

Condition Estimation in the One Norm

Suppose that we want to compute the condition number

$$\kappa_1(A) = \|A^{-1}\|_1 \|A\|_1.$$

Assume that we have

$$A = PLU \tag{1}$$

from Gaussian elimination with partial pivoting, say. *We do not wish to spend the extra $O(n^3)$ operations to compute the inverse!* For simplicity, assume that we have the results of the MATLAB command

```
[L,U,p]=lu(A,'vector').
```

Computing $\|A\|_1$ is simply

$$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}| = \max_{1 \le j \le n} \|A\mathbf{e}_j\|_1.$$

However, we cannot compute

$$\|A^{-1}\|_1 = \max_{1 \le j \le n} \|A^{-1}\mathbf{e}_j\|_1 \tag{2}$$

without computing A^{-1}. Instead, we are going to use a heuristic due to Hager(1984) (then at the PSU Math department, now at U. Florida), and N. Higham (1986).

We need a few ideas to set up this algorithm. First, a new way to view the vector one-norm and infinity-norm. For $\mathbf{x} \in \mathrm{R}^n$, the one-norm can be written

$$\|\mathbf{x}\|_1 = \sum_{j=1}^{n} |x_j| = \mathbf{z}^T \mathbf{x}$$

where $\mathbf{z} = \mathrm{sign}(\mathbf{x})$ (i.e., $z_j = \mathrm{sign}(x_j)$). The infinity-norm of $\mathbf{x}$ may be written

$$\|\mathbf{x}\|_\infty = \max_{1 \le j \le n} |x_j| = |\mathbf{e}_{j_{max}}^T \mathbf{x}|$$

where $\mathbf{e}_{j_{max}}$ is the column of the identity matrix corresponding to the largest absolute component of $\mathbf{x}$.

Second, we need two of the Hölder inequalities

$$
\begin{aligned}
|\mathbf{x}^T\mathbf{y}| &\leq \|\mathbf{x}\|_1\|\mathbf{y}\|_\infty, \\
|\mathbf{x}^T\mathbf{y}| &\leq \|\mathbf{x}\|_\infty\|\mathbf{y}\|_1.
\end{aligned}
$$

We use them over and over again!!

Third, we need a way to compute

$$\mathbf{x} = A^{-1}\mathbf{b} \tag{3}$$

and

$$\mathbf{w} = A^{-T}\mathbf{c} \tag{4}$$

from the PLU decomposition of A. We have already shown how to computing $\mathbf{x}$ in (3) from (1). We use the two steps

$$
\begin{aligned}
L\mathbf{y} &= P^T\mathbf{b}, && \text{Forward substitution} \\
U\mathbf{x} &= \mathbf{y} && \text{Back substitution}
\end{aligned}
$$

In MATLAB, this is

```
y=L\b(p)
x=U\y
```


To compute (4), we note the (1) implies

$$A^T = U^T L^T P^T$$

which is simply LU decomposition of A^T with maximal row pivoting (which works just as well as partial pivoting). We compute (4) by the sequence

$$
\begin{aligned}
U^T\mathbf{y} &= \mathbf{z} \\
L^T\mathbf{v} &= \mathbf{y} \\
\mathbf{w} &= P\mathbf{v}
\end{aligned}
$$

In MATLAB, these steps are

```
y=U'\z
w(p)=L'\y
```

[Note: This works if $\mathbf{w}$ has already been allocated, otherwise MATLAB may give you a row vector.]

Finally, we note that

$$\|A^{-1}\|_1 = \|A^{-T}\|_\infty$$

and if j_{max} is the index of the largest absolute column in A^{-1}, then

$$\|A^{-1}\|_1 = \|A^{-1}\mathbf{e}_{j_{max}}\|_1 = \mathbf{z}^T A^{-1}\mathbf{e}_{j_{max}}$$

where $\mathbf{z} = \mathrm{sign}(A^{-1}\mathbf{e}_{j_{max}})$. Thus $\mathbf{z}$ is also the sign vector for the largest absolute row of A^{-T} so that $\mathbf{z}$ is a vector such that

$$\|A^{-T}\|_\infty = \|A^{-T}\mathbf{z}\|_\infty. \quad \|\mathbf{z}\|_\infty = 1.$$

Thus $\mathbf{e}_{j_{max}}$ and $\mathbf{z}$ are yoked together in this way. Now to the algorithm.

Our algorithm uses the current guess for $\mathbf{z}$ to find the next guess for j_{max} and that next guess for j_{max} to find the next guess for $\mathbf{z}$.

Higham's initial guess is

$$\begin{aligned} \mathbf{f} &= (f_1, \ldots, f_n)^T, \quad f_k = (-1)^k(1 + (k-1)/n) \\ \mathbf{z}_0 &= \mathbf{f}/\|\mathbf{f}\|_\infty \end{aligned}$$

This chosen to be near a vector of the form

$$\mathbf{z} = \begin{pmatrix} \pm 1 \\ \pm 1 \\ \vdots \\ \pm 1 \end{pmatrix},$$

but not exactly equal to any one of them. Thus the first iteration will always generate some improvement! Another possibility is the pair of MATLAB statements

$$\begin{aligned} f &= \mathrm{randn}(n, 1); \\ z0 &= f/\mathrm{norm}(f, Inf); \end{aligned}$$

Using the PLU decomposition, we then compute

$$\mathbf{w}_0 = A^{-T}\mathbf{z}_0$$

We let j_0 be an index such that

$$\begin{aligned}
\|\mathbf{w}_0\|_\infty &= |\mathbf{e}_{j_0}^T \mathbf{w}_0| \\
&= |\mathbf{e}_{j_0}^T A^{-T} \mathbf{z}_0| \\
&= \|A^{-T} \mathbf{z}_0\|_\infty
\end{aligned}$$

By transposing the expression $\mathbf{e}_{j_0}^T A^{-T} \mathbf{z}_0$, we note that

$$\begin{aligned}
\|A^{-T} \mathbf{z}_0\|_\infty &= |\mathbf{e}_{j_0}^T A^{-T} \mathbf{z}_0| \\
&= |\mathbf{z}_0^T A^{-1} \mathbf{e}_{j_0}|
\end{aligned}$$

We then use a Hölder inequality to note that

$$\begin{aligned}
\|A^{-T} \mathbf{z}_0\|_\infty &= |\mathbf{z}_0^T A^{-1} \mathbf{e}_{j_0}| \\
&\leq \|\mathbf{z}_0\|_\infty \|A^{-1} \mathbf{e}_{j_0}\|_1 \\
&= \|A^{-1} \mathbf{e}_{j_0}\|_1
\end{aligned}$$

Thus,

$$\|A^{-T} \mathbf{z}_0\|_\infty \leq \|A^{-1} \mathbf{e}_{j_0}\|_1.$$

The latter is our first estimate of $\|A^{-1}\|_1$.

So we use our PLU decomposition, we compute $\mathbf{x}_1 = A^{-1} \mathbf{e}_{j_0}$ from the steps

We have that

$$\begin{aligned}
\|\mathbf{x}_1\|_1 &= \|A^{-1} \mathbf{e}_{j_0}\|_1 \\
&= \mathbf{z}_1^T \mathbf{y}_1 \\
&= \mathbf{z}_1^T A^{-1} \mathbf{e}_{j_0}
\end{aligned}$$

where

$$\mathbf{z}_1 = \operatorname{sign}(\mathbf{x}_1) = \operatorname{sign}(A^{-1} \mathbf{e}_{j_0}).$$

Again using a Hölder inequality

$$\begin{aligned}
\|A^{-1} \mathbf{e}_{j_0}\|_1 &= \mathbf{z}_1^T A^{-1} \mathbf{e}_{j_0} \\
&= \mathbf{e}_{j_0}^T A^{-T} \mathbf{z}_1 \\
&\leq \|\mathbf{e}_{j_0}\|_1 \|A^{-T} \mathbf{z}_1\|_\infty \\
&= \|A^{-T} \mathbf{z}_1\|_\infty
\end{aligned}$$

Thus,

$$\|A^{-T}\mathbf{z}_0\|_\infty \leq \|A^{-1}\mathbf{e}_{j_0}\|_1 \leq \|A^{-T}\mathbf{z}_1\|_\infty.$$

Next we compute

$$\mathbf{w}_1 = A^{-T}\mathbf{z}_1$$

and let j_1 be the index such that

$$\begin{aligned}\|\mathbf{w}_1\|_\infty &= |\mathbf{e}_{j_1}^T\mathbf{w}_1| \\ &= |\mathbf{e}_{j_1}^T A^{-T}\mathbf{z}_1| \\ &= \|A^{-T}\mathbf{z}_1\|_\infty\end{aligned}$$

Once again, we note that

$$\begin{aligned}\|A^{-T}\mathbf{z}_1\|_\infty &= |\mathbf{e}_{j_1}^T A^{-T}\mathbf{z}_1| \\ &= |\mathbf{z}_1^T A^{-1}\mathbf{e}_{j_1}| \\ &\leq \|\mathbf{z}_1\|_\infty \|A^{-1}\mathbf{e}_{j_1}\|_1 \\ &= \|A^{-1}\mathbf{e}_{j_1}\|_1\end{aligned}$$

Thus,

$$\|A^{-1}\mathbf{e}_{j_0}\|_1 \leq \|A^{-T}\mathbf{z}_1\|_\infty \leq \|A^{-1}\mathbf{e}_{j_1}\|_1.$$

Thus this estimate is increasing. Using the same iteration, we can go from j_1 to j_2 and from j_k to j_{k+1}, we have a sequence of indices $j_0, j_1, \ldots, j_k, j_{k+1}$ such that

$$\|A^{-1}\mathbf{e}_{j_k}\|_1 \leq \|A^{-1}\mathbf{e}_{j_{k+1}}\|_1.$$

A step in the interation is as follows;

1. Solve $A^T\mathbf{w}_k = \mathbf{z}_k$ using the PLU decomposition
2. Find j_k such that
$$\|\mathbf{w}_k\|_\infty = |\mathbf{e}_{j_k}^T\mathbf{w}_k|$$
3. Solve $A\mathbf{x}_k = \mathbf{e}_{j_k}$ using the PLU decomposition
4. Let
$$\mathbf{z}_k = \mathbf{sign}(\mathbf{x}_k)$$
5. The value $normAinv\_est = \mathbf{z}_{k+1}^T\mathbf{y}_k = \|\mathbf{y}_k\|_1$ is the current estimate of $\|A^{-1}\|_1$.

This iteration also enforces

$$\|A^{-T}\mathbf{z}_k\|_\infty \le \|A^{-1}\mathbf{e}_{j_k}\|_1 \le \|A^{-T}\mathbf{z}_{k+1}\|_\infty.$$

We stop the iteration when one of the following occurs:

1. $j_k = j_{k+1}$.
2. $\|A^{-1}\mathbf{e}_{j_k}\|_1 = \|A^{-1}\mathbf{e}_{j_{k+1}}\|_1$
3. We have done a maximum number of iterations. In practice, about 3 or 4.

We then accept $\|A^{-1}\|_1 = \|A^{-1}\mathbf{e}_{j_k}\|_1 = \|\mathbf{y}_k\|_1$ and with $\mathbf{e}_{j_k}$ and $\mathbf{z}_{k+1}$ as the "magic vector" estimates for A^{-1} and A^{-T} in the one and ∞ norms.

In practice, this iteration finds the maximum column of A^{-1} fairly quickly. It is a heuristic, it can fail on rare occasions. There are two types of "failure," both exceedingly rare.

- It can settle on a column $\mathbf{e}_j$ that is not a maximum column.
- A few iterations may not be enough. There are rare examples when the algorithm searches all n indices to find the right column.

The following 3×3 example shows how the method works.

Example 1 *Let*

$$A = \begin{pmatrix} -1 & -99 & 270 \\ -1. & -101 & 330.5 \\ 1 & 100 & -300 \end{pmatrix}$$

Its PLU decomposition is

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1.0000 & 0 \\ -1 & -0.5 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} -1 & -99 & 270 \\ 0 & -2 & 60.5 \\ 0 & 0 & 0.2500 \end{pmatrix}$$

with $\mathbf{p} = (1, 2, 3)^T$, *i.e., no pivoting is necessary.*

Suppose our initial vector $\mathbf{z}_0$ *is*

$$\mathbf{z}_0 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

Solving

$$U^T\mathbf{y}_0 = \mathbf{z}_0$$

yields

$$\mathbf{y}_0 = \begin{pmatrix} -1 \\ 49 \\ -10782 \end{pmatrix}$$

Then solving

$$L^T\mathbf{w}_0 = \mathbf{y}_0$$

yields

$$\mathbf{w}_0 = \begin{pmatrix} -5441 \\ -5342 \\ -10782 \end{pmatrix}$$

The value of j_0 *is* 3, *since the third component of* $\mathbf{w}_0$ *is the maximum and thus,*

$$\|\mathbf{w}_0\|_\infty = 10782$$

is our first estimate of $\|A^{-1}\|_1 = \|A^{-T}\|_\infty$. *Thus*

$$\mathbf{e}_{j_0} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

If we solve

$$L\mathbf{v} = P^T\mathbf{e}_{j_0}$$

we get

$$\mathbf{v} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

If we then solve

$$U\mathbf{u}_0 = \mathbf{v}$$

we get

$$\mathbf{u}_0 = \begin{pmatrix} -10899 \\ 121 \\ 4 \end{pmatrix}$$

The vector

$$\mathbf{z}_1 = \mathbf{sign}(\mathbf{u}_0) = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}.$$

And we have that

$$\|\mathbf{u}_0\|_1 = 11024$$

which is new and larger estimate of $\|A^{-1}\|_1$. *We then solve*

$$U^T \mathbf{y} = \mathbf{z}_1$$

to get

$$\mathbf{y} = \begin{pmatrix} 1 \\ -50 \\ 11024 \end{pmatrix}$$

We then solve

$$L^T \mathbf{w}_1 = \mathbf{y}$$

to get

$$\mathbf{w}_1 = \begin{pmatrix} 5563 \\ 5462 \\ 11024 \end{pmatrix}$$

Since j_1, *the largest component of* $\mathbf{w}_1$ *is* 3, *and is the same as* j_0, *we are done. Moreover,*

$$\|\mathbf{w}_1\|_\infty = \|\mathbf{u}_0\|_1 = \|A^{-1}\mathbf{e}_{j_0}\|_1 = 11024.$$

so our estimate of $\|A^{-1}\|_1$ *is*

$$\|A^{-1}\|_1 = 11024.$$

Our estimate also says that $J = j_0 = j_1 = 3$ *is the maximum column of* A^{-1}. *Thus*

$$\|A^{-1}\|_1 = \|A^{-1}\mathbf{e}_3\|_1$$

with $\|\mathbf{e}_3\|_1 = 1$ *makes* $\mathbf{e}_3$ *the "magic vector" for* A^{-1}. *We also have that*

$$\mathbf{z}_1 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

is our estimate of a vector such that $\|\mathbf{z}_1\|_\infty = 1$ *and*

$$\|A^{-1}\|_1 = \|A^{-T}\|_\infty = \|A^{-T}\mathbf{z}_1\|_\infty.$$

Since

$$\|A\|_1 = 900.5$$

then our estimate for the condition number is

$$\kappa_1(A) = \|A^{-1}\|_1 \|A\|_1 = 9.927112 \cdot 10^6 = 9927112.$$

The idea behind our estimation algorithm is to avoid computing the inverse, but the inverse of A *is*

$$A^{-1} = \begin{pmatrix} -5500 & -5400 & -10899 \\ 61 & 60 & 121 \\ 2 & 2 & 4 \end{pmatrix}$$

By inspection, you can see that its maximum column is the third one and that

$$\|A^{-1}\|_1 = 11024$$

and that

$$\mathbf{z}_1 = \mathbf{sign}(A^{-1}\mathbf{e}_3).$$