Online Lecture Notes

Prof. Boris Houska

April 19, 2022

1 Linear Equation Systems

Goal of this lecture is to analyze equations of the form

$$Ax = b$$

with $A \in A^{n\times n}$ and $b \in \mathbb{R}^n$.

1.1 Condition Numbers

In order to analyze the conditioning of a linear equation system, it is helpful to introduce induced matrix (operator matrix norms) of the form

$$\|A\| = \sup_{x\in\mathbb{R}^n} \frac{\|Ax\|}{\|x\|}$$

for any given vector norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$. Properties of matrix norms

- Absolute Homogeniety: We have $\|\alpha A\| \leq |\alpha|\|A\|$ for $\alpha \in \mathbb{R}$.
- Triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$.
- Positive Definiteness: We have $\|A\| \geq 0$ and $\|A\| = 0$ if and only if $A = 0$.
- Submultiplicativity: $\|AB\| \leq \|A\|\|B\|$. Proof: we may assume $Bx \neq 0$:

$$\|AB\| = \sup_{x\in\mathbb{R}^n} \frac{\|ABx\|}{\|x\|} = \sup_{x\in\mathbb{R}^n} \frac{\|ABx\|}{\|Bx\|}\frac{\|Bx\|}{\|x\|} \leq \|A\|\|B\| \ .$$

- Natural Scaling: we have $\|I\| = 1$.

Our first technical statement that we want to prove is that if $\|A\| < 1$, then $I + A$ is invertible and we have

$$\|(I + A)^{-1}\| \leq \frac{1}{1 - \|A\|} \ .$$

Proof: We start with the triangle inequality for norms:

$$\|-Ax\| + \|(I + A)x\| \geq \|x\| \quad \Longleftrightarrow \quad \|(I + A)x\| \geq \|x\| - \|Ax\| \qquad (1)$$

$$\overset{\text{Submultiplicativity}}{\Longrightarrow} \quad \|(I + A)x\| \geq (I - \|A\|)\|x\| \qquad (2)$$

This means that $(I + A)x \neq 0$ if $x \neq 0$. This is the same as saying that $I + A$ is invertible. Additionally, we have

$$1 = \|(I + A)^{-1}(I + A)\| \geq \|(I + A)^{-1}\|(1 - \|A),$$

where the last step follows again from the triangle inequality.

1.2 Conditioning of linear equation systems

Let A be an invertible, $Ax = b$, let $\|\delta A\| \leq \frac{1}{\|A^{-1}\|}$ and

$$(A + \delta A)(x + \delta x) = b + \delta b\ .$$

Then we have

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\text{cond}(A)}{1 - \text{cond}(A)\frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|} \right)$$

with $\text{cond}(A) = \|A\|\|A^{-1}\|$.

Proof. We start by eliminating δx from the above equation

$$(A + \delta A)\delta x = b + \delta b - (A + \delta A)x = \underbrace{(Ax + b)}_{=0} + \delta b - \delta A x$$

This means that we have

$$\delta x = (A + \delta A)^{-1} \left[\delta b - \delta A x\right]$$

This yields

$$\begin{aligned}
\|\delta x\| &= \|(A + \delta A)^{-1} \left[\delta b - \delta A x\right] \| \\
&\leq \|(A + \delta A)^{-1}\| \|\delta b - \delta A x\| \\
&\leq \|(A + \delta A)^{-1}\| \left[\|\delta b\| + \|\delta A\|\|x\|\right] \\
&= \|A^{-1}(I + A^{-1}\delta A)^{-1}\| \left[\|\delta b\| + \|\delta A\|\|x\|\right] \\
&\leq \|(I + A^{-1}\delta A)^{-1}\|\|A^{-1}\| \left[\|\delta b\| + \|\delta A\|\|x\|\right]
\end{aligned}$$

In the next step we use the lemma from the previous subsection, which yields

$$\|(I + A^{-1}\delta A)^{-1}\| \leq \frac{1}{1 - \|A^{-1}\|\|\delta A\|}$$

If we substitute this, we get

$$\|\delta x\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|\delta A\|} \left[\|\delta b\| + \|\delta A\|\|x\|\right]$$

Next,

$$\|\delta x\| \leq \frac{\|A^{-1}\|\|A\|\|x\|}{1 - \|A^{-1}\|\|\delta A\|} \left[\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|}\right] \tag{3}$$

Devide this equation by $\|x\|$ to find

$$\begin{aligned}
\frac{\|\delta x\|}{\|x\|} &\leq \frac{\|A^{-1}\|\|A\|}{1 - \|A^{-1}\|\|\delta A\|} \left[\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|}\right] \\
&= \frac{\|A^{-1}\|\|A\|}{1 - \|A^{-1}\|\|A\|\frac{\|\delta A\|}{\|A\|}} \left[\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|}\right] \\
&= \frac{\text{cond(A)}}{1 - \text{cond}(A)\frac{\|\delta A\|}{\|A\|}} \left[\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|}\right]\ . \qquad (4)
\end{aligned}$$

This completes our proof.

2 Gauss Elimination

The goal of this section is to develop a numerical algorithm for solving the linear equations system $Ax = b$.

2.1 Triangular Equation Systems

Let us first have a closer look at the upper-triangular case

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n-1} & a_{1,n} \\ 0 & a_{2,2} & \dots & a_{2,n-1} & a_{2,n} \\ 0 & 0 & \ddots & & \\ 0 & 0 & \dots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & \dots & 0 & a_{n,n} \end{pmatrix}$$

This means that the last equation has the form

$$a_{n,n}x_n = b_n \qquad \Longrightarrow \qquad \frac{b_n}{a_{n,n}}$$

The corresponding second last equation has the form

$$a_{n-1,n-1}x_{n-1}+a_{n-1,n}x_n = b_{n-1} \qquad \Longrightarrow \qquad x_{n-1} = \frac{1}{a_{n-1,n-1}}\left[b_{n-1} - a_{n-1,n}x_n\right] \ .$$

Similarly, we can solve all the other equations by a backward recursion,

$$\forall j \in \{n, n-1, \dots, 1\}, \qquad x_j = \frac{1}{a_{j,j}}\left[b_j - \sum_{k=j+1}^{n} a_{k,j}x_k\right] \ .$$

This means that in term of the computational complexity, we need

- n division operations (assuming $a_{j,j} \neq 0$)
- we need $0 + 1 + 2 + \dots (n-1) = \frac{n(n-1)}{2}$ minus operations
- we need $0 + 1 + 2 + \dots (n-1) = \frac{n(n-1)}{2}$ product operations

In total this gives $O(n^2)$ operations. Notice that this also the minumum possible computation complexity that we can expect, since it is equal to the storage complexity (recall that A has also $O(n^2)$ coefficients that need to be stored).