5374 Fall '22

Numerical Linear Algebra Lecture 5

(cf. Solomon's chapter 4, section on "Sensibility Analysis")

Condition number

Let's consider

$$Ax = b$$

for some A which is $n \times n$ and invertible.

How does the solution x vary when we vary A or b? Let us fix a $n \times n$ matrix $\dot{A}$ and a $\mathbb{R}^n$ vector $\dot{b}$, and set for all small enough ε, the matrix and vector

$$A_\varepsilon = A + \varepsilon \dot{A}, \quad b_\varepsilon = b + \varepsilon \dot{b}$$

(so, note $A_0 = A$, $b_0 = b$)

Exercise : Using Cramer's rule (namely, $\det(A) \neq 0$ $\Longleftrightarrow$ A is invertible) show there is a $\varepsilon_0 > 0$ depends on A and $\dot{A}$ such that if $|\varepsilon| < \varepsilon_0$ then A_ε is also invertible.

(this means that, when thought of as a subset of $\mathbb{R}^{n^2}$, the set of invertible matrices is an open set)

Then for those small ε we define x_ε as the unique vector solving

$$A_\varepsilon x_\varepsilon = b_\varepsilon$$

Question: What can we say about

$$\| x - x_\varepsilon \|_2$$

as $\varepsilon \to 0$?

We are going to estimate the derivative of x_ε with respect to the parameter $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$

This is indeed a differentiable funct. of ε since

$$x_\varepsilon = A_\varepsilon^{-1} b_\varepsilon$$

and $\varepsilon \mapsto A_\varepsilon$, $\varepsilon \mapsto b_\varepsilon$ are differentiable and A_0 is invertible so $\varepsilon \to A_\varepsilon^{-1}$ is differentiable (in fact, it is infinitely differentiable in ε as long as $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$)

Let us write $\dot{x} := \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} x_\varepsilon$, what can we say about $\dot{x}$?

$$\frac{d}{d\varepsilon}\left(A_\varepsilon x_\varepsilon\right) = \frac{d}{d\varepsilon}\left(b_\varepsilon\right) \qquad \forall \varepsilon$$

$$\frac{d}{d\varepsilon}\left((A+\varepsilon\dot{A})x_\varepsilon\right) = \frac{d}{d\varepsilon}\left(b+\varepsilon\dot{b}\right)$$

$$\frac{d}{d\varepsilon}\left(A x_\varepsilon + \varepsilon\dot{A}x_\varepsilon\right) = \dot{b}$$

$$A\frac{d}{d\varepsilon}x_\varepsilon + \dot{A}x_\varepsilon + \varepsilon\dot{A}\frac{d}{d\varepsilon}x_\varepsilon = \dot{b}$$

For $\varepsilon=0$, we obtain

$$A\dot{x} + \dot{A}x = \dot{b}$$

Solving for $\dot{x}$,

$$\dot{x} = A^{-1}\left(\dot{b} - \dot{A}x\right)$$

From here we can estimate the size of $x - x_\varepsilon$ since

$$x_\varepsilon = x + \varepsilon\dot{x} + O(\varepsilon^2)$$

$$\Longrightarrow \quad \|x - x_\varepsilon\| = \|\varepsilon \dot{x} + O(\varepsilon^2)\|$$
$$\leq \|\varepsilon \dot{x}\| + O(\varepsilon^2)$$
$$= |\varepsilon| \, \|\dot{x}\| + O(\varepsilon^2)$$

In scientific computing we care about two ways of estimating errors: absolute error and relative error.

$$\text{Absolute error} := \|x - x_\varepsilon\|$$

$$\text{Relative error} := \frac{\|x - x_\varepsilon\|}{\|x\|}$$

Let's estimate the relative error for our problem.

$$\frac{\|x - x_\varepsilon\|}{\|x\|} \leq \frac{|\varepsilon| \, \|\dot{x}\|}{\|x\|} + O(\varepsilon^2)$$

Recall our formula for $\dot{x}$,

$$\dot{x} = A^{-1}(\dot{b} - \dot{A}x)$$

$$\|\dot{x}\|_2 = \|A^{-1}(\dot{b} - \dot{A}x)\|_2$$
$$\leq \|A^{-1}\|_{op} \, \|\dot{b} - \dot{A}x\|_2$$

$$\Rightarrow \quad \|\dot{x}\|_2 \le \|A^{-1}\|_{op} \left(\|\dot{b}\|_2 + \|\dot{A}x\|_2 \right)$$
$$\le \|A^{-1}\|_{op} \left(\|\dot{b}\|_2 + \|\dot{A}\| \, \|x\|_2 \right)$$

Putting this estimate back in our relative error estimation,

$$\frac{\|x - x_\varepsilon\|_2}{\|x\|_2} \le |\varepsilon| \left(\frac{\|A^{-1}\|_{op} \, \|\dot{b}\|_2}{\|x\|_2} + \frac{\|A^{-1}\|_{op} \|\dot{A}\|_{op} \cancel{\|x\|_2}}{\cancel{\|x\|_2}} \right) + O(\varepsilon^2)$$

We want to estimate the LHS in terms of the relative error made when A, b are replaced with $A_\varepsilon, b_\varepsilon$, i.e. we want to estimate things by

$$\frac{\|\dot{A}\|_{op}}{\|A\|_{op}} \quad \text{and} \quad \frac{\|\dot{b}\|_2}{\|b\|_2}$$

For the second term, observe that

$$\|A^{-1}\|_{op} \|\dot{A}\|_{op} = \left(\|A\|_{op} \cdot \|A^{-1}\|_{op} \right) \cdot \frac{\|\dot{A}\|_{op}}{\|A\|_{op}}$$

Meanwhile, for the first term, we observe that

$$b = Ax$$
$$\Longrightarrow \quad \|b\|_2 \le \|A\|_{op} \cdot \|x\|_2$$

i.e.
$$\frac{1}{\|x\|_2} \le \frac{\|A\|_{op}}{\|b\|_2}$$

We conclude that

$$\frac{\|x - x_\varepsilon\|}{\|x\|} \leq |\epsilon| \left(\|A\|_{op} \|A^{-1}\|_{op} \frac{\|\mathring{b}\|_2}{\|b\|_2} + \|A\|_{op} \|A^{-1}\|_{op} \cdot \frac{\|\mathring{A}\|_{op}}{\|A\|_{op}} \right) + O(\varepsilon^2)$$

Gathering the common factor,

$$\frac{\|x - x_\varepsilon\|_2}{\|x\|_2} \leq \left(\|A\|_{op} \|A^{-1}\|_{op} \right) \left(\frac{|\epsilon| \|\mathring{b}\|_2}{\|b\|_2} + \frac{|\epsilon| \|\mathring{A}\|_{op}}{\|A\|_{op}} \right) + O(\varepsilon^2)$$

(Alternative way of writing this, using that

$$A_\varepsilon - A = \varepsilon \mathring{A}, \quad b_\varepsilon - b = \varepsilon \mathring{b}$$

$$\frac{\|x - x_\varepsilon\|_2}{\|x\|_2} \leq \left(\|A\|_{op} \|A^{-1}\|_{op} \right) \left(\frac{\|b - b_\varepsilon\|_2}{\|b\|_2} + \frac{\|A - A_\varepsilon\|_{op}}{\|A\|} \right) + O(\varepsilon^2)$$

Definition. Given a $n \times n$ matrix A we define its condition number, $\mathrm{cond}(A)$, by

$$\mathrm{cond}(A) = \begin{cases} \|A\|_{op} \|A^{-1}\|_{op} & \text{if } A \text{ is invertible} \\ +\infty & \text{if } A \text{ is not invertible.} \end{cases}$$

Remarks: * If $A = \lambda I$, $\lambda \neq 0$, then

$$\mathrm{cond}(A) = \|\lambda I\|_{op} \cdot \|\lambda^{-1} I\|_{op}$$
$$= |\lambda| \cdot |\lambda^{-1}| = 1$$

* More generally, if A is invertible and $\lambda \neq 0$, then

$$\mathrm{cond}(\lambda A) = \mathrm{cond}(A)$$

Why! Since $\|\lambda A\|_{op} = |\lambda| \, \|A\|_{op}$

$\|(\lambda A)^{-1}\|_{op} = |\lambda|^{-1} \, \|A^{-1}\|_{op}$