EECS 275 Matrix Computation

Ming-Hsuan Yang

Electrical Engineering and Computer Science
University of California at Merced
Merced, CA 95344
http://faculty.ucmerced.edu/mhyang

Lecture 16

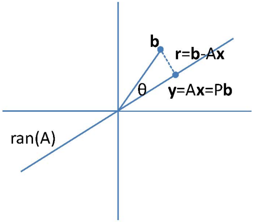
Overview

- Conditioning of least squares problems
- Perturbation
- Stability

Reading

- Chapter 18 of *Numerical Linear Algebra* by Llyod Trefethen and David Bau
- Chapter 2 of *Matrix Computations* by Gene Golub and Charles Van Loan

Conditioning of least squares problems



- Assume A is full rank and consider 2-norm for analysis

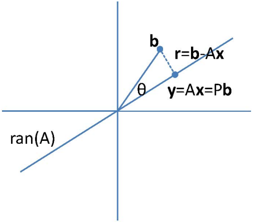
$$\text{Given} \quad A \in \mathbb{C}^{m\times n} \text{ of full rank}, m \geq n, \mathbf{b} \in \mathbb{C}^m$$
$$\text{Find} \quad \mathbf{x} \in \mathbb{C}^n, \text{ such that } \quad \|\mathbf{b} - A\mathbf{x}\| \text{ is minimized}$$

- The solution $\mathbf{x}$ and the corresponding $\mathbf{y} = A\mathbf{x}$ that is closest to $\mathbf{b}$ in ran(A) are given by

$$\mathbf{x} = A^\dagger \mathbf{b} \quad \mathbf{y} = P\mathbf{b}$$

where $A^\dagger = (A^H A)^{-1} A^H \in C^{n\times m}$ is the pseudoinverse of A and $P = AA^\dagger \in \mathbb{C}^{m\times m}$ is the orthogonal projector onto ran(A)

Conditioning of least squares problems (cont'd)



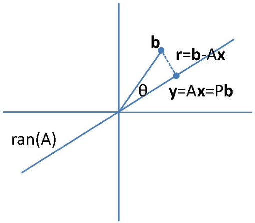
- Recall for rectangular matrix A,

$$\kappa(A) = \|A\|\|A^\dagger\| = \frac{\sigma_1}{\sigma_n}$$

- Another measure of closeness of the fit

$$\theta = \cos^{-1}\frac{\|\mathbf{y}\|}{\|\mathbf{b}\|}$$

Conditioning of least squares problems (cont'd)



- The third is a measure of how much $\|\mathbf{y}\|$ falls short of its maximum possible value, given $\|A\|$ and $\|\mathbf{x}\|$

$$\eta = \frac{\|A\|\|\mathbf{x}\|}{\|\mathbf{y}\|} = \frac{\|A\|\|\mathbf{x}\|}{\|A\mathbf{x}\|}$$

- These parameters lie in the ranges

$$1 \leq \kappa(A) < \infty, \quad 0 \leq \theta \leq \pi/2, \quad 1 \leq \eta \leq \kappa(A)$$

Conditioning of least squares problems (cont'd)

Theorem

Let $\mathbf{b} \in C^m$ and $A \in \mathbb{C}^{m \times n}$ be full rank. The least squares has the following 2-norm relative condition numbers describing the sensitivities of $\mathbf{y}$ and $\mathbf{x}$ to perturbations in $\mathbf{b}$ and A:

	$\mathbf{y}$	$\mathbf{x}$
$\mathbf{b}$	$\frac{1}{\cos\theta}$	$\frac{\kappa(A)}{\eta\cos\theta}$
A	$\frac{\kappa(A)}{\cos\theta}$	$\kappa(A) + \frac{\kappa(A)^2\tan\theta}{\eta}$

The results in the first row are exact, being attained for certain perturbations $\delta\mathbf{b}$, and the results in the second row are upper bounds

- When $m = n$, the problem reduces to a square, nonsingular system with $\theta = 0$
- The numbers in the second column reduce to $\kappa(A)/\eta$ and $\kappa(A)$

Conditioning of least squares problems (cont'd)

- Let $A = U\Sigma V^H$ where Σ is an $m \times n$ diagonal matrix
- Since perturbations are measured in 2-norm, their sizes are unaffected by a unitary change of basis, so the perturbation behavior of A is the same as that of Σ
- Without loss of generality, we can deal with Σ directly
- In the following analysis, we assume $A = \Sigma$ and write

$$A = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & & \\ & & & \end{bmatrix} = \begin{bmatrix} A_1 \\ 0 \end{bmatrix}$$

where A_1 is $n \times n$ and diagonal and the rest of A is zero

Conditioning of least squares problems (cont'd)

- The orthogonal projection of $\mathbf{b}$ onto $\operatorname{ran}(A)$ is now

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$$

  where $\mathbf{b}_1$ contains the first n entries of $\mathbf{b}$, then the projection $\mathbf{y} = P\mathbf{b}$ is

$$\mathbf{y} = \begin{bmatrix} \mathbf{b}_1 \\ 0 \end{bmatrix}$$

- To find the corresponding $\mathbf{x}$ we can write $A\mathbf{x} = \mathbf{y}$ as

$$\begin{bmatrix} A_1 \\ 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{b}_1 \\ 0 \end{bmatrix}$$

  which implies $\mathbf{x} = A_1^{-1}\mathbf{b}_1$

- It follows that the orthogonal projector and pseudoinverse are the block 2×2 and 1×2 matrices

$$P = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \qquad A^\dagger = \begin{bmatrix} A_1^{-1} & 0 \end{bmatrix}$$

Sensitivity of $\mathbf{y}$ to perturbations in $\mathbf{b}$

- The relationship between $\mathbf{b}$ and $\mathbf{y}$ is linear $\mathbf{y} = P\mathbf{b}$
- The Jacobian of this mapping is P itself with $\|P\| = 1$
- The condition number of $\mathbf{y}$ with respect to perturbations in $\mathbf{b}$ is

$$\kappa = \frac{\|J(\mathbf{x})\|}{\|f(\mathbf{x})\|/\|\mathbf{x}\|}, \quad \kappa_{\mathbf{b}\mapsto\mathbf{y}} = \frac{\|P\|}{\|\mathbf{y}\|/\|\mathbf{b}\|} = \frac{1}{\cos\theta}$$

- Recall

$$\kappa = \sup_{\delta\mathbf{x}} \left(\frac{\|\delta f\|}{\|f(\mathbf{x})\|} \Big/ \frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \right)$$

  and $\delta f \approx J(\mathbf{x})\delta\mathbf{x}$

- The condition number is realized (i.e., the supremum is attained) for perturbations $\delta\mathbf{b}$ with $\|P(\delta\mathbf{b})\| = \|\delta\mathbf{b}\|$ which occurs when $\delta\mathbf{b}$ is zero except in the first n entries

Sensitivity of $\mathbf{x}$ to perturbations in $\mathbf{b}$

- The relationship between $\mathbf{b}$ and $\mathbf{x}$ is linear, $\mathbf{x} = A^\dagger \mathbf{b}$, with Jacobian $A^\dagger$
- The condition number of $\mathbf{x}$ with respect to perturbations in $\mathbf{b}$ is

$$\kappa_{\mathbf{b}\mapsto\mathbf{x}} = \frac{\|A^\dagger\|}{\|\mathbf{x}\|/\|\mathbf{b}\|} = \|A^\dagger\| \frac{\|\mathbf{b}\|\|\mathbf{y}\|}{\|\mathbf{y}\|\|\mathbf{x}\|} = \|A^\dagger\| \frac{1}{\cos\theta} \frac{\|A\|}{\eta} = \frac{\kappa(A)}{\eta\cos\theta}$$

- The condition number is realized by perturbations $\delta\mathbf{b}$ satisfying $\|A^\dagger(\delta\mathbf{b})\| = \|A^\dagger\|\|\delta\mathbf{b}\| = \|\delta\mathbf{b}\|/\sigma_n$, which occurs when $\delta\mathbf{b}$ is zero except in the n-th entry (or perhaps also in other entries if A has more than one singular value equal to σ_n)

Tilting the range of A

- The analysis of perturbations in A is a nonlinear problem
- Observe that the perturbations in A affect the last squares problem in two ways: they distort the mapping of $\mathbb{C}^m$ onto ran(A) and they alter ran(A) itself
- Consider the slight change in ran(A) as small tiltings of this space
- What is the maximum angle of tilt $\delta\alpha$ that can be imparted by a small perturbation of δA?
- The image under A of the unit n-sphere is a hyperellipse that lies flat in ran(A)
- To change ran(A) as efficiently as possible, we grasp a point $\mathbf{p} = A\mathbf{v}$ on the hyperellipse (hence $\|\mathbf{v}\| = 1$) and nudge it in a direction $\delta\mathbf{p}$ orthogonal to ran(A)
- A matrix perturbation that achieves this most efficiently is $\delta A = (\delta\mathbf{p})\mathbf{v}^H$, which gives $(\delta A)\mathbf{v} = \delta\mathbf{p}$ with $\|\delta A\| = \|\delta\mathbf{p}\|$

Tilting the range of A (cont'd)

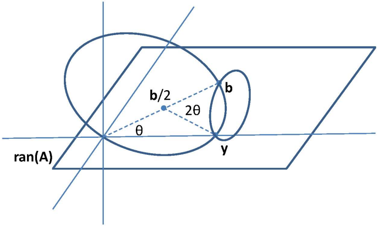
- To obtain the maximum tilt with a given $\|\delta\mathbf{p}\|$, we should take $\mathbf{p}$ to be as close to the origin as possible
- That is, $\mathbf{p} = \sigma_n \mathbf{u}_n$, where σ_n is the smallest singular value of A and $\mathbf{u}_n$ is the corresponding left singular vector
- Let $A = \begin{bmatrix} A_1 \\ \mathbf{0} \end{bmatrix}$ as before, $\mathbf{p}$ is equal to the last column of A, $\mathbf{v}^H$ is the n-vector $(0, 0, \ldots, 1)$ and δA is a perturbation of the entries of A below the diagonal in this column
- The perturbation tilts ran(A) by the angle $\delta\alpha$ given by $\tan(\delta\alpha) = \|\delta\mathbf{p}\|/\sigma_n$
- Since $\|\delta\mathbf{p}\| = \|\delta A\|$ and $\delta\alpha \leq \tan(\delta\alpha)$, we have

$$\delta\alpha \leq \frac{\|\delta A\|}{\sigma_n} = \frac{\|\delta A\|}{\|A\|}\kappa(A)$$

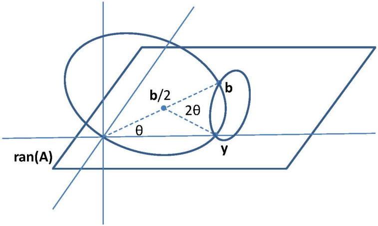
with equality attained for choices δA of the kind described above

Sensitivity of $\mathbf{y}$ to perturbations in A

- $\mathbf{y}$ is the orthogonal projection of $\mathbf{b}$ onto $\text{ran}(A)$, it is determined by $\mathbf{b}$ and $\text{ran}(A)$
- Study the effect on $\mathbf{y}$ of tilting $\text{ran}(A)$ by some angle $\delta\alpha$
- Can look at this from the geometric perspective when imaging fixing $\mathbf{b}$ and watching $\mathbf{y}$ vary as $\text{ran}(A)$ is tiled
- No matter how $\text{ran}(A)$ is tiled, the vector $\mathbf{y} \in \text{ran}(A)$ must always be orthogonal to $\mathbf{y} - \mathbf{b}$
- That is, the line $\mathbf{b} - \mathbf{y}$ must lie at right angles to the line $\mathbf{0} - \mathbf{y}$
- In other words, as $\text{ran}(A)$ is adjusted, $\mathbf{y}$ moves along the sphere of radius $\|\mathbf{b}\|/2$ centered at the point $\mathbf{b}/2$



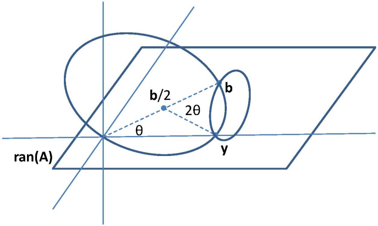
Sensitivity of $\mathbf{y}$ to perturbations in A (cont'd)



- Tilting ran(A) in the plane $\mathbf{0}-\mathbf{b}-\mathbf{y}$ by an angle $\delta\alpha$ changes the angle 2θ at the central point $\mathbf{b}/2$ by $2\delta\alpha$
- The corresponding perturbation $\delta\mathbf{y}$ is the base of an isosceles triangle with central angle $2\delta\alpha$ and edge length $\|\mathbf{b}\|/2$, thus $\|\delta\mathbf{y}\| = \|\mathbf{b}\| \sin(\delta\alpha)$
- For arbitrary perturbations by an angle $\delta\alpha$, we have

$$\|\delta\mathbf{y}\| \leq \|\mathbf{b}\| \sin(\delta\alpha) \leq \|\mathbf{b}\|\delta\alpha$$

Sensitivity of $\mathbf{y}$ to perturbations in A (cont'd)



- For arbitrary perturbations by an angle $\delta\alpha$, we have

$$\|\delta \mathbf{y}\| \leq \|\mathbf{b}\| \sin(\delta\alpha) \leq \|\mathbf{b}\|\delta\alpha$$

- Using the previous results on θ and $\delta\alpha$,

$$\begin{aligned} \delta\alpha &\leq \frac{\|\delta A\|}{\sigma_n} = \frac{\|\delta A\|}{\|A\|}\kappa(A) \\ \theta &= \cos^{-1}\frac{\|\mathbf{y}\|}{\|\mathbf{b}\|} \end{aligned}$$

- We have

$$\|\delta \mathbf{y}\| \leq \|\delta A\|\kappa(A)\|\mathbf{y}\|/(\|A\|\cos\theta)$$

and

$$\frac{\|\delta \mathbf{y}\|}{\|\mathbf{y}\|} \Big/ \frac{\|\delta A\|}{\|A\|} \leq \frac{\kappa(A)}{\cos\theta}$$

Sensitivity of $\mathbf{x}$ to perturbations in A

- A perturbation of δA can be split into two parts: one part δA_1 in the first n rows and another part δA_2 in the remaining $m-n$ rows

$$\delta A = \begin{bmatrix} \delta A_1 \\ \delta A_2 \end{bmatrix} = \begin{bmatrix} \delta A_1 \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \delta A_2 \end{bmatrix}$$

- A perturbation δA_1 changes the mapping of A in its range, but not $\operatorname{ran}(A)$ itself or $\mathbf{y}$
- It perturb A_1 by δA_1 in $\mathbf{x} = A_1^{-1}\mathbf{b}_1$ without changing $\mathbf{b}_1$, and the condition number is

$$\frac{\|\delta \mathbf{x}\|}{\|\delta \mathbf{x}\|} \Big/ \frac{\|\delta A_1\|}{\|A\|} \leq \kappa(A_1) = \kappa(A)$$

- A perturbation δA_2 tilts $\operatorname{ran}(A)$ without changing the mapping of A within this space
- This corresponds to perturbing $\mathbf{b}_1$ in $\mathbf{x} = A_1^{-1}\mathbf{b}_1$ without changing A_1, and the condition number is

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \Big/ \frac{\|\delta \mathbf{b}_1\|}{\|\mathbf{b}_1\|} \leq \frac{\kappa(A_1)}{\eta(A_1; \mathbf{x})} = \frac{\kappa(A)}{\eta}$$

Sensitivity of $\mathbf{x}$ to perturbations in A (cont'd)

- Need to relate $\delta \mathbf{b}_1$ and δA_2
- The vector $\mathbf{b}_1$ is $\mathbf{y}$ expressed in the coordinates of $\text{ran}(A)$
- Thus, the only changes in $\mathbf{y}$ that are realized as changes in $\mathbf{b}_1$ are those that lie parallel to $\text{ran}(A)$; orthogonal changes have no effect
- If $\text{ran}(A)$ is tilted by an angle $\delta\alpha$ in the plane $\mathbf{0}-\mathbf{b}-\mathbf{y}$, the resulting perturbation $\delta \mathbf{y}$ lies not parallel to $\text{ran}(A)$ but at an angle $\pi/2 - \theta$
- Thus, the changes in $\mathbf{b}_1$ satisfies $\|\delta \mathbf{b}_1\| = \sin\theta \|\delta \mathbf{y}\|$, and

$$\|\delta \mathbf{b}_1\| \leq (\|\mathbf{b}\| \delta\alpha) \sin\theta$$

- Since $\|\mathbf{b}_1\| = \|\mathbf{b}\| \cos\theta$, we have

$$\frac{\|\delta \mathbf{b}_1\|}{\|\mathbf{b}_1\|} \leq (\delta\alpha) \tan\theta$$

thus

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|\delta \mathbf{b}_1\|}{\|\mathbf{b}_1\|} \frac{\kappa(A)}{\eta} \leq \frac{\kappa(A)}{\eta} (\delta\alpha) \tan\theta$$

Sensitivity of $\mathbf{x}$ to perturbations in A (cont'd)

- Relate A_2 to early results,

$$\delta\alpha \leq \frac{\|\delta A_2\|}{\sigma_n} = \frac{\|\delta A_2\|}{\|A\|}\kappa(A)$$

- Put things together,

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \Big/ \frac{\|\delta A_2\|}{\|A\|} \leq \frac{\kappa(A)^2 \tan\theta}{\eta}$$

- Combine the perturbations caused by A_1 and A_2

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \Big/ \frac{\|\delta A\|}{\|A\|} \leq \kappa(A) + \frac{\kappa(A)^2 \tan\theta}{\eta}$$

Floating point and stability I

- Machine precision

$$\varepsilon_{\text{machine}} = \frac{1}{2}\beta^{1-t}$$

  where β is usually 2 and t is 24 and 53 for IEEE single and double precision
- A mathematical problem is a function $f : X \to Y$
- An algorithm is another map $\tilde{f} : X \to Y$ (e.g., an implementation on computer)
- An algorithm $\tilde{f}$ for a problem f is accurate if for each $x \in X$

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = O(\varepsilon_{\text{machine}})$$

- $O(\varepsilon)$ means "on the order of machine epsilon"

Floating point and stability II

- An algorithm $\tilde{f}$ for a problem f is stable if for each $x \in X$

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = O(\varepsilon_{\text{machine}})$$

for each $\tilde{x}$ with

$$\frac{\|\tilde{x} - x\|}{\|x\|} = O(\varepsilon_{\text{machine}})$$

- In other words, a stable algorithm gives nearly the right answer to nearly the right question
- For a nonsingular $m \times m$ system of equations $A\mathbf{x} = \mathbf{b}$, we have

$$\frac{\|\tilde{x} - x\|}{\|x\|} = O(\kappa(A)\varepsilon_{\text{machine}})$$