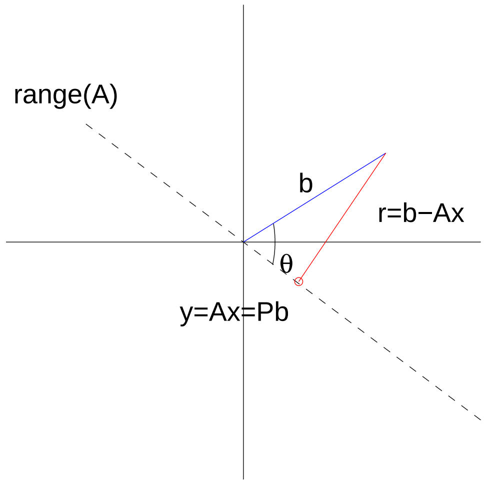
Lecture 18 - *Conditioning of Least-Squares Problems*

OBJECTIVE:

The conditioning of least-squares problems is subtle: it combines the conditioning of linear systems of equations with the geometry of orthogonal projection. It is important because of the implications for the stability of least-squares algorithms.

◇ FOUR CONDITIONING PROBLEMS

Recall the least-squares problem (Lecture 11), illustrated geometrically as follows:



We assume $\mathbf{A}$ has full rank, and let $\|\cdot\| = \|\cdot\|_2$.

Then, given $\mathbf{A} \in \mathbb{R}^{m\times n}$, $m \geq n$, and $\mathbf{b} \in \mathbb{R}^m$, find $\mathbf{x} \in \mathbb{R}^n$ such that $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|$ is minimized.

The solution $\mathbf{x}$ and the closest point $\mathbf{y} = \mathbf{A}\mathbf{x}$ to $\mathbf{b}$ are

$$\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b} \qquad \text{and} \qquad \mathbf{y} = \mathbf{P}\mathbf{b}$$

where $\mathbf{A}^{\dagger} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the pseudoinverse of $\mathbf{A}$ and $\mathbf{P} = \mathbf{A}\mathbf{A}^{\dagger}$ is the orthogonal projector onto $\mathrm{range}(\mathbf{A})$.

We consider the conditioning of the least-squares problems with respect to perturbations.

(Recall conditioning pertains to the sensitivity of solutions to perturbations in the data).

We will look at this in two ways for the least-squares problem. In either case, the data are $\mathbf{A}$ and $\mathbf{b}$
$\rightarrow$ we can perturb one of these.

The solution can be considered to be either $\mathbf{x}$ or $\mathbf{y}$.

(This is how we get 4 conditioning problems!)

$\diamond$ THEORETICAL RESULT

We have theory that provides answers to these questions.

The results involve 3 (dimensionless) parameters that appear repeatedly in analyzing least-squares problems.

They are

1. $\kappa(\mathbf{A})$ – the condition number of $\mathbf{A}$.

   Recall, if $\mathbf{A}$ is square, $\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$.

   If $\mathbf{A}$ is rectangular, this definition generalizes to

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

   Note: $1 \leq \kappa(\mathbf{A}) \leq \infty$.

2. angle θ (recall figure)

   This is a measure of how close $\mathbf{b}$ is to $\mathbf{y}$:

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

   Note: $0 \leq \theta \leq \frac{\pi}{2}$.

3. η, a measure of how much $\|\mathbf{y}\|$ falls short of its maximum value, given $\|\mathbf{A}\|$ and $\|\mathbf{x}\|$

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

   Note: $1 \leq \eta \leq \kappa(\mathbf{A})$.

Theorem 1. *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, $m \geq n$, with full rank and $\mathbf{b} \in \mathbb{R}^m$ be given.*
The least-squares problem has the following 2-norm relative condition numbers describing the sensitivities of $\mathbf{y}$ and $\mathbf{x}$ with respect to perturbations in $\mathbf{b}$ and $\mathbf{A}$:

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

Note 1. *The results in the first row are exact (tight), while the results in the second row are upper bounds.*

Note 2. *When $m = n$ (i.e., we have a square, nonsingular system and $\theta = 0$) the bounds in the second column reduce to $\frac{\kappa(\mathbf{A})}{\eta}$ and $\kappa(\mathbf{A})$.*
The results in the first column are no longer relevant.

$\diamond$ TRANSFORMATION TO A DIAGONAL MATRIX

We will provide simple proofs of the conditioning results with respect to perturbations in $\mathbf{b}$ only.

(The full proofs of conditioning results with respect to perturbations in $\mathbf{A}$ are in the text, pp. 133-135.)

First, we transform the general least-squares problem to a convenient diagonal problem.

How do we do this? SVD!

Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

Because perturbations are measured by their 2-norms, they are unaffected by the orthogonal transformations $\mathbf{U}, \mathbf{V}^T$.

i.e., the perturbation behaviour of $\mathbf{A}$ is the same as that of $\mathbf{\Sigma}$!

Thus, without loss of generality, we can forget about $\mathbf{A}$ and deal with $\mathbf{\Sigma}$ directly.

So, let

$$\mathbf{A} = \mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & \mathbf{0} & \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{0} \end{bmatrix}$$

The orthogonal projection of $\mathbf{b}$ onto $\mathrm{range}(\mathbf{A})$ is now trivial.

If

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

where $\mathbf{b}_1 \in \mathbb{R}^n$, then

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

The corresponding $\mathbf{x}$ can be found from $\mathbf{A}\mathbf{x} = \mathbf{y}$.

i.e.,

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

or

$$\mathbf{x} = \mathbf{A}_1^{-1}\mathbf{b}_1.$$

From the results, it is easy to see

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{A}^\dagger = \begin{bmatrix} \mathbf{A}_1^{-1} & \mathbf{0} \end{bmatrix}$$

◇ SENSITIVITY OF $\mathbf{y}$ TO PERTURBATIONS IN $\mathbf{b}$

$$\mathbf{y} = \mathbf{P}\mathbf{b}$$

∴ Jacobian of mapping $\mathbf{b} \rightarrow \mathbf{y}$ is $\mathbf{J} = \mathbf{P}$
and $\|\mathbf{P}\| = 1$. (why?)

Thus, the condition number of $\mathbf{y}$ with respect to perturbations in $\mathbf{b}$ is

$$\kappa_{\mathbf{b}\rightarrow\mathbf{y}} = \frac{\|\mathbf{P}\|}{\|\mathbf{y}\|/\|\mathbf{b}\|} = \frac{1}{\cos(\theta)}$$

$\diamond$ SENSITIVITY OF $\mathbf{x}$ TO PERTURBATIONS IN $\mathbf{b}$

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

$\therefore$ Jacobian of mapping $\mathbf{b} \rightarrow \mathbf{x}$ is $\mathbf{J} = \mathbf{A}^{\dagger}$.

Thus, the condition number of $\mathbf{x}$ with respect to perturbations in $\mathbf{b}$ is

$$\begin{aligned}
\kappa_{\mathbf{b}\rightarrow\mathbf{x}} &= \frac{\|\mathbf{A}^{\dagger}\|}{\|\mathbf{x}\|/\|\mathbf{b}\|} \\
&= \frac{\|\mathbf{A}^{\dagger}\|\|\mathbf{b}\|\|\mathbf{y}\|}{\|\mathbf{y}\|\|\mathbf{x}\|} \\
&= \|\mathbf{A}^{\dagger}\|\frac{1}{\cos(\theta)}\frac{\|\mathbf{A}\|}{\eta} \\
&= \frac{\kappa(\mathbf{A})}{\eta\cos(\theta)}
\end{aligned}$$