Lecture 19: Conditioning and Noise Amplification

In the previous lecture, we saw that the truncated SVD

$$A_s = \sum_{j=1}^{s} \sigma_j u_j v_j^*$$

is the best rank-s approximation to A.

We now ask a different question:

What happens when we solve a linear system using noisy data?

1 Solving a Noisy Linear System

Let

$$A \in \mathbb{C}^{n \times n}$$

be invertible, and suppose

$$Ax = b.$$

The exact solution is

$$x = A^{-1}b.$$

Suppose instead that we observe

$$b' = b + \varepsilon,$$

where ε is a small noise vector. If we solve

$$Ay = b',$$

then

$$y = A^{-1}(b + \varepsilon),$$

so

$$y - x = A^{-1}\varepsilon.$$

Therefore,

$$\boxed{\|y - x\|_2 \leq \left\|A^{-1}\right\|_{2\to 2} \|\varepsilon\|_2 .}$$

Thus even small noise can be amplified if

$$\left\|A^{-1}\right\|_{2\to 2}$$

is large.

2 What the SVD Says

Let

$$A = U\Sigma V^*$$

be a full SVD, with

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0.$$

Then

$$A^{-1} = V\Sigma^{-1}U^*,$$

where

$$\Sigma^{-1} = \operatorname{diag}\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_n}\right).$$

Hence

$$\boxed{\left\|A^{-1}\right\|_{2\to 2} = \frac{1}{\sigma_n}.}$$

Also,

$$Av_j = \sigma_j u_j$$

implies

$$A^{-1}u_j = \frac{1}{\sigma_j}v_j.$$

Thus noise in the direction u_j is amplified by the factor

$$\frac{1}{\sigma_j}.$$

The smallest singular value produces the largest possible amplification.

3 Geometric Interpretation

The matrix A maps the unit sphere to an ellipsoid with principal semi-axis lengths

$$\sigma_1, \ldots, \sigma_n.$$

If σ_n is very small, then A nearly collapses one direction. The inverse must undo that collapse, so it stretches that direction by

$$\frac{1}{\sigma_n}.$$

Thus

a very small singular value creates a very sensitive inverse.

4 Relative Error and the Condition Number

Since

$$y - x = A^{-1}\varepsilon,$$

we have

$$\frac{\|y-x\|_2}{\|x\|_2} \le \left\|A^{-1}\right\|_{2\to 2} \frac{\|\varepsilon\|_2}{\|x\|_2}.$$

Also,

$$\|b\|_2 = \|Ax\|_2 \le \|A\|_{2\to 2} \|x\|_2 .$$

Therefore,

$$\frac{1}{\|x\|_2} \le \frac{\|A\|_{2\to 2}}{\|b\|_2}.$$

Combining these,

$$\boxed{\frac{\|y-x\|_2}{\|x\|_2} \le \|A\|_{2\to 2} \left\|A^{-1}\right\|_{2\to 2} \frac{\|\varepsilon\|_2}{\|b\|_2}.}$$

Definition (Condition Number)**.** *For an invertible matrix A,*

$$\boxed{\kappa(A) = \|A\|_{2\to 2} \left\|A^{-1}\right\|_{2\to 2} = \frac{\sigma_1}{\sigma_n}.}$$

Thus

$$\boxed{\frac{\text{relative solution error}}{\text{relative data error}} \lesssim \kappa(A).}$$

A matrix with a large condition number is called *ill-conditioned*.

Absolute vs. Relative Error Amplification

Consider a linear system

$$Ax = b.$$

If b is perturbed to $b + \delta b$, the solution changes to $x + \delta x$, where

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

Absolute error.

$$\boxed{\|\delta x\|_2 \le \|A^{-1}\|_2 \|\delta b\|_2.}$$

The worst-case amplification of absolute error is $\|A^{-1}\|_2 = 1/\sigma_{\min}(A)$.

Relative error. Since $\|b\|_2 \le \|A\|_2 \|x\|_2$,

$$\boxed{\frac{\|\delta x\|_2}{\|x\|_2} \le \kappa(A) \frac{\|\delta b\|_2}{\|b\|_2},}$$

where

$$\kappa(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}.$$

Example. Let

$$A = \begin{pmatrix} 10 & 0 \\ 0 & 0.1 \end{pmatrix}, \qquad \|A^{-1}\|_2 = 10, \qquad \kappa(A) = 100.$$

Choose

$$x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad b = Ax = \begin{pmatrix} 10 \\ 0 \end{pmatrix}.$$

Suppose

$$\delta b = \begin{pmatrix} 0 \\ 0.01 \end{pmatrix}.$$

Then

$$\delta x = A^{-1}\delta b = \begin{pmatrix} 0 \\ 0.1 \end{pmatrix}.$$

Absolute error:

$$\frac{\|\delta x\|_2}{\|\delta b\|_2} = \frac{0.1}{0.01} = \boxed{10}.$$

Relative error:

$$\frac{\|\delta b\|_2}{\|b\|_2} = \frac{0.01}{10} = 0.001,$$

while

$$\frac{\|\delta x\|_2}{\|x\|_2} = \frac{0.1}{1} = 0.1.$$

Therefore,

$$\boxed{\frac{\|\delta x\|_2/\|x\|_2}{\|\delta b\|_2/\|b\|_2} = 100 = \kappa_2(A).}$$

Interpretation. The matrix A stretches the first coordinate by 10 and compresses the second coordinate by 10.

The original solution x lies in the direction of maximum stretching, while the perturbation δb lies in the direction of maximum amplification by A^{-1}.

Thus, the absolute error is amplified by 10, while the relative error is amplified by 100.

Takeaway:

$$\boxed{\begin{array}{ll} \text{Absolute error:} & \|A^{-1}\|_2 = 10, \\ \text{Relative error:} & \kappa_2(A) = 100. \end{array}}$$

Both are worst-case amplification factors, not amplification factors for every perturbation.

5 Some Basic Properties

Since

$$\sigma_1 \geq \sigma_n > 0,$$

we have

$$\boxed{\kappa(A) \geq 1.}$$

If U is unitary, then

$$\boxed{\kappa(U) = 1.}$$

For every nonzero scalar α,

$$\boxed{\kappa(\alpha A) = \kappa(A).}$$

Scaling changes all singular values by the same factor and therefore does not change the condition number.

Exercise 1. *Suppose*

$$Av_j = \sigma_j u_j.$$

Show that

$$A^{-1}u_j = \frac{1}{\sigma_j} v_j.$$

Use this to explain why small singular values are dangerous when solving a noisy system.

Lecture 20: Regularization Through the SVD

6 Why Do We Need Regularization?

Recall from the previous lecture that, for an invertible matrix $A \in \mathbb{C}^{n\times n}$ with SVD

$$A = \sum_{j=1}^{n} \sigma_j u_j v_j^*, \qquad \sigma_1 \geq \cdots \geq \sigma_n > 0,$$

the solution of $Ax = b$ is

$$\boxed{x = A^{-1}b = \sum_{j=1}^{n} \frac{u_j^* b}{\sigma_j} v_j.}$$

We have seen that small singular values can strongly amplify noise in b.

But there is another problem. Since the right singular vectors are orthonormal,

$$\boxed{\|x\|_2^2 = \sum_{j=1}^{n} \frac{|u_j^* b|^2}{\sigma_j^2}.}$$

As $\sigma_n \to 0$, if $|u_n^* b|$ does not approach zero sufficiently fast, then

$$\|x\|_2 \to \infty.$$

Thus, small singular values may lead to both

- strong amplification of noise;
- a solution with an excessively large norm.

This suggests a new goal:

Find a solution that fits the data without having an excessively large norm.

7 Least Squares with a Norm Penalty

We now allow $A \in \mathbb{C}^{m\times n}$ to be rectangular or rank-deficient, and suppose that the observed data is

$$b' = Ax_{\text{true}} + \varepsilon.$$

The ordinary least-squares problem is

$$\min_x \left\|Ax - b'\right\|_2^2.$$

Recall that its normal equation is

$$\boxed{A^*Ax = A^*b'.}$$

If A is rank-deficient, this problem has infinitely many solutions. Even if A has full column rank, small singular values may produce a solution with a very large norm.

To discourage large solutions, we add a penalty:

$$\boxed{x_\alpha = \arg\min_x \left\{ \left\| Ax - b' \right\|_2^2 + \alpha \left\| x \right\|_2^2 \right\}, \qquad \alpha > 0.}$$

This is called *Tikhonov regularization*.

The two terms have different purposes:

- $\|Ax - b'\|_2^2$ measures the data misfit.
- $\|x\|_2^2$ penalizes large solutions.

The parameter α controls the tradeoff.

8 The Regularized Normal Equation

Setting the gradient of the objective to zero gives

$$2A^*(Ax_\alpha - b') + 2\alpha x_\alpha = 0.$$

Therefore,

$$\boxed{(A^*A + \alpha I)x_\alpha = A^*b'.}$$

For any nonzero $z \in \mathbb{C}^n$,

$$\begin{aligned} z^*(A^*A + \alpha I)z &= \|Az\|_2^2 + \alpha \|z\|_2^2 \\ &> 0. \end{aligned}$$

Thus $A^*A + \alpha I$ is positive definite, even when A is rank-deficient.

Hence the regularized problem has a unique solution:

$$\boxed{x_\alpha = (A^*A + \alpha I)^{-1} A^*b'.}$$

9 Tikhonov Regularization Through the SVD

Let A have rank r and compact SVD

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*, \qquad \sigma_1 \geq \cdots \geq \sigma_r > 0.$$

Recall that

$$A^*A = \sum_{j=1}^{r} \sigma_j^2 v_j v_j^*, \qquad A^*b' = \sum_{j=1}^{r} \sigma_j (u_j^* b') v_j.$$

Using the regularized normal equation, we obtain

$$\boxed{x_\alpha = \sum_{j=1}^{r} \frac{\sigma_j}{\sigma_j^2 + \alpha} (u_j^* b') v_j.}$$

Compare this with the minimum-norm least-squares solution

$$x^\dagger = A^\dagger b' = \sum_{j=1}^{r} \frac{u_j^* b'}{\sigma_j} v_j.$$

We can rewrite the Tikhonov coefficients as

$$\frac{\sigma_j}{\sigma_j^2 + \alpha} = \underbrace{\frac{\sigma_j^2}{\sigma_j^2 + \alpha}}_{\text{shrinkage factor}} \frac{1}{\sigma_j}.$$

Therefore, Tikhonov regularization shrinks each pseudoinverse coefficient by a factor

$$\boxed{0 < \frac{\sigma_j^2}{\sigma_j^2 + \alpha} < 1.}$$

If $\sigma_j^2 \gg \alpha$, then

$$\frac{\sigma_j^2}{\sigma_j^2 + \alpha} \approx 1,$$

so the corresponding direction is almost unchanged.

If $\sigma_j^2 \ll \alpha$, then

$$\frac{\sigma_j^2}{\sigma_j^2 + \alpha} \approx 0,$$

so the corresponding direction is strongly suppressed.

Thus Tikhonov regularization reduces the influence of small singular values without completely discarding their directions.

This is a form of *soft shrinkage*.

What happens as $\alpha \to 0$?

For every nonzero singular value,

$$\lim_{\alpha \to 0^+} \frac{\sigma_j}{\sigma_j^2 + \alpha} = \frac{1}{\sigma_j}.$$

Therefore,

$$\boxed{\lim_{\alpha \to 0^+} x_\alpha = A^\dagger b'.}$$

So Tikhonov regularization approaches the minimum-norm least-squares solution as $\alpha \to 0^+$.

This remains true when A is rank-deficient.

10 Hard Thresholding: Truncated SVD

Tikhonov regularization smoothly reduces the contributions from small singular values.

There is another possibility: instead of shrinking these contributions, we can discard them completely.

Recall the compact SVD

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*.$$

Choose $s \leq r$ and keep only the largest s singular values:

$$A_s = \sum_{j=1}^{s} \sigma_j u_j v_j^*.$$

Recall that A_s is the best rank-s approximation to A in both the operator norm and the Frobenius norm.

We define the truncated SVD solution by

$$\boxed{x_s = A_s^\dagger b' = \sum_{j=1}^{s} \frac{u_j^* b'}{\sigma_j} v_j.}$$

Equivalently, the inverse factors are replaced by

$$\boxed{f_j = \begin{cases} 1/\sigma_j, & j \leq s, \\ 0, & j > s. \end{cases}}$$

This is called *hard thresholding*.

Unlike Tikhonov regularization, which smoothly reduces each coefficient, truncated SVD either keeps a singular direction or removes it completely.

If $s = r$, then

$$x_s = A^\dagger b'.$$

If $s < r$, we deliberately discard some nonzero singular directions to improve stability.

11 The Tradeoff Between Bias and Noise

Why not simply discard all small singular values?

Suppose

$$b' = Ax_{\text{true}} + \varepsilon.$$

Write

$$x_{\text{true}} = \sum_{j=1}^{r} \xi_j v_j + z, \qquad z \in \text{Null}(A).$$

The truncated SVD solution is

$$x_s = \sum_{j=1}^{s} \xi_j v_j + \sum_{j=1}^{s} \frac{u_j^* \varepsilon}{\sigma_j} v_j.$$

Thus

$$x_{\text{true}} - x_s = z + \sum_{j=s+1}^{r} \xi_j v_j - \sum_{j=1}^{s} \frac{u_j^* \varepsilon}{\sigma_j} v_j.$$

The three terms are orthogonal, giving

$$\|x_{\text{true}} - x_s\|_2^2 = \underbrace{\|z\|_2^2}_{\text{unrecoverable component}} + \underbrace{\sum_{j=s+1}^{r} |\xi_j|^2}_{\text{discarded signal}} + \underbrace{\sum_{j=1}^{s} \frac{|u_j^* \varepsilon|^2}{\sigma_j^2}}_{\text{amplified noise}}.$$

Keeping more singular directions preserves more information, but may also amplify more noise.

Discarding more directions reduces noise amplification, but introduces additional bias.

Neither method can recover components in $\text{Null}(A)$ without additional information.

12 Example: Soft Shrinkage vs. Hard Thresholding

Consider

$$A = \begin{pmatrix} 10 & 0 \\ 0 & 0.1 \end{pmatrix}, \qquad b' = \begin{pmatrix} 10 \\ 0.01 \end{pmatrix}.$$

The ordinary solution is

$$x = A^{-1} b' = \begin{pmatrix} 1 \\ 0.1 \end{pmatrix}.$$

Tikhonov Regularization

With $\alpha = 0.01$,

$$x_\alpha = \begin{pmatrix} \dfrac{100}{100.01} \\ \dfrac{0.001}{0.02} \end{pmatrix} \approx \begin{pmatrix} 0.9999 \\ 0.05 \end{pmatrix}.$$

The contribution from the small singular value is reduced from 0.1 to 0.05.

Truncated SVD

If we keep only the largest singular value,

$$x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The small singular direction is discarded completely.

13 Comparison

Both methods modify the SVD inverse factors:

Method	Inverse factor
Pseudoinverse	$\dfrac{1}{\sigma_j}$
Tikhonov	$\dfrac{\sigma_j}{\sigma_j^2 + \alpha}$
Truncated SVD	$\begin{cases} \dfrac{1}{\sigma_j}, & j \le s, \\ 0, & j > s. \end{cases}$

Regularization controls noise amplification by suppressing small singular directions, at the cost of introducing bias.

Exercises

Exercise 2. *Let*

$$A = \begin{pmatrix} 10 & 0 \\ 0 & 0.1 \end{pmatrix}, \qquad b = \begin{pmatrix} 10 \\ 0.01 \end{pmatrix}.$$

Compute the Tikhonov solution with $\alpha = 0.1$ and compare its norm with that of $A^{-1}b$.

Exercise 3. *Suppose A has singular values*

$$10, \quad 1, \quad 0.01.$$

For $\alpha = 0.01$, compute the Tikhonov shrinkage factors

$$\frac{\sigma_j^2}{\sigma_j^2 + \alpha}.$$

Which singular direction is suppressed the most?

Exercise 4. *Let A have rank r. Show that the truncated SVD solution with $s = r$ equals the minimum-norm least-squares solution.*

Exercise 5. *Explain why Tikhonov regularization has a unique solution even if A is rank-deficient.*

14 The Moore–Penrose Pseudoinverse

For an invertible square matrix,

$$Ax = b$$

has the unique solution

$$x = A^{-1}b.$$

But many matrices are rectangular or rank-deficient, so A^{-1} does not exist.

The SVD gives us a natural generalization of the inverse.

Let

$$A \in \mathbb{C}^{m \times n}$$

have rank r and compact SVD

$$A = U_r \Sigma_r V_r^*,$$

where

$$\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r), \qquad \sigma_j > 0.$$

Since all entries of Σ_r are positive, we can invert them.

Definition (Moore–Penrose Pseudoinverse)**.** *The pseudoinverse of A is*

$$\boxed{A^\dagger = V_r \Sigma_r^{-1} U_r^*,}$$

where

$$\Sigma_r^{-1} = \operatorname{diag}\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r}\right).$$

Thus,

$$\boxed{A^\dagger = \sum_{j=1}^{r} \frac{1}{\sigma_j} v_j u_j^*.}$$

The idea is simple:

$$Av_j = \sigma_j u_j$$

and therefore the pseudoinverse reverses this action:

$$\boxed{A^\dagger u_j = \frac{1}{\sigma_j} v_j, \qquad j = 1, \ldots, r.}$$

It inverts A on the directions where A is nonzero.

14.1 Pseudoinverse and Projections

Using the compact SVD,

$$A^\dagger A = V_r \Sigma_r^{-1} U_r^* U_r \Sigma_r V_r^* = V_r V_r^*.$$

Since the columns of V_r form an orthonormal basis for $\mathcal{C}(A^*)$,

$$\boxed{A^\dagger A = P_{\mathcal{C}(A^*)}.}$$

Similarly,

$$AA^\dagger = U_r U_r^*,$$

so

$$\boxed{AA^\dagger = P_{\mathcal{C}(A)}.}$$

Thus $A^\dagger$ is not generally a true inverse:

$$A^\dagger A \neq I, \qquad AA^\dagger \neq I.$$

Instead, these products are orthogonal projections.

14.2 Connection to Least Squares

For any

$$b \in \mathbb{C}^m,$$

define

$$\hat{x} = A^\dagger b.$$

Then

$$A\hat{x} = AA^\dagger b = P_{\mathcal{C}(A)} b.$$

Therefore,

$$b - A\hat{x} = b - P_{\mathcal{C}(A)} b \in \mathcal{C}(A)^\perp.$$

Hence

$$\boxed{\hat{x} = A^\dagger b}$$

is a least-squares solution of

$$Ax \approx b.$$

If the least-squares solution is not unique, $A^\dagger b$ is the one with smallest Euclidean norm.

14.3 Special Cases

If A is square and invertible, then

$$\boxed{A^\dagger = A^{-1}.}$$

If A has full column rank, then

$$\boxed{A^\dagger = (A^*A)^{-1}A^*.}$$

Thus the familiar least-squares formula

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

is simply

$$\boxed{\hat{x} = A^\dagger b.}$$

14.4 Example

Let

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Its nonzero singular values are

$$2, \qquad 1,$$

so

$$A^\dagger = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then

$$A^\dagger A = I_2,$$

while

$$AA^\dagger = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which is the orthogonal projection onto $\mathcal{C}(A)$.