Lecture 20: Regularization Through the SVD

1 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.

2 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.

3 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:

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

4 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.

5 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

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

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.

6 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.

7 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.